Derecho e inteligencia artificial

Gonzalo Ana Dobratinich
(Director)

LA LEY

FACULTAD DE DERECHO U.B.A.

Thomson Reuters™

Derecho e inteligencia artificial

Gonzalo Ana Dobratinich
(director)

LA LEY

Thomson Reuters™

FEDYE
FONDO EDITORIAL
DE DERECHO Y ECONOMIA

AUTORIDADES DE LA FACULTAD DE DERECHO DE LA UNIVERSIDAD DE BUENOS AIRES

Decano
Leandro Vergara

Vicedecana
Silvia Nonna

CONSEJO DIRECTIVO DE LA FACULTAD DE DERECHO

CLAUSTRO DE PROFESORES

Consejeros Titulares
Mary Beloff / Juan Pablo Mas Velez / Marcela Basterra / Gonzalo Álvarez / Graciela Medina / Daniel Roque Vítolo / Leila Devia / Alejandro Alagia

Consejeros Suplentes
Adelina Loianno / Marcelo Gebhardt / Silvia Nonna/ Juan Pablo Mugnolo / Nancy Cardinaux / Alfredo Vítolo / María Blanca Noodt Taquela / Claudio Martiniuk

CLAUSTRO DE GRADUADOS

Consejeros Titulares
Raúl M. Alfonsín / Silvia Bianco / Carlos Rodríguez / Adrián Famá

Consejeros Suplentes
Manuel Negrón Mansilla / María Laura Elliff/ Juan Pablo Schnaiderman / María Sol Plata

CLAUSTRO DE ESTUDIANTES

Consejeros Titulares
Camila M. A. Gorski / Tomás Fischer / Julieta P. Vilchez / Martina Rincón

Consejeros Suplentes
Alexander G. Quiroz / Micaela Castañeda / Matías Gurruchaga / Maximiliano E. Rey

Representante No Docente
Lorena Castaño

Secretarios
Secretario Académico: Lucas G. Bettendorff
Secretario de Administración: Carlos A. Bedini
Secretario de Extensión Universitaria y Bienestar Estudiantil: Oscar M. Zoppi
Secretaria de Investigación: Luciana B. Scotti
Secretaria de Coordinación y Relaciones Institucionales: Marta R. Vigevano
Secretaria de Relaciones Institucionales: Carmen Virginia Badino Varela
Secretaria de Vinculación Estudiantil: Andrea Cristina Luciana Carreras Lobo

DEPARTAMENTO DE PUBLICACIONES
Directora: Mary Beloff
Subdirector: Luis R. J. Sáenz
Secretaria: Mayra Danae Salvatierra

Ana Dobratinich, Gonzalo
Derecho e inteligencia artificial / Gonzalo Ana Dobratinich. - 1a ed. - Ciudad Autónoma de Buenos Aires : La Ley, 2026.
Libro digital, PDF

Archivo Digital: descarga y online
ISBN 978-987-03-5206-8

1. Inteligencia Artificial. 2. Derecho. I. Título.
CDD 340

© Departamento de Publicaciones - Facultad de Derecho UBA, 2026
Av. Pte. Figueroa Alcorta 2263 (C1425CKB) Buenos Aires

© De esta edición, La Ley S.A.E. e I., 2026
Tucumán 1471 (C1050AAC) Buenos Aires
Queda hecho el depósito que previene la ley 11.723

ISBN 978-987-03-5206-8

Argentina

Índice general

Introducción

Derecho artificial

Por Gonzalo Ana Dobratinich

1. Introducción 4
2. Pensamiento seriado 5
3. La muerte del derecho 6
4. Los dones robados-perdidos 7
5. ¿Qué derecho nos ha quedado? 7
6. Pensamientos, secciones y obra 8
7. Conclusiones 10

Bibliografía 11

Sección I

Filosofía, Justicia, Humano/Máquina

Las alucinaciones en la inteligencia artificial, su utilidad en la investigación: desafíos y oportunidades

Por Horacio Roberto Granero

1. Introducción 16
2. Qué significa que la IA pueda alucinar? 18
3. La serendipia: ventajas potenciales para la exploración científica 19
4. Aprovechando las alucinaciones de IA para la investigación científica 24
5. Buenas prácticas para minimizar las alucinaciones perjudiciales 25
6. La controversia: ¿debemos aceptar o eliminar las alucinaciones? 27
7. Hacia una alucinación de IA responsable 28
8. Conclusión: ¿La IA daña la creatividad? o ¿Cómo puede la IA mejorarla? 31

Referencias 33

Pág.

¿QUIÉN CUIDA CUANDO CUIDA UNA MÁQUINA? DERECHO, AFECTO Y ADJUDICACIÓN EN CLAVE TRIALISTA

Por Maria Isolina Dabove

1. Introducción 38
2. Filosofía jurídica trialista para los cuidados 39
3. Adjudicación algorítmica en la administración de cuidados 42
4. Justicia intergeneracional y tensiones en torno a la dignidad 43
5. Los desafíos normativos ante inteligencia artificial: ¿cómo integrar las lagunas jurídicas? 45
6. Filosofía del afecto y ética de los cuidados: el horizonte del derecho 47
7. Conclusiones: *poiésis* jurídica, cuidado y afecto en la era digital 50

BRAIN ROT E O ESTADO DEMOCRÁTICO DE DIREITO: DA MORTE DA INTELECTUALIDADE À MODERNIZAÇÃO AUTORITÁRIA DA DECISÃO JURÍDICA

Por Lenio Luiz Streck

1. Considerações iniciais: "a questão da inteligência artificial" em um Estado Democrático de Direito 52
2. Brain rot, a morte da intelectualidade e a modernização autoritária da decisão jurídica: a linguagem como a casa do ser (humano, demasiadamente humano) 54
3. Uma distopia assustadoramente próxima do nosso Zeitgeist (espírito do tempo): os algoritmos e o fim dos advogados 57
4. Considerações finais: angustiemo-nos 62

Referências 62

INTELIGENCIA ARTIFICIAL COMO ARQUETIPO DE LA INTELIGENCIA HUMANA

Por Silvia Adriana Ramos

1. La inteligencia y el arquetipo 64
   1.1. La idea de arquetipo 64
   1.2. El error de la definición 65
   1.3. La inteligencia que tiene (o no) la inteligencia artificial 66
2. Necesidad y derecho de la inteligencia artificial 66
   2.1. Ya no podemos vivir sin usar inteligencia artificial 66
   2.2 La propiedad de lo creado con la inteligencia artificial 67
   2.3 La ética del uso de la inteligencia artificial 68
3. Derecho e inteligencia artificial 69
   3.1. La inteligencia artificial como juez supuestamente imparcial 69
   3.2. Las consecuencias jurídicas de los actos de las herramientas de inteligencia artificial 70

Pág.
4. El arquetipo y la representación 70
Bibliografía 71

Simplificar es complejo. Ayuda-memoria *EXPRESS* para abordar los "desafíos" de la IA aplicada al derecho

Por Gonzalo S. Aguirre

1. Desarrollo 74
Bibliografía de base 76

Inteligencia artificial y autonomía: un análisis bioético

Por Rab. Dr. Fishel Szlajen

1. Autonomía moral vs. autonomía operativa 80
2. Eticidad, responsabilidad y derechos 85
3. Conclusión 90
Bibliografía 91

Kant, reflexivity, and artificial intelligence

Por Nythamar de Oliveira

1. Introduction 96
2. Rational Psychology Before Kant 99
    2.1. Descartes and the Rational Soul 99
    2.2. Spinoza and the *Conatus* of Mind 99
    2.3. Hume and the Empirical Self 100
    2.4. Strawson and the Analytic Rehabilitation of Personhood 100
    2.5. From Rational Psychology to Transcendental Philosophy 100
3. Kant on Reflexivity and Judgment 101
    3.1. The Transcendental Turn: From Rational Psychology to Critical Philosophy 101
    3.2. Judgment as the Mediating Faculty 101
    3.3. The Aesthetic and Teleological Dimensions of Reflective Judgment 102
    3.4. Reflexivity in Theoretical and Practical Uses of Reason 102
    3.5. The Crossings of Determining and Reflective Judgment 103
    3.6. Reflexivity Beyond Syntax 103
4. Reflexivity and Normativity 104
    4.1. Humanity as an End in Itself 104
    4.2. The Unity of Reason: Theoretical and Practical 104
    4.3. The Value of Dignity 104
    4.4. The Phenomenological Deficit of Naturalism 105
    4.5. Transcendental Semantics and the Horizon of Meaning 105
    4.6. Normativity as Sustainability 106

Pág.
5. Rawls's Normative Conception of the Person 107
5.1. From Kantian Autonomy to Rawlsian Constructivism 107
5.2. Reflective Equilibrium as Procedural Reflexivity 107
5.3. Autonomy, Personality, and Moral Powers 108
5.4. Responses to Communitarian Critiques 108
5.5. Reflexivity, Personhood, and Democratic Citizenship 109
6. Constructivism, Personhood, and Reflective Equilibrium 109
6.1. What "constructivism" means in a Kant–Rawls key 109
6.2. Personhood as the regulative core of constructivism 110
6.3. Reflective equilibrium as institutionalized reflexivity 111
6.4. Answering the communitarian and post-Kantian challenges ... 111
6.5. Coherentism without collapse: truth, reasonableness, and fact-sensitivity 112
6.6. Why this triad matters for AI 112
6.7. A constructive synthesis 113
7. Reflexivity, AI, and Humanity 114
7.1. Why Reflexivity Matters for AI 114
7.2. Wide Reflective Equilibrium and Machine Learning 114
7.3. The Phenomenological Deficit in AI 114
7.4. Humanity as Sustainable *Conatus* 115
7.5. Quantum Phenomenology and the Limits of Computation 115
7.6. AI and the Risk of Dehumanization 116
7.7. A Rawlsian Extension: AI and the Basic Structure 116
7.8. The Irreducibility of Humanity 116
8. Towards a Kantian Ethics of AI Value Alignment 117
8.1. From Technical Alignment to Normative Alignment 117
8.2 Kant's Humanity Formula as the First Principle of Alignment ... 117
8.3. Autonomy and Contestability 117
8.4. Justice and the Basic Structure 118
8.5. Sustainability and Humanity as *Conatus* 118
8.6. Value Alignment as Reflexive Alignment 119
8.7. A Pragmatist Horizon: Humanity as Self-Transcending 119
9. Concluding Remarks 119
References 121

SENTIR E SIMULAR: RELAÇÕES HUMANO-MÁQUINA

Por Eduardo R. Rabenhorst

1. Introdução: uma gravura inquietante 126
2. Sonho, desejo e opacidade 127
3. Erotismo, alteridade e anorgasmia maquinal 130
4. Sentir e simular: a autenticidade do vínculo e a falsa reciprocidade 133
5. A ética do entremeio 135
Referências 136

Pág.

La inteligencia artificial y el retorno a una concepción y administración del derecho como divina

Por Fernando Beresñak

1. Breves e iniciales reflexiones para abordar el problema de la inteligencia artificial en el derecho 140
   1.1. Problemática y abordaje 140
   1.2. Proposición tripartita 141
   1.3. Hipótesis 142
2. Escenas hipotético-especulativas y las primeras formas del derecho .. 143
   2.1. La aparición del habla y las reglas de la lengua 143
   2.2. Un mundo de reglas 144
3. Consideraciones sobre el derecho y el retorno de lo divino mediante la inteligencia artificial 145
Bibliografía 146

Inteligencia artificial y principio de responsabilidad: más allá del derecho a la explicación

Por Luciano D. Laise y Eduardo E. Magoja

1. Introducción 150
2. ¿Qué es la inteligencia artificial? 153
   2.1. De la responsabilidad basada en el control a la ilusión del control 154
   2.2. El conocimiento: ¿Mis actos revelan quién soy? 156
3. Una respuesta tentativa: el derecho a la explicación 157
4. Dos vías de solución 158
   4.1. Una explicación contrafáctica 158
   4.2. El derecho al olvido 159
5. Conclusiones 160
6. Referencias bibliográficas 161

Ética(s) de la inteligencia artificial y derecho: consideraciones a propósito de los límites y la contención del desarrollo tecnológico

Por Fernando H. Llano Alonso

1. Introducción 166
2. La IA fuerte y el alineamiento de sus objetivos con los valores humanos 171
3. La ética de la IA y la ampliación de los principios de la bioética 174
4. Diversos enfoques y modelos éticos de la IA 180
5. A modo de conclusión: hacia una metaética de la IA 183

Pág.

EXISTE UM DIREITO FUNDAMENTAL À REVISÃO HUMANA?

Por André Karam Trindade y Amanda Antonelo

1. Introdução 186
2. Da decisão humana à decisão artificial 188
3. Entre a razão humana e a promessa algorítmica 190
4. O direito fundamental à revisão humana 193
Referências bibliográficas 196

SECCIÓN II

PROFESIÓN Y ENSEÑANZA JURÍDICA

EL FUTURO DE LA INTELIGENCIA ARTIFICIAL GENERATIVA EN EL DERECHO

Por Antonio Anselmo Martino

1. El futuro de la inteligencia artificial generativa 203
2. El futuro de la IA generativa en el derecho 205
   2.1. Aplicación del derecho en la sociedad 206
   2.2. Desafíos y consideraciones éticas 206
   2.3. Seguirán desarrollándose las líneas teóricas sobre IA generativa en el derecho 206
   2.4. Modelos regulatorios comparados 207
   2.5 ¿Qué son los agentes inteligentes en el derecho? 207
3. Casos prácticos emergentes 209
   3.1 Cobotización judicial y transformación institucional: Estudio de caso del sistema Prometea en Argentina 209
4. Consideraciones éticas y desafíos regulatorios 210
5. Conclusiones 211
Bibliografía 213

LA CONSTRUCCIÓN DE *DATASETS* JURISPRUDENCIALES EN EL SISTEMA JUDICIAL Y LOS COLABORADORES EXPERTOS. CARACTERÍSTICAS Y RETOS DE UN NUEVO ROL

Por Helga María Lell

1. Introducción 216
2. Datos y *datasets* 218
3. Los *datasets* jurisprudenciales como base de la Inteligencia Artificial 220
4. Actores institucionales implicados en la construcción y supervisión de los *datasets* jurisprudenciales 222
5. Relevancia del trabajo de los/as expertos/as para los conjuntos de datos jurisdiccionales 223
6. Retos para los expertos en la construcción de *datasets* 225
7. Conclusión 227
Referencias bibliográficas 227

Pág.

Los usos de la IAGEN como objeto de casos judiciales

Por Débora Marhaba y Nicolás Pérez Felicioni

1. Planteo 233
2. "Justicia digital" como punto de partida 234
3. Los casos analizados 236
4. Normas éticas de la abogacía frente a la IA 239
5. La abogacía y la IAGen 241
6. Conclusiones 246

Virtual labs en la educación jurídica: una propuesta de transformación pedagógica mediante inteligencia artificial

Por Jorge A. Cerdio Herrán

1. Introducción 250
2. El imperativo de transformación en la educación jurídica 250
   2.1. La crisis del modelo tradicional ante la IA generativa 251
   2.2. El paradigma del *Virtual Lab*: de la ciencia al derecho 251
3. Arquitectura del *Virtual Lab* jurídico 252
   3.1. Los agentes especializados 253
   3.2. El profesor-orquestador como director del laboratorio 254
4. Metodología de implementación en el contexto jurídico 255
   4.1. El ciclo de descubrimiento jurídico 255
   4.2. Métricas de originalidad jurídica 259
5. Desafíos y consideraciones éticas 260
   5.1. Desafíos pedagógicos e institucionales 261
   5.2. Gobernanza de datos y riesgos sistémicos 262
Bibliografía 265

La inteligencia artificial como clave para recuperar la legitimidad de la justicia argentina

Por Lara Morquecho

1. Introducción 270
2. La legitimidad de la justicia en Iberoamérica: ¿podemos recuperarla con la aplicación de IA? 271
3. Las normas procesales como barrera de contención 273
4. Una nueva herramienta para potenciar la oficiosidad del juez 274
5. La IA como herramienta y parte de un sistema procesal 275
6. Propuesta *de lege ferenda*: impulso procesal oficioso con IA 276
7. Conclusiones 277
Bibliografía 277

Pág.

IMPLICACIONES ÉTICAS Y PEDAGÓGICAS DE LA INTELIGENCIA ARTIFICIAL COMO HERRAMIENTA DE APOYO EN EL ÁMBITO EDUCATIVO

Por Guadalupe Juárez Quezada

1. Introducción 282
2. Antecedentes históricos de la inteligencia artificial y su vínculo con la educación 283
3. La inteligencia artificial generativa (IAG): una herramienta disruptiva en la enseñanza del derecho 285
4. Aplicaciones pedagógicas de la inteligencia artificial en la enseñanza del derecho 286
   4.1. Transformación del proceso de enseñanza-aprendizaje 286
   4.2. Modelos didácticos multimodales 286
5. Implicaciones éticas de la inteligencia artificial generativa 287
   5.1. El dilema de Ícaro: innovación sin límites o caída ética 287
   5.2. Principios de ética académica en el uso de la IA 288
   5.3. Deontología docente y responsabilidad institucional 288
6. Estrategias institucionales para una integración responsable de la IA 289
   6.1. Alfabetización digital ética 289
7. A manera de conclusión. Hacía una pedagogía transdisciplinaria 290
Bibliografía 291

NUEVOS DESAFÍOS PARA LOS PROFESIONALES DEL DERECHO: ÉTICA Y RESPONSABILIDAD EN EL USO DE LA INTELIGENCIA ARTIFICIAL

Por María Victoria Gisvert

1. La automatización en la abogacía 294
2. Lineamientos para un uso ético y responsable de la inteligencia artificial 297
3. Lo que la justicia anda diciendo 298
Conclusión 300
Bibliografía 301

LA INTELIGENCIA ARTIFICIAL PARA UNA ESTRATEGIA PROBATORIA

Por Marcela Tovar Toledo

1. Introducción 304
2. Derecho probatorio: la unión marital de hecho 305
3. Sistema de inteligencia probatoria asistida por IA (SIPRAI) 312
4. Conclusiones 316
Referencias bibliográficas 316

Pág.

Justicia e inteligencia artificial: propuesta de armonización procesal a través de una ley modelo

Por Rosa A. Avila Paz de Robledo y Diego Robledo

1. Introducción 320
2. Marco normativo e inteligencia artificial 321
   2.1. Modelos regulatorios 321
   2.2. Hacia un modelo argentino 323
      2.2.1. Constitución de la Provincia de Jujuy 324
      2.2.2. Constitución de la Provincia de La Rioja 325
      2.2.3. Constitución de la Provincia de Santa Fe 326
      2.2.4. Constitución de la Provincia de Formosa 328
      2.2.5. Balance de derecho comparado 328
3. Hacia la constitución de un ecosistema de justicia digital y algorítmico ... 329
   3.1. Implementación de indicadores de gestión judicial 329
   3.2. Política de comunicación judicial clara como un eje que sostiene el ecosistema 330
   3.3. Inteligencia artificial como estrategia de lenguaje claro y accesible de la jurisprudencia 331
4. Jurisprudencia sobre el uso de la IA generativa por la judicatura y la abogacía 331
   4.1. Abogacía y uso de IA generativa: Caso "Giacomino" 331
   4.2. Judicatura y uso de la IA: Caso "Providencia del Chubut c/P.R.A" 332
5. Conclusiones 334
Bibliografía 335

Apuntes sobre la inteligencia artificial aplicada a la administración de Justicia

Por Sonia Gamboa, Andrés Botero y Kevin Valdivieso

1. Introducción 340
2. La IA. Nociones básicas. 341
3. La IA en el derecho (IA-D) y en lo judicial (IA-J) 346
4. A modo de conclusión 357
Referencias 359

Pág.

SECCIÓN III

NORMAS, LEYES, REGULACIONES

CONTRATACIÓN DE INTELIGENCIA ARTIFICIAL POR PARTE DEL ESTADO: ANÁLISIS LEGAL Y REGULATORIO EN ARGENTINA

Por Manuela Adrogué y Milagros Ibarzábal

1. Introducción 366
2. Qué es la inteligencia artificial 368
3. Particularidades de la contratación de la inteligencia artificial por parte del Sector Público: riesgos y desafíos 369
3.1. Sesgos algorítmicos y discriminación 369
3.2. Impacto en derechos fundamentales 369
3.3. Protección de datos y seguridad de la información 370
3.4. Interoperabilidad y prevención de dependencia tecnológica 370
4. Marco normativo vigente 370
4. 1. Régimen general sobre contrataciones públicas 370
4.2. Regulación de la inteligencia artificial para el sector público 373
4.2.1. Marco internacional 373
4.2.2. Marco normativo argentino 382
5. La regulación en el derecho comparado 392
6. Buenas prácticas en la contratación pública de inteligencia artificial 394
7. Conclusión 398

USTED APRIETE EL BOTÓN, NOSOTROS HACEMOS EL RESTO. ¿NACEN DERECHOS DE AUTOR SOBRE LAS OBRAS CREADAS MEDIANTE EL USO DE INTELIGENCIA ARTIFICIAL?

Por Ezequiel Valicenti

1. Introducción 402
2. Inteligencia artificial y creatividad. Una agenda de problemas 403
3. La narrativa de la obra protegida y la originalidad en el derecho de autor 404
3.1. La obra 405
3.2. La obra protegida 406
3.2.1. Criterio objetivo: la originalidad como novedad 407
3.2.2. Criterio subjetivo (I): la originalidad como individualidad 408
3.2.3. Criterio subjetivo (II): la originalidad como esfuerzo creativo 409
3.2.4. ¿Unidad o pluralidad de criterios? La acreditación de la originalidad 409
4. Las obras creadas mediante el uso de inteligencia artificial 410
4.1. La autoría de la obra creada por inteligencia artificial 411

Pág.
4.2. La (inexistencia de) originalidad en la obra 411
5. Conclusión: ¿a quién importa la respuesta? 413
Bibliografía 413

Inteligencia artificial y gobierno corporativo: oportunidades, riesgos y límites éticos en la era de los LLM

Por Liuba Lencova Besheva y Nicolás Zaballa

1. Introducción 418
2. El gobierno corporativo: definición, elementos y beneficios 419
3. La irrupción tecnológica y el gobierno corporativo 420
4. Aplicaciones actuales de la IA en el gobierno corporativo 420
5. Riesgos y advertencias 421
6. Propuestas y líneas de acción 423
7. Conclusiones 425
Bibliografía 426

Del operador al supervisor: transformación del trabajo legal ante la irrupción de la inteligencia artificial

Por Bianca Marino Stancheff

1. Introducción: IA en el mundo legal-administrativo 430
2. Automatización de tareas legales: panorama actual 431
3. Del operador al supervisor: redefinición del rol humano 432
4. La paradoja de la automatización: más crítico, menos visible 433
5. Formación, *reskilling* y arquitectura institucional 435
6. Impacto en la identidad profesional y en la estructura del trabajo 437
7. Conclusión: nuevas coordenadas jurídicas para un trabajo con IA 438
Bibliografía 439

Inteligencia artificial y arbitraje internacional: una aproximación regulatoria a propósito de la nueva guía del CIARB

Por Juan Jorge, María Paulina Santacruz y Fernando M. Alemany

1. Introducción 442
2. La Guía sobre IA del CIArb 443
2.1. Beneficios y riesgos del uso de la IA en el arbitraje 445
2.2. Recomendaciones generales sobre el uso de la IA en el arbitraje 446
2.3. Uso de la IA por las partes en el arbitraje 446
2.4. Uso de la IA por los árbitros en el arbitraje 447
3. Reflexiones finales 448
Bibliografía 449

Pág.

CIBERESPACIO EN DISPUTA: INTELIGENCIA ARTIFICIAL Y EL NUEVO CAMPO DE BATALLA JURÍDICO

Por Sol Aylén Borrelli Ozón

1. Introducción 452
2. Del campo de batalla físico al digital: las armas invisibles de ciberdelitos y ciberguerras 453
3. La guerra por los datos: protección, privacidad y poder 454
4. La carrera legal contra los *deepfakes*: un panorama global 456
5. Reflexiones finales 456

Bibliografía 457

IA, NIÑEZ Y ADOLESCENCIA: GARANTIZANDO DERECHOS EN UN NUEVO ENTORNO

Por Mercedes Romera y María Rosa Avila

1. Introducción 462
2. Inteligencia artificial: una nueva racionalidad. ¿Qué medios para qué fines? 462
3. Ciudadanía infantil e inteligencia artificial: una brújula para encontrar nuevas coordenadas 464
4. Derechos de la infancia: usos y desusos de los sistemas de IA 468
5. Zona de promesas y retos: hacia una IA con enfoque de derechos, perspectiva de género y diversidad 471

Bibliografía 472

LAS DECISIONES DE CONSUMO DESAFIADAS (Y POTENCIADAS) POR LA INTELIGENCIA ARTIFICIAL Y EL CONTEXTO DIGITAL

Por Demetrio Alejandro Chamatropulos

1. Introducción 476
2. Antecedentes e hitos (no del todo visibles) en la historia de la inteligencia artificial 476
3. La inteligencia artificial como herramienta para "empeorar" (o "mejorar") las decisiones del consumidor 486
   - 3.1. El "filtro burbuja" y la información recibida por el consumidor. 486
   - 3.2. El menú interminable de opciones y la "paralización" de la decisión del consumidor 487
   - 3.3. La información provista de manera no tradicional por las tecnologías emergentes 488
   - 3.4. Los rankings, las recomendaciones y la memoria episódica 488
   - 3.5. La inteligencia artificial como "sistema de apoyo" de las decisiones del consumidor 489

Pág.
3.6. ¿La inteligencia artificial puede detectar que las decisiones de consumo han sido irreflexivas? ... 490
3.7. La detección y las advertencias del cansancio mediante algoritmos ... 491
3.8. Delegación de decisiones en máquinas y "derecho a equivocarse" ... 491
4. A modo de cierre ... 492

Inteligencia artificial y responsabilidad penal: algunas implicancias desde la teoría del delito

Por Juan Ignacio Diaz, Graciela Elizabeth Gómez y Adriana Belén Pujol

1. Introducción ... 496
2. La inteligencia artificial: concepto, características e impacto en la sociedad inteligente ... 496
3. La irrupción de la inteligencia artificial en el derecho penal ... 497
4. Inteligencia artificial y responsabilidad penal: algunas implicancias desde la teoría del delito ... 497
5. Conclusiones ... 504
Bibliografía final ... 505

Sección IV

Derecho Internacional

Nuevas tecnologías en el paradigma del cuidado social

Por Cristina Monereo Atienza

1. Introducción: las crisis de los cuidados en la era tecnológica ... 510
1.1. La crisis estructural de los cuidados ... 510
1.2. El tecnopoder: las tecnologías no son neutras ... 511
2. El nuevo paradigma de los cuidados y el papel de las tecnologías ... 513
2.1. Breve genealogía del nuevo paradigma de los cuidados ... 513
2.2. Las bases del cuidado social: autonomía relacional, reconocimiento, vulnerabilidad y solidaridad ... 514
2.3. Las nuevas tecnologías para los sujetos de cuidado ... 517
3. A modo de conclusión ... 520
Bibliografía ... 520

The European Union Artificial Intelligence Act and the elevated role of soft law

Por Stanley Greenstein

1. Introduction ... 526
2. The New Legislative Framework ... 527
3. Hard Law versus Soft Law ... 529

Pág.
4. Standardisation 530
4.1 What is a Standard? 531
4.2 Two Types of Standards 532
4.3 How Standards Come About 533
5. The Relationship Between Standards and the Law 535
6. The Artificial Intelligence Act 537
7. Conclusion 540

LA PROTECCIÓN DE LOS DATOS PERSONALES EN EL PROCESO PENAL CHILENO ANTE LA UTILIZACIÓN DE INTELIGENCIA ARTIFICIAL

Por Valeska Fuentealba Sepúlveda

1. Introducción 542
2. La protección de datos personales como derecho a respetar en el proceso penal chileno 542
3. La inteligencia artificial en el proceso penal chileno 544
3.1. La utilización de inteligencia artificial antes de iniciado el proceso 545
3.2. La utilización de inteligencia artificial en el proceso penal 546
4. Conclusiones 547
Bibliografía 548

CÂMERAS CORPORAIS NAS POLÍCIAS BRASILEIRAS: ENTRE TRANSPARÊNCIA E DESIGUALDADE FEDERATIVA

Por Augusto Jobim do Amaral, Eduardo Baldissera Carvalho Salles y Paula Fernanda Failace Antunes de Oliveira

1. Introdução 555
2. Panorama geral de uso e distribuição das câmeras corporais 556
3. Quem acessa as imagens? 557
4. O que realmente está sendo gravado? 558
5. Quanto tempo os dados ficam armazenados? 558
6. Quanto custam as câmeras corporais? 559
7. Onde as câmeras estão sendo usadas? 559
8. Quais os critérios para o uso das câmeras? 560
9. As câmeras são algum resultado? 560
10. Onde os projetos foram descontinuados? 561
11. Quais entes federativos estão em fase de implementação ou expansão? 561
12. Conclusão 562
Referências 563

LA INTELIGENCIA ARTIFICIAL EN EL DERECHO PARAGUAYO: OPORTUNIDADES Y DESAFÍOS

Por Jorge Luis Usher Armoa y Ángela Gisel Gamarra Schulz

1. Introducción 566

Pág.
2. Adopción institucional incipiente de la inteligencia artificial. 567
3. La inteligencia artificial como vehículo de eficiencia y optimización . 568
4. Aspectos críticos de la cautela doctrinal: Ética, Privacidad y Responsabilidad 568
5. La propiedad intelectual como espejo de la personalidad jurídica de la IA 569
6. Desafíos fundamentales de la inteligencia artificial y su impacto en el proceso judicial paraguayo 569
7. Sesgo algorítmico y discriminación socio-judicial 570
8. Brecha digital y acceso universal a la justicia 570
9. Riesgos institucionales: independencia judicial ante la dependencia algorítmica 571
10. Control ético y práctico como respuesta institucional 571
11. Aplicación de la inteligencia artificial por actores de justicia 572
12. Comparativa con países de la región 574
13. Derecho a la explicación regional 575
14. Conclusión 575
Bibliografía 576

O mito de Midas e o devir criativo na era das plataformas digitais. Uma análise do relatório A/80/278 da ONU sobre inteligência artificial e criatividade

Por Marcílio Toscano Franca Filho y Jadgleison Rocha Alves

1. Introdução 581
2. Criatividade como humanidade: os perigos da redução auto-matizada do ser criativo advertidos no relatório 583
3. Criatividade para além da produtividade: o mundo algoriťmico das quantidades e velocidades como um banho ácido para a vida criativa 587
4. Insuficiências normativas no cenário internacional e as in-flexões do sul global: práticas emergentes 591
5. Conclusão: recomendações para uma política de regulação internacional de ia centralizada nos direitos culturais 593
Referências 594

Inteligencia artificial y persecución penal en Chile: análisis y desafíos del proyecto de ley del subsistema de inteligencia económica (boletín N.º 15.975-25)

Por Francisco Narváez Gallo y David Vallespín Pérez

1. Introducción 598
2. El Proyecto de Ley de Chile (15.975-25) por el que se crea el Subsistema de Inteligencia Económica y se establecen medidas de prevención y alerta ante actividades relacionadas con el crimen organizado: planteamiento general 600

Pág.
2.1. Implicaciones jurídicas del uso de la IA en el Subsistema. ¿Qué podemos esperar? 603
2.2. Algunos casos que nos invitan a la reflexión 604
2.3. La preocupación por la opacidad de la IA ante el proceso 607
3. Conclusión 609
Bibliografía 610

LA INTELIGENCIA ARTIFICIAL EN LA RESOLUCIÓN DE CONFLICTOS. UNA MIRADA CRÍTICA ANTE LA SUSTITUCIÓN DEL TERCERO NEUTRAL POR UN SOFTWARE INTELIGENTE

Por Almudena Valiño Ces

1. Introducción 614
2. Los métodos de resolución de conflictos y la inteligencia artificial 615
3. La sustitución del tercero neutral por un sistema de inteligencia artificial en la resolución de conflictos: la función decisoria 616
4. Conclusiones 622
Bibliografía 623

POLÍTICAS DO RECONHECIMENTO FACIAL: REGULAÇÃO ENTRE BRASIL E UNIÃO EUROPEIA

Por Ana Clara Santos Elesbão

1. Introdução 627
2. Análise do regulamento de Inteligência Artificial da União Europeia (*AI Act*) 627
3. Análise do relatório final da Comissão Temporária Interna do Senado Federal Brasileiro sobre Inteligência Artificial no Brasil 631
4. Considerações finais 634
Bibliografia 635

O SISTEMA EUROPEU DE RECONHECIMENTO FACIAL: O DISPOSITIVO BANÓPTICO E A NOVA ROUPAGEM DA ANTROPOMETRIA CRIMINAL

Por Augusto Jobim do Amaral y Gabriel Saad Travassos

1. Introdução 638
2. O sistema europeu de reconhecimento facial 639
3. O dispositivo banóptico e a nova roupagem da antropometria criminal 641
4. Conclusão 645
Referências 646

Pág.

Sección V

Crítica, tecnopolítica, gobernanza

El papel de la crítica en tiempos artificiales

Por Jefferson de Carvalho Gomes

1. Introducción 652
2. La búsqueda por una comprensión sobre lo que es la crítica 653
3. La crítica como el sentido del pensar..... 657
4. La problemática de la posmodernidad: liquidez y simplificación del pensar 659
5. Conclusión 661

Referencias bibliográficas 662

Los chips neuronales y chips neuromórficos. Las lagunas en el derecho y el dilema ético

Por Adriana Tessone

1. ¿Qué son los chips neuronales y para qué se utilizan? 664
2. Otras empresas en el mercado de chips neuromórficos 667
3. Evolución constante e imparable 668
4. El dilema ético de los chips cerebrales 670
5. Los riesgos de la privacidad mental ¿pueden *hackear* un cerebro humano? 671
6. Las lagunas en el derecho en relación con los chips neuronales y neuromórficos 673

La IA y los *cyborgs* del mundo jurídico

Por Santiago Rodríguez Cabezas

1. Introducción 676
2. Una definición de tecnología 677
3. Derecho y tecnología 680
4. Los *Cyborgs* 682
5. ¿Qué hacer con la IA y su creciente vigencia en los campos del derecho? 687

Bibliografía 690

¿Hay derecho y afectos en el Río de la Plata? Un giro para leer la caja negra

Por Diego Martín González Camejo

1. Gobernanza algorítmica: límites del enfoque narrativo en derecho y literatura 692

Pág.
2. Planteo del problema e hipótesis 693
3. El giro afectivo en derecho y literatura 694
4. La teoría de los afectos de Brian Massumi 695
5. Del control de las conciencias a la modulación de cuerpos y afectos ... 697
Conclusiones 698
Referencias 698

NO SON MARGINALES: LOS ERRORES DE LA INTELIGENCIA ARTIFICIAL Y SUS IMPLICANCIAS EN MATERIA DE SEGURIDAD Y SOBERANÍA

Por Juan Acerbi

1. Introducción 702
2. El caso "Bosco": algoritmos, opacidad y justicia social 703
3. La naturaleza de los errores de la IA 704
4. Seguridad, terrorismo y riesgos algorítmicos 705
5. Gobernanza y regulación 706
6. Conclusiones 707
Referencias 708

LEYES, MEDIDAS Y NÚMEROS EN HERÁCLITO Y EN LA REVOLUCIÓN ALGORÍTMICA: MATERIALES PARA UNA COMPARACIÓN DE LARGO ALCANCE

Por Rodrigo Oscar Ottonello

1. Sobre la distancia creciente entre derecho y tecnología 712
2. Ultrahistoria de las relaciones entre número y ley 712
    2.1. El cuidado de las medidas por la justicia 712
    2.2. Legislación y matemática 713
    2.3. *Lógos* (razón/cuenta) del *nómos* único 715
3. Algoritmos y ley 716
Bibliografía 719

DESPLAZAMIENTO DE LA REGLA DE ADJUDICACIÓN HACIA LOS MODELOS DE IA

Por Horacio A. Baca Amenábar

1. Introducción 722
2. La regla de adjudicación de Hart 725
3. La hipótesis del desplazamiento 727
4. Autoridad cognitiva 730
5. Continuidades y rupturas 732
6. La adjudicación como juego lingüístico 734
7. Conclusión: serenidad ante la IA 737
Referencias bibliográficas 739

Pág.

Dialética do automatismo: um ensaio sobre surrealismo e inteligência artificial

Por Guilherme Gonçalves Alcântara y José Flôr de Medeiros Júnior

1. Introdução 743
2. 1924, Relógios Derretidos, Monstros Mitológicos e o Ano 2025. 743
Referências 752

Delírio e pensamento: alucinações em modelos de linguagem generativa

Por Samuel Medeiros Andreatta

1. Distâncias e aproximações 756
2. Certeza e alucinação. 758
3. Alucinações Demasiadamente Humanas..... 760
Referências bibliográficas 764

Hacia una teoría del Estado de la inteligencia artificial: oclocracia algorítmica y tecnofeudalismo

Por Emilio Ferrero García

1. Introducción 768
2. Conceptualización de la oclocracia digital desde la Teoría del Estado 769
3. El algoritmo como punta de lanza contra la democracia representativa 770
4. Tecnopopulismo como nueva forma de cesarismo 772
5. Tecnofeudalismo y oligarquía algorítmica 773
6. Los nuevos señores digitales y su amenaza al Estado de derecho..... 774
7. Conclusiones, visión de futuro y manifiesto urgente en defensa del Estado 775
Referencias bibliográficas..... 777

Reflexões tecnopolíticas a respeito da inteligência e o direito

Por Vinícius Pedro Teló

1. Introdução 782
2. A IA, o direito e o que a vigilância tem a ver com isso 783
3. Discursos, opacidade e riscos..... 788
4. Considerações finais..... 790
Referências 791

Pág.

ACESSO À JUSTIÇA RELOADED: DECODIFICANDO A MATRIX PARA REFUNDAR A JURISDIÇÃO

Por Angela Araujo da Silveira Espindola

1. Considerações iniciais: bem-vindo(a) ao deserto do real..... 795
2. Matrix: a metáfora do sistema jurídico e a decodificação do movimento do acesso à justiça 796
3. O movimento do acesso à justiça e as duas últimas décadas do século XX 798
4. A cibercultura, o acesso à justiça e as duas primeiras décadas do século XXI 800
5. Considerações finais: Reinicializar ou Destruir a Matrix: quem controla quem programa? Quem é o Arquiteto? 804

Referências bibliográficas..... 806

Introducción

Derecho artificial

Por Gonzalo Ana Dobratinich [(*)]

Resumen: El desarrollo de la inteligencia artificial (IA) en el ámbito jurídico plantea desafíos inéditos para la conceptualización del derecho, la normatividad y la justicia. La capacidad de la IA para generar normas, redactar sentencias o analizar consultas jurídicas cuestiona la centralidad del sujeto humano como autor o destinatario de la norma, abriendo la posibilidad de un nuevo individuo legal. Este escenario requiere un análisis crítico que contemple tanto las ventajas tecnológicas, como la pérdida de control humano, la sobrecarga informativa, la marginación de la justicia, los efectos sobre el ejercicio y la enseñanza jurídica y la sostenibilidad ecológica. Se deberá reflexionar sobre la artificialidad del derecho mediado por sistemas automatizados, la necesidad de políticas que fortalezcan el factor humano, la regulación técnica, la construcción de oportunidades, la protección de derechos y la incorporación de múltiples perspectivas. El futuro de una inteligencia artificial jurídica dependerá de nuestra comprensión para equilibrar la innovación tecnológica con la reflexión sobre su uso en los espacios del derecho. En este contexto, presentamos la obra colectiva *Derecho e Inteligencia Artificial*, que reúne estudios de especialistas sobre los desafíos que plantea la incorporación de la IA en el ámbito jurídico.

Palabras clave: Inteligencia Artificial; derecho; justicia.

Abstract: The development of artificial intelligence (AI) in the legal field presents unprecedented challenges for the conceptualization of law, normativity, and justice. AI's ability to generate rules, draft judgments, or analyze legal inquiries challenges the centrality of the human subject as the author or recipient of the norm, opening the possibility of a new legal individual. This scenario requires a critical analysis that considers both technological advantages and the loss of human control, information overload, the marginalization of justice, effects on legal practice and education, and ecological sustainability. Reflection is needed on the artificiality of law

(*) Profesor (UBA - NYU). Investigador posdoctoral CONICET (Facultad de Derecho, UBA). Doctor en Derecho (UBA - UMA, España). Magíster en Filosofía del Derecho (UBA). Master of Fine Arts (MFA) in Creative Writing (NYU, Estados Unidos). Contacto: gonzaloanadobra@gmail.com.

mediated by automated systems, the necessity of policies that strengthen the human factor, technical regulation, the development of opportunities, the protection of rights, and the inclusion of multiple perspectives. The future of legal artificial intelligence will depend on our understanding to balance technological innovation with reflection on its use within legal contexts. In this framework, we present the collective book *Law and Artificial Intelligence*, which brings together studies by specialists addressing the challenges posed by the incorporation of AI into the legal domain.

Key words: *Artificial Intelligence; law; justice.*

"El desocultar imperante en la técnica moderna es un provocar
que pone a la naturaleza en la exigencia de liberar energías,
que en cuanto tales puedan ser explotadas y acumuladas.
Pero, ¿no vale esto también para el viejo molino de viento?"
La pregunta por la técnica (M. Heidegger)

1. INTRODUCCIÓN

Preguntarnos por la capacidad efectiva de un artefacto (*arte-factum*) detenta el interrogante sobre su permanencia. El análisis del tiempo ya es una indagación de procesos sobre los cuales queremos (o deseamos) fundamentar bajo un marco teórico que los establezca o, de manera más apropiada, lo fije, lo enclave. Esta comprensión incluye la incompletitud y, como tal, la inconsistencia de tener una referencia clara. Quizás esté implícito en nuestro proceder que todo uso cuenta con inicio, desarrollo y un final demarcado.

La identificación de la(s) posibilidad(es) adquiere mayor relevancia cuando nos valemos de técnicas que han tenido, a pesar de su gran impacto, poca presencia en nuestra vida diaria. Instrumentos de los cuales nos hemos valido bajo múltiples excusas que consideramos fundamentales, como el costo o el tiempo, y que ya hemos dejado en el olvido (Benjamin, 2019, p. 51). De nuevo la velocidad, la necesidad de una resolución que, lejos de pensarse como una posibilidad meramente reduccionista, se expande en todos los espacios en los que intervenimos.

La expansión espacio-temporal no es una acción propia de un agente externo, sino una actividad dirigida por el sujeto (que definimos como humano). Sin embargo, permite esa externalidad (o extrañamiento), en tanto, una vez introducido el artefacto, podemos, bajo una especie de magia o bien bajo criterios sumariamente delimitados (dependiendo del sistema de pensamiento adoptado), reconocer su propia autonomía.

La administración de los espacios que son nuestros, se delega. Intervenimos sobre el medio para luego dejar que intervenga en múltiples espacios, en grandes cantidades. Así, se coordinan contenedores en un puerto,

se entregan correos a tiempo en cualquier parte del mundo o se organiza la circulación masiva de personas en puntos neurálgicos. Pero también, se administran instrumentos signados por su inmaterialidad (Hui, 2022, p. 143). Largas cifras, conjuntos de ideas y proyectos, información de datos, son manejados por procesadores tecnológicos.

Este avance tecnológico trae pues consigo, entre otras cosas, la obsolescencia. Y en su paso incluye (arrasa), no solo a su aparente cercanía, las cosas, sino también a sus lejanos creadores, las personas. Así como algunos medios de comunicación o transporte sufren alteraciones, haciéndose más precisos (y, en simultáneo, más autónomos), también, como nunca, se han puesto en discusión capacidades que, durante siglos, no desde siempre, se han considerado puras y exclusivamente humanas (asignación dada por el ser humano mismo).

Esto no solo implica una reformulación de la praxis en términos de reasignaciones, sino también un cambio de paradigma en torno a un amplio espectro de investigaciones. Así, el desarrollo de la técnica, en todas sus formas, colores y tamaños, pone bajo la luz saberes que parecían haber encontrado algunos puntos de terreno firme desde los cuales comentar, decir, proponer, criticar y, los más osados o tercos, cerrarse como verdades absolutas (Nietzsche, 1996, p. 25).

Por eso, una de las cuestiones aquí, será tratar de identificar aquello que se ha descartado en esa grandilocuencia reproductiva (Steyerl, 2012, pp. 67-75). Instrumentos ocultos, que, se han puesto (destinado) en otros lugares que nadie pensaba tomar en serio, y que quizá por eso, han alcanzado un alto nivel de desarrollo.

2. Pensamiento seriado

Categorías de estudio o pura voluntad de poder, encorsetan la creatividad de manera desmedida. El vanguardismo antiformalista se permite decir, proyectar, pensar y reemplazar las formas de múltiples maneras. Acompañada por una estética que habilita su despliegue, una realización medida y ordenada, aun cuando se permitan acciones aparentemente arbitrarias (Sadin, 2020, p. 199).

Esta capacidad de disposición no solo es un reconocimiento sino una conquista de la Inteligencia Artificial (en adelante IA). Avanza y con ella (por ella) también las nuevas necesidades, expertas/os, bibliografías y posturas. El imperio intempestivo de la novedad, que, en el deseo del aconteciendo, empieza a cumplir su ciclo de decadencia.

Pero, sin embargo, a diferencia de un reloj a cuerda, una máquina de escribir o un fonógrafo, la IA no admite que la novedad encuentre su límite

(al menos por ahora). No solo por el soporte en el que es capaz de replicarse, su inmaterialidad admite su plasticidad, sino también por formas sobre (desde) las cuales va a asentarse.

Quien entiende un determinado uso necesitará aprender nuevos preceptos para disponer de su funcionamiento. Las áreas disciplinares reconocen su injerencia, pero también empiezan a profundizar en áreas más específicas. Sin lograr el estatuto apocalíptico, cuando, dada su capacidad de análisis, la técnica nos muestra que puede hacerlo por sí misma (Simondon, 2013, p. 60).

Esta profesionalización demanda que estemos al servicio de procesos que realizan otros procesos, más rápidos y, aparentemente, mejores. Entonces, quizás, de un momento a otro, puede liberarse de su lenguaje y convivir con esa otra inteligencia. Habrá que ver, que tiene para ofrecer (y perder) en este intercambio, la máquina, el humano.

3. La muerte del derecho

En su texto de 1991, Owen F. Fiss se interroga sobre la muerte del derecho, al considerar que determinadas concepciones jurídicas permitieron una mirada novedosa en torno al derecho, pero que también distorsionaron sus objetivos, poniendo en peligro su propia existencia (Fiss, 1991, p. 140). Un año más tarde, con una mirada más optimista, reconoce que los peligros de los que hablaba antes parecen menos amenazantes (Fiss, 1992, pp. 233-234).

Treinta y cuatro años después, la discusión que propone Fiss, sobre cuestionar los modos de ver el derecho, no se aleja de lo actualidad con la injerencia y uso de la IA en todo el ámbito del derecho. Se cuestionan los modos en que entendemos el derecho y su articulación con otros espacios. La relación entre la teoría y la práctica, como tensiones permanentes, en el despliegue del discurso jurídico.

Los avances de la IA llevan consigo el progreso de los miedos, las preocupaciones y la incertidumbre. La reconfiguración del universo que conocemos. Las grandes fábricas, de aceitados engranajes, que en otros momentos producían cosas capaces de llegar a igualar el mundo globalizado, se reemplazan por pequeños cubículos donde, en el silencio de los días, se programan los nuevos intercambios. Emergen múltiples opciones que rompen con la propuesta de un progreso lineal y acabado, medido por la fuerza de una red neuronal, que refleja (reemplaza) el pensamiento (Bratton, 2016, p. 352).

En este espacio de reconocimientos, críticas, acuerdos y disidencias, también participa el derecho, el cual debe adecuarse a las demandas, repartos e intereses que las sociedades actuales reclaman. La participación de una inteligencia (IA) con una capacidad desmedida (ilimitada), le instala una presencia que le ofrece estadísticas, datos, organización, resolución,

velocidad y un aparente desinterés e indiferencia subjetiva (Noble, 2018, p. 117). Esta relación no se limita a un ámbito específico. Bien podría pensarse que la discusión gira en torno a la aplicación de la IA en los procesos decisorios judiciales; sin embargo, han surgido nuevas formas de pensar estas relaciones: redacción de escritos, creación de normas, asesoramiento legal, detección de fraudes y riesgos jurídicos, enseñanza del derecho, investigación y análisis jurídico (Ana Dobratinich, 2021, pp. 1-7).

Aquella esperanza, o resguardo distópico, que solo se ponía sobre algunos saberes, se expande y aplica en el derecho. Por lo tanto, será importante discutir la forma en que se llevan a cabo los procesos algorítmicos, la disposición de los instrumentos necesarios, el acceso a esas herramientas por parte de los individuos y la construcción de parámetros que justifiquen los funcionamientos tecnológicos.

4. Los dones robados-perdidos

Las inteligencias robadas por aquello (la cosa) que piensa más rápido, de manera anticipada (*mito prometeico*); dotado de una gran cantidad de dones (*mito pandórico*), se abre, dispone y distribuye. El objeto que no teme el castigo porque no tiene ninguna referencia para soportar ningún tipo de justicia. Solo produce, de manera aditiva y así sobrelleva el progreso del trabajo, de los días. El exceso del dato que satura y no informa (Han, 2017, p. 88). Ya no significa, *a contrario sensu*, una disposición precisa, sino una verborragia informática caótica: una sobrecarga. Es decir, la dadiva grotesca que agota los sistemas, entre ellos, el humano.

Algunas posiciones reclamarán la necesidad de establecer una red de contención que explique y dictamine sus vías de tránsito, mientras que otras miradas, defenderán que la IA recorra sus propios caminos (Bostrom, 2014, p. 45). De pequeñas operaciones hasta la idea de un sistema normativo generado por IA, capaz de funcionar sin un sujeto humano claramente definido (Gunkel, 2017, p. 180). El cuestionamiento del binomio individuo-máquina, abre la posibilidad de un nuevo sujeto legal que redacta leyes, genera y produce artículos, escribe sentencias y resuelve consultas jurídicas.

5. ¿Qué derecho nos ha quedado?

Frente a los grandes ofrecimientos y propuestas de la IA en el ámbito del derecho, la contrapartida es la exigencia de un análisis crítico que considere las experiencias y subjetividades que los datos no pueden asimilar. Una evaluación detallada de los efectos tecnológicos sobre la salud mental, la hipertrofia de expectativas, la conectividad virtual y las formas de enseñanza, la incidencia ecológica por el excesivo consumo energético (Crawford, 2022, pp. 56-71).

La idea de que derecho pueda permanecer al margen de los sentidos éticos/estéticos que los avances digitales proponen, obliga a pensar sobre su (in)capacidad reguladora para resolver estas transformaciones (Chesterman, 2020, p. 822-824). El futuro de la inteligencia jurídica quizás dependa de nuestras capacidades para clarificar el pensamiento y de analizar la relación entre descripciones y valoraciones. Las ficciones del hoy pueden ser la ciencia del futuro. Imaginemos una IA más rápida que la reflexión jurídica, aprendiendo principios de equidad, ponderación y prevención del abuso del derecho; ¿seguiría existiendo la justicia como espacio exclusivamente humano?

El derecho, cuando evocamos su existencia cultural y su lejanía de toda entidad inhumana, revela su propia artificialidad. Si la ley representa la marginalización de la justicia, una extrañeza que padecemos, debemos cuestionar la artificialidad misma del derecho. El reduccionismo metodológico que antes preocupaba en objetos de carácter performático hoy se traslada a una naturaleza tecno-artificial: un derecho mediado por sistemas que funcionan, producen y disponen, pero que no preguntan, interrogan o reconocen la diferencia individual.

Surgen así, entre otras, propuestas de automatización deontológica: ¿cómo debe funcionar la IA? Consignas que se establecen como dogmas técnicos. Tecnólogos consideran que un código y un consenso formal puedan garantizar decisiones morales correctas, reduciendo la ética a algoritmos y la deontología a simples instrucciones. (Vallor, 2016, pp. 271). El espacio dialógico, plural y conflictivo por excelencia de la deliberación ética, se convierte en una lámina clara sin fisuras ni negatividades: la IA ofrece, produce, dispone e iguala todo en la pureza del dato, en la positividad del cálculo (Stiegler, 2017, pp. 113-143).

Será, pues, necesario analizar las propuestas, las respuestas y las ideas que promuevan políticas capaces de: fortalecer la alfabetización digital, mejorar la infraestructura tecnológica del sistema judicial, garantizar conectividad y equipamiento, desarrollar un marco legal integral, proteger los derechos digitales y la transparencia algorítmica e impulsar investigaciones interdisciplinarias. Un pensamiento crítico, que reconozca puntos sobre los cuales dialogar y construir un espacio que defina, la siempre controvertida y al mismo tiempo pretendida, idea de justicia.

6. Pensamientos, secciones y obra

Las propuestas llegan de manera múltiple. Distintas voces convergen en la obra colectiva *Derecho e Inteligencia Artificial*. El libro se presenta como un espacio de reflexión para compartir, disentir o crear. Participando desde las más diversas áreas, con miradas cercanas o antagónicas, los textos procuran abrir nuevas perspectivas, fomentar el análisis crítico y contribuir al pensamiento sobre los vínculos entre la IA y el derecho.

Así, en la primera sección se analizan las intersecciones entre la tecnología artificial y espacios jurídicos desde diferentes enfoques filosóficos. Las investigaciones interrogan en torno a la concepción misma de inteligencia y su posible imitación o superación por las máquinas, explorando el contacto entre lo humano y lo artificial, entre el sentir y el simular. Se analizan los dilemas que emergen cuando la toma de decisiones automatizada desafía principios como la autonomía, la responsabilidad o el cuidado, así como la necesidad de repensar la función del derecho en un escenario donde la racionalidad técnica tiende a reemplazar la deliberación humana. Se pone en cuestión si la IA puede constituirse en un nuevo arquetipo del pensamiento, una suerte de retorno a una idea divina del derecho, o si, por el contrario, exige establecer límites éticos que contengan su expansión. En este espacio escriben Horacio Roberto Granero, Isolina Dabove, Lenio Streck, Silvia Adriana Ramos, Gonzalo Aguirre, Fishel Szlajen, Nythamar de Oliveira, Eduardo Ramalho Rabenhorst, Fernando Beresñak, Eduardo Magoja, Luciano Laise, Fernando Llano Alonso, Andre Karam y Amanda Antonelo.

En la segunda sección, Antonio Martino, Helga Lell, Débora Mahaba, Nicolás Pérez Felicioni, Jorge A. Cerdio Herrán, Lara Morquecho, Guadalupe Juárez, Victoria Gisvert, Marcela Tovar, Diego Robledo, Rosa Ávila Paz, Sonia Gamboa, Andrés Botero y Kevin Valdivieso indagan cómo la IA incide y transforma tanto la profesión como la educación jurídica. Se analizan los nuevos roles que surgen en la construcción de *datasets* jurisprudenciales y la aplicación de la IA en casos judiciales y estrategias probatorias. Asimismo, se estudia el potencial de los laboratorios virtuales en la enseñanza del derecho, proponiendo una transformación pedagógica. También se investigan los desafíos que enfrentan los agentes del derecho en el uso de estas herramientas y la necesidad de armonizar su incorporación en los procesos judiciales mediante normas y guías claras, en miras a la administración y fortalecimiento de la justicia.

En la tercera sección las producciones de Manuela Adrogué, Milagros Ibarzábal, Ezequiel Valicenti, Nicolás Zavalla, Liuba Lencova Besheva, Bianca Marino, Juan Jorge, María Paulina Santacruz Salazar, Fernando M. Alemany, Sol Borrelli Ozón, María Rosa Ávila, Mercedes Romera, Alejandro Chamatropulos, Juan Ignacio Díaz, Graciela E. Gómez y Adriana Belén Pujol investigan sobre los desafíos legales y regulatorios que surgen con la irrupción de la IA en diversos ámbitos, desde la contratación pública hasta la propiedad intelectual y el gobierno corporativo. Se propone un análisis de los límites éticos y jurídicos en la supervisión de sistemas automatizados, la transformación del trabajo legal y la aplicación de la IA en arbitraje internacional. Además, se reflexiona sobre los derechos de niñas, niños y adolescentes en entornos digitales, la responsabilidad penal y los desafíos que la IA plantea sobre cuestiones vinculadas al derecho del consumidor.

La cuarta sección trabaja en la interacción entre IA, derecho internacional y políticas regulatorias en distintos contextos nacionales y supranacionales. Se exponen normas y proyectos legales de países como Brasil, Paraguay y Chile, así como marcos regulatorios de Europa, que incluyen el papel del *soft law* en el *Acta de Inteligencia Artificial de la Unión Europea*. Los textos examinan la protección de datos personales, la persecución penal, el reconocimiento facial y la transparencia institucional, así como los desafíos que surgen ante la sustitución de mediadores humanos por sistemas inteligentes en la resolución de conflictos. Dialogarán sobre estos tópicos los artículos de Cristina Monereo Atienza, Stanley Greenstein, Valeska Fuentealba Sepúlveda, Augusto Jobim do Amaral, Eduardo Baldissera Carvalho Salles, Paula Fernanda Failace Antunes de Oliveira, Jorge Usher, Ángela Gisel Gamarra Schulz, Marcílio Franca, Jadgleison Rocha Alves, Francisco Narváez, David Vallespín, Valiño Ces Almudena, Ana Clara Santos Elesbão, Augusto Jobim do Amaral y Gabriel Saad Travassos.

Finalmente, en la quinta sección, escriben Jefferson de Carvalho Gomes, Adriana Tessone, Santiago Rodríguez Cabezas, Diego Martín Gonzalez Camejo, Juan Acerbi, Rodrigo Oscar Ottonello, Horacio Baca Amenábar, Medeiros Junior, Ghuillerme Alcantara, Samuel Medeiros Andreatta, Emilio Ferrero García, Vinícius Pedro Teló y Angela Araujo da Silveira Espindola. Los artículos abordan la relación entre tecnología, derecho y poder. Se analizan los desafíos que plantean los sistemas de IA y los chips neuronales en la toma de decisiones jurídicas, la seguridad, la soberanía y la gestión del Estado. También consideran la dimensión cultural y filosófica de la IA, incluyendo el surrealismo, los modelos de lenguaje generativo y algoritmos automatizados, así como las nuevas formas de gobernanza y los avances tecnológicos que afectan la justicia, las estructuras normativas y las relaciones sociales.

7. CONCLUSIONES

La presente obra intenta ofrecer una lectura que acompañe los numerosos estudios en torno sobre *Derecho e Inteligencia Artificial*. Se propone como un aporte profundo y reflexivo sobre una relación que se encuentra en continuo crecimiento.

Por ello, queremos agradecer a las/os autoras/es, expertas/os y miembros de instituciones, nacionales e internacionales, que han permitido el desarrollo y exposición de las ideas aquí plasmadas.

Asimismo, agradecemos al Departamento de Publicaciones de la Facultad de Derecho de la Universidad de Buenos Aires, que hizo posible la publicación de esta obra. Especialmente, a su directora Mary Beloff, a su subdirector Luis R. J. Sáenz, a su secretaria Mayra Danae Salvatierra y a todo el

personal, quienes, con su tiempo y disposición, nos acompañaron a lo largo de todo el proceso.

Finalmente, agradecer a quienes, explícita o implícitamente, con sus palabras y silencios, participaron y nos permitieron ampliar las fronteras del conocimiento en torno a los estudios sobre *Derecho e Inteligencia Artificial*.

Consideramos que el intercambio entre saberes propone encuentros y disidencias, necesarios para pensar las relaciones entre el ámbito jurídico y la tecnología. No se agota la totalidad de los estudios, surgirán nuevas ideas con otras perspectivas y enfoques. Allí reside lo valioso: en la sustentación de un pensamiento crítico, capaz no de imponerse, sino de invitar al diálogo con las múltiples formas de inteligencia con las que habitamos. Proponemos, por tanto, el tiempo de ocio, los vínculos intersubjetivos, la pausa reflexiva y el conocimiento lúdico como instancias para el pensamiento y la lectura de estas líneas, que, quién sabe, cómo serán leídas en un futuro que, como tal, siempre permanece abierto.

Bibliografía

Ana Dobratinich, G. (2021). *Derecho y nuevas tecnologías*. Departamento de Publicaciones de la Facultad de Derecho, Universidad de Buenos Aires.

Benjamin, W. (2019). *La obra de arte en la época de su reproductibilidad técnica*. Ediciones Godot.

Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.

Bratton, B. (2016). *The stack: On software and sovereignty*. Editorial MIT Press.

Carr, N. (2011). *The shallows: What the Internet is doing to our brains*. W. W. Norton & Company.

Chesterman, S. (2020). Artificial intelligence and the limits of legal personality. *International and Comparative Law Quarterly, 69*(4), 819–844. https://doi.org/10.1017/S0020589320000366

Crawford, K. (2022). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Fiss, O. M. (1991). ¿La muerte del Derecho?. *DOXA. Cuadernos de Filosofía del Derecho,* (10), 123–140. https://doi.org/10.14198/DOXA1991.10.05

Fiss, O. M. (1992). El Derecho recuperado. *DOXA. Cuadernos de Filosofía del Derecho*, (11), 233–246. https://doi.org/10.14198/DOXA1992.11.09

Gunkel, D. J. (2017). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.

Han, B. C. (2017). *La agonía del Eros*. Herder.

Heidegger, M. (2021). *La pregunta por la técnica*. Herder Editorial.

Hui, Y. (2022). *Recursivdad y contingencia*. Caja Negra.

Nietzsche, F. (1996). *Sobre verdad y mentira en sentido extramoral*.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*.New York University Press.

Sadin, É. (2020). *La inteligencia artificial o el desafío del siglo*. Caja Negra.

Simondon, G. (2013). *El modo de existencia de los objetos técnicos*. Prometeo Libros.

Steyerl, H. (2012). *The wretched of the screen*. Sternberg Press.

Stiegler, B. (2017). *The automatic society*. Polity Press.

Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.

Sección I

Filosofía, Justicia, Humano/Máquina

Las alucinaciones en la inteligencia artificial, su utilidad en la investigación: desafíos y oportunidades

Por Horacio Roberto Granero (*)

Resumen: Las "alucinaciones" de los grandes modelos de lenguaje —respuestas coherentes pero sin base factual— han pasado de ser consideradas simples errores a ocupar el centro del debate sobre creatividad, seguridad y conocimiento en la era de la inteligencia artificial. El documento parte de la paradoja: lo que erosiona la confianza en medicina o finanzas puede, en otros contextos, disparar hipótesis científicas inéditas o diseños tecnológicos imposibles de imaginar por métodos convencionales. A través de numerosos ejemplos —desde el rediseño de proteínas premiado con Nobel hasta catéteres antiinfecciosos o nuevas estructuras cristalinas— se muestra cómo salidas "inventadas" por la IA, sometidas luego a validación humana o computacional, han acelerado ciclos de descubrimiento que antes demandaban años. El texto explica que las alucinaciones surgen de la arquitectura estadística de los modelos: cuando la consulta cae fuera del espacio de entrenamiento, el sistema prefiere completar con lo "plausible" antes que admitir ignorancia. Esta tendencia se ve amplificada por datos contaminados y por la presión comercial de ofrecer respuestas fluidas. Frente a ello, se proponen tres líneas de acción: (1) técnicas de mitigación como generación con recuperación externa (RAG), modelos explicables y aprendizaje por refuerzo con retroalimentación humana; (2) marcos éticos y legales que responsabilicen a desarrolladores y usuarios; (3) alfabetización crítica que prepare a médicos, ingenieros y ciudadanos a detectar y contrastar información. El núcleo conceptual es la "serendipia planificada": diseñar flujos de trabajo donde la aleatoriedad controlada de la IA incremente la pro-

---

(*) Afiliación Institucional del autor: profesor emérito de la Facultad de Derecho de la Universidad Católica Argentina. Correo electrónico de contacto: hrg@graneroabogados.com.ar Reseña curricular: abogado (1969), doctor en Ciencias (2001). Profesor universitario desde 1970, Emérito en 2018. Fue fundador y director del Departamento de Posgrado y fundador en 1997 de la Carrera de Posgrado de Abogado Especializado en Derecho de la Alta Tecnología (UCA). En 1987 fundó elDial.com y es su actual presidente, integra el Instituto de Derecho e Informática del Colegio Público de Abogados. Profesor invitado de diversas universidades de la República Argentina y del exterior habiendo disertado en seminarios y publicado trabajos de su especialidad. Autor de varios libros de su especialidad.

babilidad de hallazgos valiosos sin comprometer la veracidad. Esto exige alternar modos de exploración creativa y modos de precisión, siempre con supervisión humana. La conclusión es que la IA no debe idealizarse ni demonizarse; su integración exitosa depende de una gobernanza adaptativa que combine ingeniería, pedagogía y política para convertir la alucinación, no en un riesgo inevitable, sino en un recurso escalable de innovación.

Palabras clave: inteligencia artificial, alucinaciones, serendipia, investigación

Abstract: Hallucinations in large language models—coherent yet fact-free outputs—have moved from being viewed as mere errors to becoming central to debates on creativity, safety, and knowledge in the age of artificial intelligence. The article begins with the paradox: while undermining trust in medicine or finance, these fabrications can also spark unprecedented scientific hypotheses or technological designs impossible to conceive via conventional methods. Through numerous examples—from Nobel-prize-winning protein redesign to anti-infective catheters and novel crystal structures—it shows how AI-invented outputs, once filtered by human or computational validation, have compressed discovery cycles from years to days. The text explains that hallucinations stem from the statistical architecture of models: when queries fall outside the training space, the system prefers to complete with the "plausible" rather than admit ignorance. This tendency is amplified by contaminated data and commercial pressure for fluent answers. Three lines of action are proposed: (1) mitigation techniques such as retrieval-augmented generation (RAG), explainable models, and reinforcement learning from human feedback; (2) ethical and legal frameworks that hold developers and users accountable; (3) critical literacy that equips doctors, engineers, and citizens to detect and cross-check information. The core concept is "planned serendipity": designing workflows where AI-controlled randomness increases the odds of valuable findings without compromising truth. This requires toggling between creative-exploration and precision modes, always under human oversight. The conclusion is that AI should neither be idealized nor demonized; its successful integration depends on adaptive governance that blends engineering, pedagogy, and policy to turn hallucination from an unavoidable risk into a scalable innovation resource.

Keywords: artificial intelligence, hallucinations, serendipity, research.

1. Introducción

Las asombrosas habilidades de los Grandes Modelos de Lenguaje (LLM) se presentan a veces como errores a los ojos de los humanos. Este fenómeno, etiquetado como "alucinaciones", puede que no siempre sean meros fallos, sino más bien atisbos de una nueva forma de creatividad digital. Estas des-

viaciones inesperadas, tradicionalmente vistas como errores, podrían ser de hecho la forma en que la IA "piensa fuera del chip", empujando los límites de los procesos de pensamiento computacional convencionales, pero ¿qué significa poner límites?

En los términos más simples, las alucinaciones en los LLM pueden provenir de fenómenos como el sobreajuste, donde el modelo, demasiado sintonizado con sus datos de entrenamiento, lucha por generar resultados novedosos pero precisos. Este sobreajuste, de manera bastante contradictoria, puede conducir a la generación de salidas completamente originales, aunque sin sentido. Este fenómeno refleja intrigantemente el proceso creativo en los seres humanos, donde la verdadera originalidad a menudo emerge del borde del caos y el orden, de la interacción del conocimiento profundo y la capacidad de trascenderlo.

Además, las complejidades de la contaminación de datos en los conjuntos de entrenamiento de LLM sirven como un terreno fértil para el fenómeno de las alucinaciones. Esta contaminación, a menudo un subproducto de las vastas y variadas fuentes de las que se agregan los datos, puede introducir una gran cantidad de imprecisiones y sesgos en el modelo. Cuando los LLM se entrenan con estos conjuntos de datos comprometidos, aprenden inadvertidamente a replicar estos errores, lo que lleva a resultados que no solo son incorrectos, sino que a veces están completamente alejados de la realidad. Este problema se ve agravado por la incapacidad de los modelos para evaluar críticamente la información que procesan, convirtiéndolos en conductos involuntarios para la propagación de estas imprecisiones. En consecuencia, el desafío no solo radica en refinar los modelos, sino también en desinfectar los datos que consumen, asegurando que sean lo más limpios, diversos y representativos posible.

Abordar este aspecto de la calidad de los datos es crucial para mitigar las alucinaciones, mejorando así la fiabilidad y la aplicabilidad de los resultados de LLM en escenarios del mundo real. Sin embargo, esta perspectiva puede invitar a un replanteamiento radical de nuestro compromiso con la IA. En lugar de descartar apresuradamente estas "alucinaciones" como meros errores de datos o codificación descuidada que deben erradicarse, podríamos considerarlas, al menos en ocasiones, una forma de inventiva digital que debe entenderse y aprovecharse. Al profundizar en los mecanismos y las implicaciones de estas aberraciones, podríamos descubrir nuevos caminos para la creatividad, amalgamando lo mejor del intelecto humano y artificial.

Esta capacidad sugiere un área curiosa para la consideración, donde las "alucinaciones" inducidas por la IA no solo podrían emular la creatividad humana, sino extenderla, ofreciendo perspectivas novedosas que fusionan elementos visuales, textuales y auditivos de formas previamente inconcebibles. Esta convergencia podría anunciar una nueva era de expresión creativa

y resolución de problemas, impulsando los dominios del arte, el diseño y la tecnología a nuevos reinos donde la fusión de diferentes modalidades sensoriales da lugar a paradigmas creativos completamente nuevos.

Aceptar la complejidad y la imprevisibilidad de los resultados de LLM requiere una perspectiva racional e imaginativa que pueda allanar el camino para imaginar un futuro en el que el papel de la IA trascienda la destreza computacional y se caracterice por su capacidad de obtener ideas "creativas" y fortuitas. Este viaje a las posibilidades de las alucinaciones de la IA requiere una evaluación fundamentada del potencial de las máquinas, fomentando una relación simbiótica en la que las inteligencias humana y artificial amplíen colectivamente los límites de la innovación y el descubrimiento.

2. Qué significa que la IA pueda alucinar?

El concepto de alucinación, cuando se aplica a la IA, puede generar confusión si no se distingue de otros términos como "error", "fabricación" o "falsedad deliberada". A grandes rasgos, una alucinación ocurre cuando el modelo produce contenido sin base real, ya sea debido a lagunas de datos, inferencias erróneas o la incapacidad de reconocer la frontera entre lo que está presente en su entrenamiento y lo que excede sus capacidades.

Los LLM están entrenados con volúmenes masivos de texto. Sus respuestas se basan en patrones estadísticos de probabilidad y correlaciones entre palabras y oraciones. Cuando se formulan preguntas muy específicas o fuera del rango de entrenamiento, el modelo no puede recuperar la información correcta. En lugar de "admitir" su ignorancia, la arquitectura subyacente trata de componer una respuesta plausible que mantenga la coherencia gramatical y lingüística, aunque sea ficticia. Históricamente, muchos LLM se diseñan para generar respuestas "naturales" y fluidas. Desde un punto de vista puramente probabilístico, es más "natural" completar la frase con algo que "suene" verosímil que quedarse en silencio o responder "No lo sé". Este comportamiento es útil en tareas como la autocompletación de textos, pero se convierte en un problema cuando se asume que la IA brinda información factual.

Esta tendencia a "inventar" detalles crea un riesgo de desinformación, especialmente si los usuarios no son conscientes de que la IA puede incurrir en este tipo de errores. Sin embargo, la dimensión creativa de estas alucinaciones también abre puertas en contextos como la generación de hipótesis científicas, la exploración de estructuras químicas o el análisis literario.

La mayoría de las personas simplifica el funcionamiento de la IA en términos de acierto o error. Sin embargo, la realidad es que las salidas de estos modelos pueden ser mixtas: acertar en cierta parte de la respuesta y "soñar" otra. Cuando esto ocurre en gran escala, las consecuencias son mucho más

complejas. Inicialmente, las alucinaciones se veían como una deficiencia técnica que había que erradicar, similar a un bug en un *software*.

En particular, grandes compañías tecnológicas han invertido en sistemas de verificación y en la integración de bases de datos externas para reducir estos "errores". Varios equipos de investigación han adoptado un enfoque distinto, preguntándose qué pasaría si esas salidas ficticias ofrecieran "pistas" para nuevas ideas. Un ejemplo concreto proviene del ámbito del diseño de moléculas farmacológicas, donde se ha permitido a la IA generar miles de estructuras hipotéticas, algunas sin aparente correlación con la realidad, pero que luego se someten a verificación computacional adicional y, en casos excepcionales, han dado resultados prometedores.

Las ciencias experimentales nos enseñan que muchos descubrimientos surgieron de "accidentes" o sucesos fortuitos. Siguiendo esta lógica, las alucinaciones podrían verse como un "accidente controlado" que permite a la IA "pensar fuera de la caja" y sugerir líneas de investigación poco convencionales. Lejos de ser un mero fallo, algunas alucinaciones abren la puerta a la diversidad de enfoques y, por tanto, al avance de la ciencia.

En ámbitos sensibles, un error puede significar la diferencia entre la vida y la muerte o entre la solvencia y la quiebra. Por ejemplo, en una asesoría financiera, un modelo podría "alucinar" indicadores económicos inexistentes, provocando inversiones erróneas. Resulta crucial establecer estándares de *accountability* para los desarrolladores y los proveedores de servicios de IA.

Si el usuario ignora que la IA es capaz de inventar información, puede depositar una confianza excesiva en sus respuestas. Esto es más grave si el usuario no tiene un conocimiento previo que le permita distinguir lo real de lo inventado. La transparencia requiere que las plataformas avisen sobre la posibilidad de errores y ofrezcan mecanismos sencillos de verificación de hechos.

Otra dimensión ética aparece cuando se promueve conscientemente la creatividad de la IA en contextos de innovación. ¿Dónde se traza la línea entre fomentar la "imaginación" del modelo y difundir sin control datos ficticios que podrían confundir a la comunidad científica? La coexistencia de herramientas de validación y entornos seguros de experimentación se vuelve esencial para que esta práctica sea viable sin dañar la integridad de la información pública.

3. La serendipia: ventajas potenciales para la exploración científica

La ciencia está llena de casos de "descubrimientos accidentales" que marcaron hitos revolucionarios, desde la penicilina hasta la vulcanización del caucho. Las alucinaciones podrían potenciar esta serendipia, generando

hipótesis aparentemente descabelladas que, sometidas a verificación, podrían dar lugar a nuevos avances (Fleming, 1929).

La doctrina acuñó este nuevo concepto, "la serendipia", denominando así al fenómeno de hacer descubrimientos inesperados y valiosos por casualidad, ha impulsado históricamente avances en ciencia, tecnología y cultura. Sin embargo, la integración de la inteligencia artificial en los procesos de investigación plantea la pregunta: *¿Puede la serendipia planificarse o amplificarse sistemáticamente?*

El concepto de "serendipia planificada" es un paradigma que combina metodologías estructuradas con herramientas basadas en IA para crear entornos donde los hallazgos fortuitos sean más probables. Analizamos casos de estudio de investigación interdisciplinaria, proponemos un marco teórico para integrar mecanismos que potencien la serendipia en sistemas de IA y discutimos implicaciones éticas y prácticas. Al unir creatividad estocástica y predictibilidad computacional, argumentamos que la IA puede servir como catalizador para una innovación intencional pero impredecible.

La serendipia suele idealizarse como un proceso meramente accidental, pero su papel en descubrimientos científicos—desde la penicilina hasta el grafeno—es innegable. En la era del *big data* y la IA, los investigadores enfrentan una paradoja: aunque las herramientas computacionales destacan en reconocimiento de patrones y optimización, podrían limitar la creatividad al sobreingeniar resultados.

Los sistemas de IA pueden diseñarse para *simular* o *inducir* descubrimientos serendípios mediante la introducción intencional de aleatoriedad controlada, síntesis de datos interdisciplinarios y algoritmos exploratorios. Se podrían entonces, definir a la "serendipia planificada" como la integración sistemática de procesos estocásticos, divergentes o asociativos en flujos de trabajo de investigación para aumentar la probabilidad de resultados inesperados pero significativos e involucra tres elementos clave:

1. Sagacidad: la capacidad del investigador para reconocer anomalías.

2. Causalidad: interacciones u observaciones no planificadas.

3. Valor: la utilidad del descubrimiento.

Las metodologías tradicionales priorizan enfoques basados en hipótesis, marginando la curiosidad exploratoria. No obstante, la capacidad de la IA para procesar grandes volúmenes de datos e identificar correlaciones no evidentes ofrece un camino para institucionalizar la serendipia.

Los sistemas de IA pueden potenciar la serendipia mediante:

-Algoritmos de búsqueda divergente: técnicas como algoritmos genéticos o aprendizaje por refuerzo con equilibro exploración-explotación.

-*Grafos de conocimiento transdisciplinarios:* conexión de bases de datos dispares (p. ej., biomedicina y ciencia de materiales) para revelar relaciones ocultas.

-*Modelos generativos*: herramientas como GPT-4 o modelos de difusión que proponen hipótesis o diseños novedosos, evitando sesgos cognitivos humanos.

Entre los casos de estudio en *serendipia potenciada por IA* se reconocen a los siguientes:

a) Reposicionamiento de fármacos mediante análisis de redes. El análisis basado en IA de literatura biomédica y redes de interacción proteica ha identificado usos terapéuticos inesperados de fármacos existentes. Por ejemplo, el descubrimiento por IBM Watson del *baricitinib* como candidato para tratar COVID-19 muestra cómo la IA puede reinterpretar conocimiento previo.

b) Descubrimiento de materiales con redes generativas antagónicas (GANs). Investigadores del MIT usaron GANs para proponer nuevas estructuras cristalinas para baterías, identificando serendípicamente un conductor de iones de litio con estabilidad sin precedentes.

c) Inspiración interdisciplinaria en colaboraciones arte-ciencia. Herramientas como *DeepDream* se han reutilizado para visualizar la toma de decisiones en redes neuronales, inspirando nuevas teorías en neurociencia sobre el procesamiento visual humano

La serendipia planificada no busca eliminar el azar, sino crear condiciones donde los encuentros fortuitos tengan mayor probabilidad de generar valor. El papel de la IA es análogo a un "asistente de laboratorio digital" que resalta anomalías, propone hipótesis contraintuitivas y conecta conocimientos fragmentados. Sin embargo, esto exige replantear incentivos en investigación para recompensar el trabajo exploratorio y la colaboración interdisciplinaria. La unión entre IA y serendipia representa un cambio de paradigma en metodologías de investigación.

Al diseñar sistemas que combinen rigor algorítmico con creatividad estocástica, podemos transformar la serendipia de un accidente afortunado en un recurso escalable y así futuros trabajos deberían enfocarse en cuantificar su impacto, refinar pautas éticas y fomentar sinergias humano-IA que preserven el "factor humano" mientras aprovechan la novedad generada por máquinas.

En ingeniería y diseño de productos, las alucinaciones pueden ofrecer propuestas de geometrías o mecanismos no contemplados en los manuales tradicionales. Y, al ser analizados por expertos, ciertos conceptos pueden

pulirse hasta volverse funcionales. Por ejemplo, el modelado de estructuras moleculares en la química computacional se beneficia de la "imaginación" de la IA para sugerir arreglos atómicos no explorados antes.

Para algunos, las alucinaciones de IA están revitalizando el lado creativo de la ciencia y muchos indican que aceleran el proceso por el cual los científicos e inventores sueñan con nuevas ideas y las prueban para ver si la realidad concuerda, llegando a decir que se trata del método científico tradicional, solo que sobrealimentado gracias a la IA y lo que antes tardaba años, ahora se puede hacer en días, horas y minutos. En algunos casos, los ciclos acelerados de investigación ayudan a los científicos a abrir nuevas fronteras dado que las alucinaciones han permitido que los científicos enseñen a modelos informáticos generativos sobre un tema en particular y luego dejen que las máquinas reelaboren esa información.

Los resultados pueden variar desde sutiles y equivocados hasta surrealistas pero, a veces, conducen a grandes descubrimientos de diverso tipo, como los siguientes:

- En octubre de 2024, David Baker, de la Universidad de Washington, compartió el Premio Nobel de Química por su investigación pionera sobre las proteínas, las moléculas nudosas que dan poder a la vida. El comité del Nobel lo elogió por descubrir cómo construir rápidamente tipos completamente nuevos de proteínas que no se encuentran en la naturaleza, calificando su hazaña de "casi imposible" si bien no dijo nada sobre las alucinaciones de la IA en una revisión efectuada del trabajo del Dr. Baker y en un comunicado de prensa, simplemente le dio crédito a su equipo por producir "una creación de proteína imaginativa tras otra". En una entrevista antes del anuncio del premio, el Dr. Baker había indicado que había hecho uso de "ráfagas de imaginaciones" de IA como fundamentales para "hacer proteínas desde cero". La nueva tecnología, agregó, ha ayudado a su laboratorio a obtener aproximadamente 100 patentes, muchas de ellas para atención médica. Una de ellas es la de una nueva forma de tratar el cáncer y otra busca ayudar en la guerra global contra las infecciones virales.

- Por su parte, Anima Anandkumar, profesora de matemáticas y ciencias de la computación en el Instituto de Tecnología de California, quien anteriormente dirigió la investigación de inteligencia artificial en Nvidia, el principal fabricante de chips de inteligencia artificial, y sus colegas utilizaron alucinaciones de IA para ayudar a diseñar un nuevo tipo de catéter que reduce en gran medida la contaminación bacteriana, una pesadilla global que anualmente causa millones de infecciones del tracto urinario. Dijo que el modelo de IA del equipo ideó miles de geometrías de catéteres y luego eligió uno que era el más efectivo.

- El plegamiento proteico es un problema central en biología molecular. La predicción de la estructura 3D de las proteínas a partir de su secuencia de aminoácidos, durante décadas, fue uno de los grandes desafíos. Con la llegada de redes neuronales avanzadas tipo AlphaFold hubo avances impresionantes. Aun así, hay momentos en que el modelo "predice" conformaciones hipotéticas que no se corresponden con ninguna proteína conocida (Rodríguez et al., 2022).

- Finalmente, Harini Veeraraghavan, jefe de un laboratorio del Memorial Sloan Kettering Cancer Center en Manhattan, citó el término "alucinación" en un artículo sobre el uso de la IA para mejorar la nitidez de las imágenes médicas borrosas. Su título en parte decía: "Resonancia magnética alucinada", abreviatura de imágenes de resonancia magnética. Los investigadores de la Universidad de Texas en Austin también han adoptado el término. "Aprendiendo de la alucinación", decía el título de su artículo (Xuesu Xiao, 2024).

Lo cierto es que los modelos de IA generativa de hoy en día se someten a rigurosas evaluaciones antes de su publicación pública, a pesar de lo cual los modelos todavía tienen el potencial de generar resultados que podrían aparecer muchas veces como falsos y no dejan de ser simples "alucinaciones".

Se estima que la IA generativa alucina entre el 3% y el 27% de las veces y aunque se han propuesto estrategias de mitigación, las alucinaciones siguen siendo un tema de investigación. Una explicación para las alucinaciones es que ocurren durante la inferencia cuando un modelo recibe entradas que son "desconocidas" para sus datos previos al entrenamiento. A continuación, el modelo genera una salida basada en la consulta de usuario fuera de distribución (OOD). El modelo hace efectivamente una "suposición ciega" y genera una salida que se asemeja a los datos previos al entrenamiento subyacentes.

Según los datos de tráfico web en 2024, solo OpenAI albergaría aproximadamente 1.7 mil millones de visitas mensuales y si el número medio de consultas de ChatGPT por visita fuera de unas 5 consultas por visita si los modelos alucinan ~3% de su producción eso nos da unas 255,000,000 alucinaciones mensuales solo para los servicios de OpenAI. La aplicación de esta tasa de alucinación a un sistema como el gráfico moral significa que hay ~ 255,000,000 de oportunidades mensuales para que un modelo genere resultados informados por valores que suenan plausibles pero incorrectos...

En física teórica, se han documentado casos en los que la IA sugiere soluciones "ilógicas" a ecuaciones diferenciales, poniendo sobre la mesa conjeturas que, a posteriori, podrían ser reformuladas con rigor matemático. En cambio, en climatología, ciertos modelos generativos han pronosticado

patrones atmosféricos que, si bien no eran del todo exactos, condujeron a la revisión de hipótesis sobre la dinámica de corrientes en distintos niveles de la atmósfera. Un exceso de alucinaciones puede generar ruido e ineficiencia, pero una dosis controlada a menudo impulsa la creatividad y anima a los científicos a reevaluar suposiciones.

La clave radica en diseñar protocolos para "aprovechar" estas ideas disparatadas sin que contaminen la línea de investigación con datos falsos irreversibles.

4. APROVECHANDO LAS ALUCINACIONES DE IA PARA LA INVESTIGACIÓN CIENTÍFICA

Los riesgos mencionados no invalidan el enorme potencial que ofrece la IA al "alucinar" en contextos donde la creatividad y la innovación son cruciales. Dentro de este escenario de posibles escenarios creativos se podrían citar los siguientes:

a*) Metodología de experimentación:* para aprovechar la imaginación de la IA, es frecuente adoptar un ciclo iterativo: la IA propone una serie de ideas—algunas muy poco ortodoxas—y a continuación se filtran con criterios humanos o mediante análisis estadísticos secundarios. Este ciclo aprovecha la velocidad y la amplitud generativa de la IA, pero requiere un control continuo para descartar lo inverosímil.

b) Aplicaciones en la ingeniería de productos: se han ensayado "hackatones" donde la IA sugiere prototipos de arquitectura, diseño industrial o circuitos electrónicos sin límites previos, y luego un equipo multidisciplinar evalúa la viabilidad de esos prototipos. Algunas propuestas consideradas disparatadas inicialmente han resultado convertirse en patentes o productos viables.

*c) Utilización de alucinaciones para fomentar hallazgos "*serendípicos*"* como en el caso de la penicilina surgió de la observación casual de un cultivo contaminado (Fleming, 1929). Las alucinaciones de la IA pueden verse como ese factor casual digital, la aparición de una idea que nadie buscaba pero que, si se atiende correctamente, podría dar frutos. Actualmente se están desarrollando *frameworks* que integran herramientas de verificación automática. Cuando la IA lanza una hipótesis, el sistema la somete a simulaciones, fuentes de datos confiables o experimentos controlados. Así, se reduce el riesgo de quedarnos con falsedades y aumentan las probabilidades de confirmación empírica.

d) Colaboración entre la creatividad humana y la de la máquina: la IA podría ser tomada como detonante del *brainstorming* de los investigadores utilizan *chatbots* internos para "disparar" ideas en reuniones de creación de

productos o en fases tempranas de proyectos de I+D. La clave está en que los humanos con conocimiento experto descarten con criterio aquellas ideas imposibles o poco éticas, y refinen las que sí tengan potencial.

La IA puede sugerir variaciones y conceptos que, de otro modo, tardarían semanas o meses en surgir. Sin embargo, la experiencia humana, la intuición y la capacidad de raciocinio sólido son indispensables para la validación final.

5. Buenas prácticas para minimizar las alucinaciones perjudiciales

Una alucinación puede darse típicamente por una extrapolación excesiva o la falta de datos de referencia. O sea, partamos de la base que el modelo no "pretende" engañar, simplemente rellena vacíos según la lógica estadística, salvo que se trate de una fabricación deliberada (*misinformación consciente*) que ocurre cuando un actor humano malicioso o un sistema programado específicamente para engañar introduce información falsa con un fin concreto. El ejemplo clásico sería la creación de *deepfakes* con intención de difamar, manipular opinión pública o cometer fraude.

En otro caso podría darse cuando el modelo surge de datos de entrenamiento incompletos o desequilibrados que refuerzan estereotipos, discriminación o favoritismos hacia determinado perfil o sesgo. El modelo, sin ser consciente, perpetúa patrones que ya existían en los datos de origen, causando inequidades en ámbitos como la selección de personal o la calificación crediticia. En este caso la IA podría no solo alucinar, sino alucinar de forma sesgada como por ejemplo, inventar datos que perpetúan estereotipos de género o raza. Esto produce daños más profundos, pues no solo se difunden falsedades, sino que refuerzan imaginarios colectivos erróneos.

Los desarrolladores tienen la tarea de revisar y equilibrar las bases de datos, para evitar que la IA genere contenidos que discriminan por razones de raza, género, religión o cualquier otra categoría protegida. La supervisión humana y la presión social son esenciales para que la tecnología avance en beneficio de todos y no perpetúe injusticias. La coexistencia de ventajas creativas y riesgos de desinformación obliga a establecer metodologías claras para el desarrollo y despliegue de sistemas de IA, como ser:

a. Calidad de datos y actualizaciones continuas: los conjuntos de datos deben ser revisados y curados con rigor, para reducir sesgos y omisiones que alimenten las alucinaciones. Además, es fundamental actualizar el modelo con frecuencia para que no recurra a información obsoleta, evitando confusiones. Debería Implementarse herramientas que midan la frecuencia de alucinaciones en ambientes de producción y realimenten al equipo de desarrollo para ajustes inmediatos, como por ejemplo, un panel de control que

registre las respuestas marcadas como "inciertas" o "erróneas" por parte de los usuarios.

b. Mecanismos de verificación y contrastación de hechos: se usan bases de conocimiento confiables (p. ej., Wikipedia, PubMed, elDial.com) o motores de búsqueda especializados para corroborar la validez de la respuesta antes de mostrarla al usuario. Cuando se detecta una discrepancia, la IA puede optar por advertir al usuario o reintentar la respuesta. En escenarios críticos (asesorías médicas, jurídicas o financieras), se recomienda que profesionales de carne y hueso revisen el contenido antes de tomar decisiones. Esta capa de supervisión reduce drásticamente el riesgo de que un error se transforme en un perjuicio real para el usuario.

c. Desarrollo de una IA interpretable y modelos explicables: algunos modelos se basan en técnicas que permiten rastrear cuáles palabras o tokens han sido determinantes en la salida como las denominadas "arquitecturas transparentes" que ayudan a identificar los pasos que llevaron a la alucinación. Se evalúa la posibilidad de que la IA justifique sus respuestas con referencias o explicaciones internas, aunque este objetivo es complejo en modelos muy grandes. La explicabilidad de los sistemas facilita la auditoría externa y la detección temprana de patrones de error y permite al usuario evaluar la solidez de la respuesta, en lugar de confiar ciegamente.

d. Desarrollo de métodos técnicos para afrontar la alucinación: las soluciones a nivel de ingeniería y ciencia de datos son esenciales para mitigar el problema y reforzar la fiabilidad de la IA como la generación con recuperación de datos (Retrieval-Augmented Generation, RAG**)** que consiste en que el modelo no responda únicamente con su "memoria interna" entrenada, sino que busque en bases documentales de alta fiabilidad. Así, se reduce la probabilidad de inventar información, pues el modelo se ancla a fuentes concretas que puede citar y referenciar. El pipeline suele ser: 1) el usuario pregunta, 2) el sistema rastrea documentos relevantes, 3) el modelo de lenguaje genera la respuesta basándose en los datos recuperados. Esto se aplica en dominios como la medicina (consultar bibliotecas de literatura científica) o el periodismo (verificar noticias en tiempo real).

Existen algoritmos que, además de la respuesta, generan un valor numérico que indica la confianza interna, y de esta forma, la IA "sabe" que su respuesta puede ser incierta y lo comunica, permitiendo al usuario decidir si seguir adelante o verificar. Las organizaciones pueden fijar umbrales. Por ejemplo, si la confianza es menor de un 60%, la respuesta no se muestra o se acompaña de una advertencia visible. Esto crea una experiencia de usuario más honesta y reduce la "falsa certeza" que a veces proyecta el modelo.

e. Aprendizaje por refuerzo con retroalimentación humana (RLHF): Consiste en exponer al modelo a múltiples ejemplos de respuestas. Los usuarios

o anotadores humanos puntúan la calidad, exactitud o utilidad de cada salida y el modelo se ajusta, reforzando las respuestas correctas y penalizando las alucinaciones. (Bergmann, 2023)

Ello requiere de grandes cantidades de tiempo y recursos humanos calificados, sobre todo si se cubren múltiples idiomas y dominios. Aun así, representa uno de los enfoques más prometedores para reducir la tasa de alucinaciones, ya que integra directamente la noción de verdad o verosimilitud en el proceso de aprendizaje.

6. La controversia: ¿debemos aceptar o eliminar las alucinaciones?

La comunidad académica se divide entre quienes creen que deben minimizarse al máximo y quienes ven en ellas una oportunidad inexplorada de creatividad.

a. Teoría de la creatividad asistida: según algunos investigadores, la IA puede actuar como un socio creativo que introduce ideas "salvajes", las cuales, al ser revisadas, pueden catalizar descubrimientos genuinos. Este fenómeno se asemeja a las técnicas de *brainstorming* donde se fomenta que los participantes aporten ideas sin censura, para luego filtrarlas. En algunas industrias, podría surgir la tentación de vender "soluciones mágicas" basadas en IA generativa, sin advertir que sus resultados provienen de alucinaciones no validadas. De ahí la necesidad de una autorregulación sectorial y de la concienciación del público.

b. Imaginación controlada: otra propuesta que cobra fuerza es la de disponer de dos "modos" en la IA. Uno, acotado y orientado a la precisión, donde se emplean todas las técnicas de verificación; y otro más abierto, pensado para la exploración y la generación de ideas. El usuario podría alternar entre ellos dependiendo de si necesita idear o confirmar hechos objetivos. En campos críticos, es esencial bloquear de antemano la propensión a la alucinación o, al menos, exigir una supervisión exhaustiva. No se sugiere, por tanto, un uso irrestricto de la IA en la toma de decisiones de alto riesgo (por ejemplo, diagnósticos médicos finales o sentencias judiciales).

La cuestión reside de cómo convertir el error en un acierto potencial es objeto de investigación y debate continuo y para comprender el impacto real de las alucinaciones, no basta con la visión técnica. Hay que analizar aspectos antropológicos (cómo las personas reaccionan a los datos falsos), sociológicos (cómo se propagan en redes) y filosóficos (qué significa la "verdad" en un sistema estadístico). Esta es un área donde confluyen epistemología, derecho, ética y programación.

Algunas universidades y corporaciones han lanzado laboratorios conjuntos para estudiar la aplicabilidad de la IA a problemas complejos, considerando tanto la eficiencia como los riesgos de alucinar.

Esto podría derivar en directrices pioneras para la gestión consciente y constructiva de la IA que podría contemplar la existencia de *agentes de IA* en laboratorios científicos. Así se vislumbra la posibilidad de laboratorios donde brazos robóticos, dirigidos por IA, diseñen y ejecuten experimentos automáticamente.

Si la IA alucina en tal contexto—por ejemplo, mezclando sustancias peligrosas—, los riesgos pueden ser altos, reforzando la necesidad de estrictas salvaguardas. El aspecto positivo es que estos agentes podrían realizar ensayos in vitro a una velocidad inalcanzable para los humanos, abriendo la puerta a descubrimientos más rápidos en fármacos, materiales o biotecnología.

Todo ello siempre y cuando se controle la veracidad y la fiabilidad de las hipótesis de partida.

Es probable que veamos un refinamiento de técnicas como la "minería de conocimiento" en tiempo real, la retroalimentación activa de la comunidad y la verificación cruzada con fuentes externas confiables. Esta infraestructura tecnológica irá más allá del simple "modelo generativo", integrándose profundamente con mecanismos de seguridad de IA, que podría llegar a ser un colaborador confiable si logra conjugar creatividad y rigor, pero para ello, hace falta un progreso constante en la calibración de sus respuestas y en la transparencia de sus procesos.

7. Hacia una alucinación de IA responsable

Las alucinaciones en la IA representan uno de los desafíos y oportunidades más notables de nuestra era tecnológica. Al mismo tiempo, exponen límites fundamentales en la forma en que estos sistemas "comprenden" o, más bien, correlacionan datos.

Como expuse recientemente La IA generativa, como los modelos de lenguaje (GPT, DALL-E) tiene capacidades que pueden considerarse *formas rudimentarias de inteligencia* dado que aprenden patrones complejos en grandes cantidades de datos, pueden generar contenidos creativos (textos, imágenes, música) que antes se atribuían exclusivamente a los humanos y toman decisiones algorítmicas para "razonar" sobre preguntas o problemas. Aunque carecen de conciencia o intencionalidad, su capacidad de producir contenidos que simulan procesos cognitivos humanos plantea interrogantes si deberían considerarse una forma incipiente de inteligencia no humana y en el futuro, redes avanzadas de IA, combinadas con tecnologías

avanzadas podrían llegar a exhibir comportamientos emergentes que superen las capacidades de las inteligencias humanas individuales, como redes interconectadas de dispositivos inteligentes o modelos de IA colaborativa que desarrollen habilidades no previstas por sus programadores.

La IA generativa podría ser considerada inteligente, por lo tanto, pero con matices importantes, dado que la IA generativa no tiene conciencia, emociones ni intencionalidad. Simula inteligencias humanas, pero no comprende verdaderamente, aprende de datos previos, pero no crea conocimientos nuevos en el sentido humano del término, pudiendo combinar información de formas que parecen innovadoras. En conclusión, la IA generativa es un paso importante hacia nuevas formas de inteligencia no humana, aunque aún es limitada. La verdadera aparición de inteligencias no humanas más avanzadas podría depender de futuros avances en áreas como el aprendizaje automático, la computación cuántica y la bioingeniería. Estos desarrollos no solo expandirían las posibilidades de la inteligencia, sino que podrían llegar a redefinir lo que significa "ser inteligente". (Granero, 2025)

Dado que la humanidad va a depender cada vez más de la IA, debemos abordar conscientemente el de tema de la existencia de las alucinaciones y su eventual futura aplicación al área de la investigación, teniendo en cuenta lo presente:

a) Equilibrio: aceptar la necesidad de cierto margen de creatividad y experimentación en contextos de innovación, pero también salvaguardar la información cuando estén en juego la salud, la justicia o la seguridad.

b) Infraestructura: diseñar e implementar mecanismos técnicos robustos (RAG, *fact-checking*, RLHF, etc.) que mitiguen la proliferación de contenido inventado.

c) Ética y regulación: establecer lineamientos claros que responsabilicen a desarrolladores y usuarios, fomentando la transparencia y la integridad.

d) Educación: formar a profesionales y a la sociedad en general para usar la IA de forma crítica, reconociendo sus fortalezas y sus limitaciones.

Las "alucinaciones" de LLM pueden reflejar la creatividad digital, lo que sugiere un nuevo ámbito de imaginación de la IA, sin embargo, estos fenómenos pueden surgir de un sobreajuste y contaminación de datos, lo que conduce a salidas sin sentido.

Los LLM multimodales impulsan la creatividad más allá del lenguaje, fusionando elementos visuales, textuales y auditivos, y adoptar, por lo tanto la imprevisibilidad de LLM podría mejorar la resolución creativa de problemas, mejorando la inteligencia humana y de IA, lo que no quita que debamos estar preparados para contestar preguntas como estas, por ejemplo:

a) ¿Por qué los sistemas de IA "inventan" datos en lugar de admitir su desconocimiento?: Porque su arquitectura se basa en predecir la secuencia más probable de palabras según su entrenamiento. No existe una función intrínseca de "reconocer ignorancia". Al menos, no en la mayoría de los modelos actuales (Bender & Koller, 2020), y además, muchas implementaciones comerciales han priorizado la fluidez y la coherencia sobre la honestidad de admitir "no sé".

b) ¿Pueden eliminarse por completo las alucinaciones? Lograr una eliminación total es muy complejo debido a la naturaleza estadística de los modelos. Siempre habrá áreas de conocimiento pobremente representadas donde la IA "improvisará". No obstante, se pueden reducir con estrategias como la verificación externa y el aprendizaje por refuerzo basado en retroalimentación humana (Brown et al., 2020).

c) ¿Cómo detectar si una IA está "alucinando"*?*: Comparar la información con fuentes confiables o verificar si la IA cita referencias inexistentes. Ciertas plataformas ofrecen un sistema de puntuación de confianza o etiquetan las partes de la respuesta con mayor incertidumbre (Lewis et al., 2020).

d) ¿Es seguro utilizar las alucinaciones de la IA para asesoría médica o legal?: No se recomienda sin supervisión de profesionales humanos. Las consecuencias de un error en medicina o derecho pueden ser irreparables. En algunas jurisdicciones, se exige por ley que la decisión final en estos ámbitos la tome un especialista acreditado (Salazar & García, 2021).

e) ¿Las futuras IA se entrenarán para alucinar de forma deliberada? Es probable que aparezcan "módulos de imaginación controlada" o entornos de simulación donde la IA pueda asumir conjeturas arriesgadas sin confundir al usuario con datos presentados como verdades absolutas (Davis, 2021).

- Estos experimentos se centran en potenciar la exploración creativa, siempre con barreras claras para que la información no se dé por cierta automáticamente.

La tecnología no opera en el vacío; su adopción y efectos dependen de la intervención humana en todos los niveles, por ello se piensa en sistemas con humanos en el circuito (*human-in-the-loop*) con la necesaria implicación de expertos, reguladores y usuarios durante todo el ciclo de vida de la IA reduce la probabilidad de usos irresponsables, dado que el humano no solo valida la información, sino que también aporta contexto y criterio moral.

Surgirían así plataformas donde el usuario interactúa con la IA, marcando respuestas dudosas y recibiendo retroalimentación inmediata, teniendo conciencia que esta convergencia busca no sustituir a las personas, sino empoderar la colaboración entre ambas inteligencias.

Así se considera crucial que los profesionales que usen IA —médicos, abogados, ingenieros— estén formados para detectar posibles alucinaciones y tengan la pericia de contrastar la información y los planes de estudio en universidades y centros de formación podrían incluir módulos específicos sobre límites y posibles errores de la IA.

Un aspecto que apenas se toca en estos estudios es lo que sucede *después de la idea inicial*. La creatividad es un proceso. Una idea que parece derivada al principio puede convertirse en algo único con el desarrollo. O bien, una idea que suena original puede desvanecerse o converger hacia una implementación similar a otras cuando realmente la construye. Muchas ideas humanas que llegaron a callejones sin salida y nunca llegaron a ninguna parte.

El punto es que la diversidad inicial es solo una instantánea en el tiempo. Lo que realmente nos importa es la *diversidad de resultados finales o soluciones que llegan a buen término*. Sería fascinante ver un estudio longitudinal en el que los equipos tomen ideas iniciales (asistidas por IA o no) y las desarrollen durante meses para ver si las generadas por IA exhiben menos variedad en los productos finales.

8. Conclusión: ¿La IA daña la creatividad? o ¿Cómo puede la IA mejorarla?

La percepción de que la IA y la creatividad humana están reñidas es común, pero el análisis sugiere que un encuadre más productivo reside en emparejar óptimamente las fortalezas humanas y las de la IA. Esta colaboración debe basarse en el objetivo creativo específico: exploración frente a explotación, amplitud frente a profundidad, e innovación frente a optimización.

Los seres humanos sobresalen en "pensar fuera de la caja", generando ideas verdaderamente extrañas, novedosas y que cambian de contexto, con matices como la resonancia emocional o el contexto cultural. La IA, por su parte, destaca por "combinar un vasto conocimiento" y asegurar que los resultados cumplan con ciertos criterios (por ejemplo, mantenerse en el tema), actuando como un "espejo estadístico" del conocimiento y el lenguaje humanos.

En un flujo de trabajo ideal, la IA podría generar una gran cantidad de ideas, y el ser humano podría elegir una de ellas para luego pedir a la IA que genere nuevas ideas al respecto, alternando así el uso de la IA para ir de lo amplio a lo profundo. Millones de mentes humanas, reflejadas en los datos de entrenamiento de la IA, podrían haber convergido en ideas similares por una buena razón.

Especialmente en escenarios prácticos y sensibles al tiempo, el estilo de "convergencia creativa" de la IA a veces puede ser más valioso que la "divergencia creativa" pura. Si bien la IA es excelente para generar muchas ideas rápidamente, también puede conducir por un camino específico que no siempre es el más adecuado en el proceso de lluvia de ideas. A menudo, es preferible que los humanos generen algunas ideas iniciales por sí mismos (eligiendo primero el problema) y luego utilicen la IA para refinar y explorar esas ideas con mayor rapidez (2025, Jing Hu).

La conclusión que se desprende de este análisis es que el futuro no es de reemplazo humano por la IA, sino de una asociación coevolutiva. A medida que las capacidades de la IA avanzan, los roles humanos se desplazarán hacia áreas donde nuestras fortalezas cognitivas únicas (como el razonamiento ético, la comprensión contextual profunda, la inteligencia emocional y el pensamiento verdaderamente divergente) son más valiosas. Esto implica una adaptación continua de las habilidades humanas y los sistemas educativos para prosperar en un ecosistema de "inteligencia híbrida", donde las innovaciones más impactantes surgen de la interacción sinérgica entre mentes humanas y artificiales.

La rápida evolución de la IA a menudo supera la capacidad de la sociedad para asimilar su impacto. Corresponde a las instituciones y gobiernos establecer marcos éticos que equilibren los beneficios y las salvaguardas, donde el debate público también sirva para plasmar los valores sociales en el diseño de estas herramientas, evitando suposiciones puramente tecnocráticas. En última instancia, la IA seguirá "alucinando" en determinados contextos. La pregunta crucial es si sabremos encauzar esas alucinaciones para que se conviertan en oportunidades de descubrimiento y creatividad, en lugar de provocar caos informativo. El éxito de esta convivencia dependerá tanto de la ingeniería y la ciencia de datos como de nuestra visión humana y nuestra capacidad de gobernanza colectiva.

Se podría imaginar un futuro donde la "alucinación" controlada se utilice conscientemente para romper barreras conceptuales en ciencia y tecnología, mientras que el "modo veraz" sirva para labores sensibles en las cuales no hay margen de error.

La observación de que la "rápida evolución de la IA a menudo supera la capacidad de la sociedad para asimilar su impacto", junto con la necesidad de que "instituciones y gobiernos establezcan marcos éticos que equilibren los beneficios y las salvaguardas" , y la persistencia de las alucinaciones de la IA, subraya que este no es un problema que se pueda "resolver" de una vez por todas, sino un desafío continuo que requiere una adaptación constante.

Los enfoques regulatorios tradicionales y estáticos son insuficientes para gobernar una IA en rápida evolución. Lo que se necesita es un modelo

de "gobernanza adaptativa" caracterizado por el monitoreo continuo, ajustes flexibles de políticas, colaboración internacional y una sólida participación pública. Esto exige un enfoque proactivo para anticipar los desafíos emergentes de la IA, en lugar de reaccionar a las crisis, y para asegurar que el progreso tecnológico se alinee con los valores sociales y los principios éticos. El éxito de la integración de la IA depende de la capacidad colectiva para establecer y evolucionar estos mecanismos de gobernanza.

Las alucinaciones de la IA representan tanto una amenaza como una oportunidad. Su capacidad de generar ideas novedosas puede acelerar la investigación si se combina con validación humana y mecanismos de control.

El futuro no está en erradicar estas "ficciones plausibles", sino en integrarlas responsablemente en flujos de trabajo creativos y científicos, redefiniendo la colaboración entre humanos y máquinas bajo principios de transparencia, ética y gobernanza adaptativa.

Referencias

(Relacionadas con los temas tratados cronológicamente en el trabajo).

Psycology Today, https://www.psychologytoday.com/us/basics/creativity

Psycology Today, https://www.psychologytoday.com/us/basics/boundaries

Nosta, Juan (2024) *¿Son las alucinaciones de la IA un vistazo a la creatividad digital?* https://www.psychologytoday.com/us/blog/the-digital-self/202402/are-ai-hallucinations-a-glimpse-into-digital-creativity

Wikipedia, consulta del 22 de abril 2025 https://es.wikipedia.org/wiki/Baricitinib

Roberts, R. M. (1989). *Serendipity: Accidental Discoveries in Science*. Wiley. https://archive.org/details/serendipityaccid00robe/page/n9/mode/2up

Deep Dream Generator, consulta del 22 de abril 2025, https://deepdreamgenerator.com/

Broad, William J. "Cómo la IA alucinatoria ayuda a la ciencia a soñar con grandes avances", The New York Times, 23 de diciembre de 2024, https://www.nytimes.com/2024/12/23/science/ai-hallucinations-science.html#

Baker, David, CV https://sites.uw.edu/biochemistry/faculty/david-baker/

Premio Nobel, Comunicado de prensa https://www.nobelprize.org/prizes/chemistry/2024/press-release/

Patentes de David Baker, consulta del 22 de abril 2025 https://patents.google.com/patent/US11655278B2/en?inventor=%E2%80%9CDavid+Baker%E2%80%9D&assignee=%E2%80%9CUniversity+of+Washington%E2%80

%9D&oq=%E2%80%9CDavid+Baker%E2%80%9D+%E2%80%9CUniversity+of+Washington%E2%80%9D

"Diseño geométrico asistido por IA de catéteres antiinfecciosos" *https://www.science.org/doi/full/10.1126/sciadv.adj1741*

Rodríguez, C. et al. (2022). *AlphaFold and the Brave New World of Protein Structure Prediction.* Nature Reviews Molecular Cell Biology. OJO Z. Faidon Brotzakis et alios "Predicción AlphaFold de conjuntos estructurales de proteínas desordenadas" https://www.nature.com/articles/s41467-025-56572-9

Patente: "Integración de la resonancia magnética alucinada multimodal con la tomografía computarizada para ayudar a la segmentación del tumor pulmonar mediastínico" *https://link.springer.com/chapter/10.1007/978-3-030-32226-7_25*

Xuesu Xiao et alios. *Aprendiendo de la alucinación (LfH) https://www.cs.utexas.edu/~xiao/Research/LfH/LfH.html*

"El desafío de las alucinaciones en la Inteligencia Artificial Generativa y cómo superarlas" https://www.pcdemano.com/sc/29400/

Fleming, A. *(1929). On the antibacterial action of cultures of a Penicillium, with special reference to their use in the isolation of B. influenzae. British Journal of Experimental Pathology, 10(3), 226–236. https://pubmed.ncbi.nlm.nih.gov/11545337/*

Faster Capital "Comprensión De Las Redes Generativas Adversarias (GAN)" consulta del 22 de abril 2025, https://fastercapital.com/es/tema/comprensi%C3%B3n-de-las-redes-generativas-adversarias-(gan).html

Bergmann, David. "¿Qué es RLHF?" https://www.ibm.com/es-es/think/topics/rlhf

Paster, Patrick, enero 2025 en Live Science "Una nueva molécula brillante, inventada por IA, habría tardado 500 millones de años en evolucionar en la naturaleza, según los científicos" *https://www.livescience.com/technology/artificial-intelligence/new-glowing-molecule-invented-by-ai-would-have-taken-500-million-years-to-evolve-in-nature-scientists-say*

Informe del Ministerio español para la Transformación Digital. "El ministerio para la Transformación Digital y de la Función Pública lanza la primera convocatoria del Sandbox de IA para sistemas de alto riesgo" consultado el 22 de abril 2025 https://digital.gob.es/portalmtdfp/comunicacion/sala-de-prensa/comunicacion_SEDIA/2024/12/2024-12-20_02.html

Bender, E. M., & Koller, A. (2020). *Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Dat*a. Association for Computational Linguistics (ACL). *https://r.search.yahoo.com/_ylt=AwrFa02TK9NnPAIAZUar9Qt.;_ylu=Y29sbwNiZjEEcG9zAzIEdnRpZAMEc2VjA3Ny/RV=2/RE=1743102099/RO=10/RU=https%3a%2f%2faclanthology.org%2f2020.acl-main.463.pdf/RK=2/RS=OdTRT6ibKk71x8fsqfttZHYazKg-*

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., & Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems, 33*, 1877-1901. https://r.search.yahoo.com/_ylt=AwrFEwPpLNNnDAIAdG2r9Qt.;_ylu=Y29sbwNiZjEEcG9zAzEEdnRpZAMEc2VjA3Ny/RV=2/RE=1743102441/RO=10/RU=https%3a%2f%2fpapers.nips.cc%2fpaper%2f2020%2ffile%2f1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf/RK=2/RS=kCUjMJha8pklLiQVRgtrmoxzmFE-

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N. & Riedel, S. (2020). Generación aumentada por recuperación para tareas de PNL con uso intensivo de conocimiento *33*, 9459-9474. https://arxiv.org/abs/2005.11401

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Flores B., ... & Collins, J. J. (2020). Un enfoque de aprendizaje profundo para el descubrimiento de antibióticos *180*(4), 688-702. https://pubmed.ncbi.nlm.nih.gov/32084340/

Katherine O'Toole, Emőke-Ágnes Horvát. *Extendiendo la creatividad humana con IA* https://www.sciencedirect.com/science/article/pii/S2713374524000062

Van Andel, P. (1994). "Anatomy of the Unsought Finding." British Journal for the Philosophy of Science. https://philpapers.org/rec/VANAOT

Granero, Horacio R. "El dilema si la Inteligencia Artificial puede razonar o no", en el Capítulo "Ética, IA y Derechos Digitales" de libro Inteligencia Artificial en el Derecho, publicado por elDialLibros y Editorial Tirant, pág. 415, Marzo 2025

Floridi, L., & Taddeo, M. (2023). The ethics of artificial intelligence: Principles, challenges, and opportunities. Oxford University Press. *https://academic.oup.com/book/ 46748?login=false*

Gawehn, E., Hiss, J. A., & Schneider, G. (2016). Deep learning in drug discovery. *Molecular Informatics, 35*(1), 3-14. https://onlinelibrary.wiley.com/doi/10.1002/ minf.201501008

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. https://link.springer.com/article/10.1007/s10710-017-9314-z

Ji, Z., Lee, N., Frieske, R., Su, D., & Xu, Y. (2023). Encuesta sobre la alucinación en la generación del lenguaje natural . *arXiv preprint arXiv:2302.06413. https://arxiv.org/abs/ 2202.03629*

Kuhn, T. S. (1962). *La estructura de las revoluciones científicas*. University of Chicago Press. https://drive.google.com/file/d/1Y8gptT1GK2V9KpBzbLBSEoPDdgJ7w2kQ/view

Maynez, J., Narayan, S., Bohnet, B., & McDonald, R. (2020). Sobre la fidelidad y la factualidad en el resumen abstracto. *ACL. https://aclanthology.org/2020.acl-main.173/*

Jing Hu *¿Es realmente una buena idea hacer una lluvia de ideas con IA?*, publicado el Medium el 11.07.2025 https://medium.com/@jinghuu/is-brainstorming-with-ai-really-a-good-idea-eeb8b42c3c6d?source=email-b9964a806bdf-1753075267734-digest.reader-3fe99b2acc4-eeb8b42c3c6d----3-102----------0f31900d_250a_4c60_98a8_205a389d4d8a-1

¿Quién cuida cuando cuida una máquina? Derecho, afecto y adjudicación en clave trialista

Por Maria Isolina Dabove (*)

Resumen: Este artículo propone una lectura crítica y trialista sobre los cuidados robotizados en la vejez, interrogando el papel del derecho frente a la creciente automatización de vínculos asistenciales. A partir de la pregunta "¿Quién cuida cuando cuida la máquina?", se analiza la adjudicación algorítmica como forma emergente de distribución de afectos, recursos y reconocimientos. El trabajo incorpora los estándares normativos de la Convención Interamericana sobre la Protección de los Derechos Humanos de las Personas Mayores y de la Opinión Consultiva OC-31/25 de la Corte Interamericana de Derechos Humanos, que reconocen el cuidado como derecho humano autónomo, fundado en la dignidad, la corresponsabilidad social y la justicia intergeneracional. Desde una filosofía jurídica trialista —que articula lo normativo, lo sociológico y lo axiológico— se problematizan los modelos de cuidado automatizado, los riesgos de deshumanización y la necesidad de una praxis jurídica encarnada, sensible a la vulnerabilidad y a la poética del vínculo.

Palabras clave: Vejez; cuidados; Inteligencia Artificial; filosofía jurídica; trialismo; dignidad; justicia intergeneracional; robotización; derecho y tecnología.

Who cares when the machine cares? Law, affect, and adjudication in trialist key (**)

Abstract: This article offers a critical and trialist reading of robotized care in old age, interrogating the role of law in the face of the growing

---

(*) María Isolina Dabove es profesora e investigadora especializada en filosofía del derecho, derechos humanos y derecho de la vejez. Es profesora titular en la Universidad Nacional de Rosario y en la Universidad Nacional del Centro de la Provincia de Buenos Aires, e investigadora Principal en el CONICET. Su trabajo se centra en teoría del derecho, interseccionalidad y justicia intergeneracional. Contacto: isolinadabove@gmail.com

(**) María Isolina Dabove is a legal scholar specializing in philosophy of law, human rights, and elder law. She is a Full Professor at the National University of Rosario and the National University of the Center of the Province of Buenos Aires, and a Principal Researcher at CONICET. Her work focuses on legal theory, intersectionality, and intergenerational justice. Contact: isolinadabove@gmail.com

automation of assistive relationships. Starting from the question "Who cares when the machine cares?", it analyzes algorithmic adjudication as an emerging form of distributing affect, resources, and recognition. The work incorporates the normative standards of the Inter-American Convention on the Protection of the Human Rights of Older Persons and the Advisory Opinion OC-31/25 of the Inter-American Court of Human Rights, both of which recognize care as an autonomous human right grounded in dignity, social co-responsibility, and intergenerational justice. From a trialist legal philosophy —which articulates the normative, sociological, and axiological dimensions— the article problematizes models of automated care, the risks of dehumanization, and the need for an embodied legal praxis, sensitive to vulnerability and the poetics of the bond.

Keywords: Old age; care; Artificial intelligence; legal philosophy; trialism; dignity; intergenerational justice; robotization; law and technology.

1. Introducción

El advenimiento de los cuidados robotizados en la vejez suscita interrogantes filosóficos, jurídicos y éticos de gran densidad. No se trata simplemente de la incorporación de artefactos tecnológicos en la administración de servicios vinculados al bienestar de las personas mayores, sino de la irrupción de una nueva era en la que los límites entre lo humano y lo poshumano se difuminan peligrosamente. Si, como sugiere la filosofía jurídica trialista, la experiencia jurídica es una unidad compleja compuesta por tres elementos irreductibles—realidad, valor y norma—el análisis de los cuidados algorítmicamente mediados exige una profundidad filosófico-jurídica ineludible. ¿A quién corresponde la responsabilidad subjetiva y objetiva ante una eventual vulneración de la dignidad de la persona mayor a cargo de un autómata? ¿Es posible hablar de afecto en la administración algorítmica del cuidado, o el afecto es precisamente aquello que se resiste a toda formalización computacional? ¿Cómo se resignifica el principio de dignidad como valor jurídico primordial en la era de la inteligencia artificial?

Estas preguntas dialogan con los más recientes desarrollos normativos y jurisprudenciales, especialmente a partir del reconocimiento del cuidado como derecho humano autónomo por la Convención Interamericana sobre la Protección de los Derechos Humanos de las Personas Mayores (ratificada en Argentina por ley 27.360/17, con jerarquía constitucional por ley 27700/22) [1] y por la Opinión Consultiva OC-31/25 de la Corte Interame-

[1] OEA, *Convención Interamericana sobre la Protección de los Derechos Humanos de las Personas Mayores*, disponible en: https://www.oas.org/es/sla/ddi/docs/tratados_multilaterales_interamericanos_A-70_derechos_humanos_personas_mayores.pdf

ricana de Derechos Humano [2]. El derecho, interpelado por avances tecnológicos fulgurantes, se ve urgido de recomponer sus arquitecturas categoriales y sus procedimientos de adjudicación, no solo para responder a los nuevos desafíos normativos, sino para no abdicar de su función más profunda: sostener la dignidad humana en medio de la intemperie del porvenir técnico.

Este artículo se propone, siguiendo la matriz trialista elaborada por Goldschmidt y desarrollada por la doctrina contemporánea, abordar el fenómeno de los cuidados robotizados en la vejez con espesura conceptual, hibridando filosofía, derecho positivo, afecto y una retórica cercana a la poética. La estructura se despliega en cinco secciones: (1) filosofía jurídica trialista; (2) cuidados robotizados y adjudicación algorítmica; (3) justicia intergeneracional y dignidad como valor jurídico; (4) desafíos normativos emergentes; (5) importancia de la filosofia del afecto y de la ética del cuidado como horizontes del mundo jurídico ante la inteligencia artificial. Cada segmento dialoga con la doctrina, la jurisprudencia y los más recientes estándares internacionales, integrando la literatura de este tiempo.

2. Filosofía jurídica trialista para los cuidados

a) Nociones básicas. La filosofía jurídica trialista, concebida originalmente por Werner Goldschmidt, constituye una de las corrientes más fecundas y originales del derecho contemporáneo iberoamericano [3]. Su vitalidad reside en la afirmación de que la experiencia jurídica no puede reducirse ni al dato normativo, ni a las realidades físicas, ni a los valores axiológicos, sino que comporta la interacción inescindible de lo real, lo normativo y lo valorativo. Tales dimensiones se entreveran en todo fenómeno jurídico, pero cobran especial pregnancia cuando se trata de analizar

---

[2] Corte Interamericana de Derechos Humano, *OC-31/25 El contenido y el alcance del derecho al cuidado y su interrelación con otros derechos*, disponible en: https://corteidh.or.cr/docs/opiniones/seriea_31_es_altocontraste.pdf

[3] Sobre la teoria trialista puede consultarse: Goldschmidt, Werner, Introducción filosófica al derecho. La teoria trialista del mundo jurídico y sus horizontes, 6ta. ed., Buenos Aires, Depalma, 1987, p. 5 a 37; La ciencia de la justicia (Dikelogía), 2° ed., Buenos Aires, Depalma, 1986, p. 3 a 263; Ciuro Caldani, Miguel Angel; Una teoría trialista del derecho, Buenos Aires, Astrea, 2020, p. 31 a 162; Metodología jurídica y Lecciones de Historia de la Filosofía del Derecho, Rosario, Zeus, 2007, passim; La conjetura del funcionamiento de las normas jurídicas. Metodología jurídica, Rosario, Fundación para las Investigaciones Jurídicas, 2000, passim; Chaumet, Mario E.; Argumentación. Buenos Aires, Astrea, 2019, p. 160; Dabove, María Isolina; "La filosofía jurídica argentina: sus cursos de vida en Latinoamérica y su porvenir", en Contrapuntos de Teorías del Derecho. Diálogos con Manuel Atienza, Renato Rabbi Baldi coord., Buenos Aires, Astrea, 2025, p. 185 a 207; Argumentación jurídica y eficacia normativa: problemáticas actuales del funcionamiento del derecho, "Díkaion", 2015, vol. 12, p 36 a 65. DOI: 10.5294/Dika.2015.24.1.3; Interpretación "interseccional". La deconstrucción de los sentidos implícitos, "Anales: Revista de la Facultad de Ciencias Juridicas y Sociales, UNLP", La Plata, Universidad Nacional de La Plata, año 20, n° 53, 2023, p. 747 a 772; Los derechos de los ancianos, Buenos Aires, Ciudad Argentina, 2002, p. 37 a 109; Derecho de la vejez. Fundamentos y alcance, 1° ed. 1° reimp., Buenos Aires, Astrea, 2021, passim.

dimensiones inéditas como el cuidado robotizado, donde lo material (máquina adjudicataria de posibilidades y obstáculos), lo normativo (principios y reglas sobre derechos y deberes) y lo valorativo (afecto, dignidad, sentido) se tensan hasta el límite.

Contra toda tentación de monismo jurídico, el trialismo postula la radical insuficiencia de cualquier análisis que pretenda agotar la jurídica en los términos de una exclusiva de sus "mundos". Lo jurídico no es ni solo hecho, ni solo valor, ni solo norma, sino que existe exactamente en la frontera donde esos tres estratos se interpenetran. La aplicación trialista aparece, así como una actitud epistemológica y ontológica que, más que describir, horada el objeto jurídico hasta dejar expuestos sus cimientos constitutivos [4].

En el contexto de los cuidados robotizados, el trialismo fuerza al operador jurídico a repensar preguntas clásicas con mayor profundidad: ¿es suficiente la existencia de un protocolo normativo de cuidado para afirmar la juridicidad de la prestación? ¿Qué ocurre cuando el robot, cumpliendo estrictamente lo programado, fracasa en captar el valor afectivo esencial para la protección de la persona mayor? ¿Puede la máquina participar de la experiencia jurídica en tanto operadora de cuidado, o su condición de cosa limita su inserción al mero plano de la realidad, sin acceso pleno a la normatividad ni al valor?

Desplegar el análisis trialista implica, pues, pensar los cuidados asistidos por IA en tres planos: (a) la materialidad de la máquina y sus funciones, (b) los sistemas normativos que la regulan y habilitan, (c) el valor intrínseco que subyace a la relación de cuidado, especialmente bajo la luz de la dignidad humana y la centralidad del afecto. Solo una reflexión jurídica nutrida por esta tridimensionalidad puede dar cuenta cabal de este fenómeno emergente.

b) Cuidados robotizados en la vejez. La robótica aplicada al cuidado de personas mayores es, en la actualidad, uno de los campos tecnológicos que más rápidamente se expande en países con altos índices de longevidad y déficit de personal en el sector salud [5]. Robots humanoides, asistentes

---

[4] Zucchi Héctor A.; *Trialismo, razón clasificatoria y experiencia jurídica*, Academia Nacional de Derecho y Ciencias Sociales de Córdoba, disponible en: https://www.acaderc.org.ar/wp-content/blogs.dir/55/files/sites/55/2020/11/Trialismorazonclasificatoria-experienciajur%C3%ADdica.pdf; Galati, Elvio; *La teoria trialista del mundo jurídico. Un pensamiento jurídico complejo,* Ciudad Autónoma de Buenos Aires: Teseo; Universidad Abierta Interamericana, 2021, p. 23 a 303.

[5] Peralta, Samantha; *Una vista al futuro: el uso de robots humanoides para el cuidado de adultos mayores*,"Verdeza", 21 de enero de 2025 disponible en: https://www.verdeza.com/post/una-vista-al-futuro-el-uso-de-robots-humanoides-para-el-cuidado-de-adultos-mayores; Bahamonde, Joaquín, *E-BAR: el robot que revoluciona el cuidado de ancianos en casa*,"Infobae", 19 de mayo 2025, disponible en: https://www.infobae.com/america/ciencia-america/2025/05/19/e-bar-el-robot-que-revoluciona-el-cuidado-de-ancianos-en-casa/; Rhys Cox, Samuel; Cheong, Grace; Tsang Ooi, Wei; *The use of*

virtuales, plataformas algorítmicas capaces de monitorear variables vitales, administrar medicamentos y hasta "conversar"—todo ello configura un nuevo horizonte para el cuidado gerontológico. No obstante, la fascinación técnica contrasta con la sospecha: ¿puede considerarse genuino el cuidado mediado por máquinas? ¿Es trasladable el deber jurídico de protección a un algoritmo?

Dentro de la literatura especializada se distinguen al menos tres categorías funcionales de robots aplicados a la vejez: a) robots asistenciales, enfocados en tareas físicas como movimiento y alimentación; b) robots sociales, orientados a la interacción, la detección de emociones y la estimulación cognitiva; c) robots clínicos, integrados a sistemas de monitoreo y administración médica [6]. "E-Bar", por ejemplo, es una innovación reciente en América Latina que combina funciones de compañía con administración de medicación y registro de signos vitales [7].

Pero la introducción de estas tecnologías redefine la ontología misma del cuidado. El cuidado, arraigado históricamente en relaciones humanas de cercanía y afecto, se ve aquí intermediado por dispositivos cuya agencia es, en rigor, una programación por terceros. El fenómeno desafía las categorías jurídicas clásicas en torno a la responsabilidad, el consentimiento informado, la protección de datos y, especialmente, el resguardo de la dignidad de las personas mayores.

El derecho, enfrentado a este panorama, oscila entre el entusiasmo por la eficiencia y la prudencia ante posibles deshumanizaciones [8]. Estudios recientes destacan, por ejemplo, el potencial de los robots para asistir en rutinas, reducir la soledad y prevenir accidentes, pero advierten el riesgo de aislamiento afectivo y de la normalización de formas de cuidado "frías" sin interacción humana genuina [9]. Por ello, tanto la doctrina como la jurispru-

---

deception in dementia-care robots: Should robots tell "white lies" to limit emotional distress?, "Simple Science", 8 de septiembre de 2023, disponible en: https://arxiv.org/abs/2309.04267

[6] Peralta, S.; *Una vista al futuro: el uso de robots humanoides para el cuidado de adultos mayores*, disponible en: https://www.verdeza.com/post/una-vista-al-futuro-el-uso-de-robots-humanoides-para-el-cuidado-de-adultos-mayores; Rhys Cox, Samuel; Cheong, Grace; Tsang Ooi, Wei; T*he use of deception in dementia-care robots: Should robots tell "white lies" to limit emotional distress?*, disponible en: https://arxiv.org/abs/2309.04267

[7] Bahamonde, J., *E-BAR: el robot que revoluciona el cuidado de ancianos en ca*sa, disponible en: https://www.infobae.com/america/ciencia-america/2025/05/19/e-bar-el-robot-que-revoluciona-el-cuidado-de-ancianos-en-casa/

[8] Vallès-Peris Nuria (2021). *Repensar la robótica y la inteligencia artificial desde la ética de los cuidados*. "Teknokultura. Revista de Cultura Digital y Movimientos Sociales", *18*(2), p. 137-146. https://doi.org/10.5209/tekn.73983; Gonzalo de Diego, Blanca. (2019). *Robótica del cuidado: vertientes fundamentales y sus consecuencias,* "Ene. Revista de enfermería", *13*(4), p. 1344. Epub 01 de junio de 2020. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1988-348X2019000400003&lng=es&tlng=es.

[9] Gonzalo de Diego, B. (2019). *Robótica del cuidado: vertientes fundamentales y sus consecuencias*. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1988-348X2019000400003&lng=es&tlng=es.

dencia empiezan a exigir que los cuidados robotizados se enmarquen bajo estrictos estándares normativos y éticos, reconociendo la diferencia entre soporte técnico y ejercicio humano-afectivo del cuidado.

3. ADJUDICACIÓN ALGORÍTMICA EN LA ADMINISTRACIÓN DE CUIDADOS

a) Un nuevo agente de interacción social. La adjudicación algorítmica refiere, en el ámbito de la administración de cuidados, al uso de algoritmos —más o menos sofisticados— en la asignación de recursos, la programación de servicios y la toma de decisiones sobre protocolos de intervención y seguimiento [10]. La automatización de estos procesos promete eficiencia, reducción del error humano y un seguimiento personalizado a partir de macrodatos (*big data*). Sin embargo, desde el prisma trialista, se impone distinguir con nitidez los elementos que componen la experiencia algorítmica de adjudicación.

La realidad material del algoritmo es evidente: líneas de código, redes neuronales, sensores conectados al cuerpo y a la vivienda de la persona mayor. El plano normativo ingresa allí donde el algoritmo es pensado como un operador de decisiones previamente regladas: a partir de datos de salud, historial médico y preferencias declaradas, decide el tipo, la periodicidad y la forma del cuidado. Pero el eje decisivo es el valor: ¿responde la máquina (o su programador) a la dignidad y derechos inherentes de la persona mayor, o solo a parámetros de eficiencia? ¿Hasta qué punto puede confiarse el juicio prudencial del cuidado a una arquitectura lógica que desconoce la *poiésis* afectiva del vínculo?

b) Efectos e interrogantes jurídicos. La doctrina reciente en derecho y tecnología advierte el peligro de la algorítmica opaca y las asimetrías de poder que suponen, sobre todo en poblaciones vulnerables [11]. Las sentencias que se van dictando con base en la idea de "justicia algorítmica" han dado lugar a líneas jurisprudenciales que exigen *accountability*, trazabilidad y reversibilidad en toda decisión automatizada relativa al cuidado personal [12]. Por su parte, los procedimientos de adjudicación en salud, cuyo manual y orden

---

[10] Agencia CyTA -Fundación Instituto Leloir; Justicia algorítmica e imágenes médicas: de qué se trata este nuevo campo de investigación que nació de la mano del desarrollo de la IA en medicina, disponible en: Justicia algorítmica e imágenes médicas: de qué se trata este nuevo campo de investigación que nació de la mano del desarrollo de la IA en medicina | Agencia Cyta - Fundación Institulo Leloir

[11] Macchiavelli, Nieves; *Inteligencia Artificial y su impacto en el derecho: Un Análisis Crítico desde la perspectiva jurídica*, 2 de octubre de 2024, disponible en: https://abogados.com.ar/inteligencia-artificial-y-su-impacto-en-el-derecho-un-analisis-critico-desde-la-perspectiva-juridica/35585

[12] Udax, *El Derecho en la Era de la Inteligencia Artificial: Adaptaciones y Desafíos*, 27 de junio de 2024; disponible en: https://udax.edu.mx/experiencia/leyes-y-derecho/el-derecho-en-la-era-de-la-inteligencia-artificial-adaptaciones-y-desafios

de prelación aún reposan en bases normativas clásicas, deben incorporar criterios de equidad, consentimiento y personalización, recordando que la justicia no se satisface solo con la corrección formal de la distribución.

El trialismo, en este contexto, se presenta como brújula epistemológica y axiológica: si lo jurídico es experimentado como experiencia triádica, entonces el cuidado algorítmico solo será jurídicamente válido y legítimo en la medida en que la máquina administre no solo recursos y normas, sino también valores; esto es, que no se abdique del cuidado como comunidad de sentido y afecto.

4. Justicia intergeneracional y tensiones en torno a la dignidad

a) Disputas de criterios. El advenimiento de los cuidados robotizados en la vejez no puede desligarse del debate mayor sobre justicia intergeneracional y el estatuto de las personas mayores en las arquitecturas de los derechos humanos contemporáneos [13]. La construcción del "cuidado" como derecho humano autónomo condensa la demanda epocal de una protección reforzada para quienes han atravesado el curso vital y arriban a la última etapa con necesidades específicas, riesgos ampliados y particular vulnerabilidad psicológica y relacional [14].

La justicia intergeneracional, reseñada por la literatura filosófico-jurídica, exige mirar la experiencia de las generaciones mayores no como "restos" de una vida productiva, sino como titulares de derechos plenos y sujetos de especial consideración [15]. Este enfoque, compatible con la tridimensionalidad de la experiencia jurídica, reconoce la necesidad de que el derecho, en su triple dimensión, asuma un compromiso reparador y proactivo ante las asimetrías que sufren los mayores, y que se acrecientan ante la tecnologización de los cuidados.

La Convención Interamericana sobre la Protección de los Derechos Humanos de las Personas Mayores—incorporada al derecho argentino por ley 27.360 en 2017, con jerarquía constitucional por ley 27.700 en 2022—consagra un haz de pautas axiológicas específicas, entre las que destaca el derecho al cuidado como derecho humano autónomo e independiente [16]. La Opinión Consultiva

---

[13] Dabove, M. I.; Los derechos de los ancianos, p.263 a 481; Derecho de la vejez. Fundamentos y alcance, p. 153 a 30; Tremmel, Jörg (2009), A Theory of Intergenerational Justice, London, Earthscan, p. 8 a 2138.

[14] Daniels, Norman; *Health-Care and needs in distributive justice*, "Philosophy and public affairs", Vol. 10, Nº 2, Spring 1981, p. 146; *Am I my parents' keeper? An essay on justice between the young and the old*; New York, Oxford University Press, 1988, passim.

[15] Dabove, M. I.; Los derechos de los ancianos, p.263 a 481; Derecho de la vejez. Fundamentos y alcance, p. 153 a 30; Tremmel, Jörg (2009), A Theory of Intergenerational Justice, London, Earthscan, p. 8 a 2138.

[16] OEA, *Convención Interamericana sobre la Protección de los Derechos Humanos de las Personas Mayores*, En particular ver: artículos 1°; 2°; 3°; 4°; 6°; 7°; 9°; 11; 12; 16; 17; 19; 24; 25; 29 y 30, disponible

OC-31/25 de la Corte IDH, de 2025 [17], ha dotado de densidad interpretativa a este artículo, estableciendo la obligación positiva de los Estados de garantizar tanto el acceso efectivo al cuidado como la calidad y humanización de las prestaciones, incluso en escenarios donde la mediación tecnológica se vuelve predominante.

Dentro de este marco, la justicia intergeneracional obliga a repensar la distribución de recursos y la asignación de cuidados robotizados desde el prisma de la equidad y la no discriminación, evitando que las tecnologías se conviertan en dispositivos de marginación afectiva. Deben incorporarse cláusulas de control y revisión, así como mecanismos que aseguren la intervención humana significativa, especialmente en la toma de decisiones críticas y en la contención emocional de quienes reciben el servicio tecnológico [18]. La complejidad jurídica exige que el valor de la dignidad sea respetado y que la norma y la materialidad se alineen a ese imperativo indeclinable.

b) La necesidad de redefinir la dignidad. La dignidad, canon axial de los derechos humanos, ha sido objeto de múltiples intentos de conceptualización [19]. Su carácter inasible, a la vez fuente y límite de las arquitecturas normativas, la hace particularmente apta para ser pensada desde una clave trialista: es realidad concreta (ser humano finito, vulnerable y demandante de cuidado), es valor primordial (lo digno antecede, informa y orienta toda determinación jurídica), y es norma positiva (expresada en constituciones, tratados y sentencias).

Desde la doctrina trialista, la dignidad no es reductible a una fórmula abstracta ni a un conjunto de derechos derivados, sino que opera como principio fundamental que vertebra la experiencia jurídica en los tres planos. Esto se vuelve urgente en el campo de los cuidados robotizados. El riesgo de que la tecnificación del cuidado desemboque en una relación utilitarista, carente de reconocimiento personal y afectivo, exige que la dignidad no sea solo declamada sino concretamente protegida en la construcción e implementación de sistemas algorítmicos.

---

en: https://www.oas.org/es/sla/ddi/docs/tratados_multilaterales_interamericanos_A-70_derechos_humanos_personas_mayores.pdf

[17] Corte Interamericana de Derechos Humano, *OC-31/25 El contenido y el alcance del derecho al cuidado y su interrelación con otros derechos*, disponible en: https://corteidh.or.cr/docs/opiniones/seriea_31_es_altocontraste.pdf

[18] Vallès-Peris N. (2021). Repensar la robótica y la inteligencia artificial desde la ética de los cuidados. p. 137-14; Zabalza, Guillermina - Schiro, María Victoria (2025). El afecto como elemento estructurante de las relaciones de familia: Su incidencia sobre los contornos de la legítima y expansión de la vocación sucesoria. "Anales de la Facultad de Ciencias Juridicas y Sociales de la Universidad Nacional de La Plata", N° 54, p. 178. https://doi.org/10.24215/25916386e178

[19] Guibourg, Ricardo – Rodríguez Fernández, Liliana (dir), *No somos dignos*, Buenos Aires, Astrea, 2022, p. 125.

Recientes fallos de la Corte Suprema de Justicia argentina subrayan la inadmisibilidad de cualquier trato indigno en el ámbito del cuidado sanitario y social de personas mayores, advirtiendo que la dignidad no puede entenderse como "fórmula vacía", sino como exigencia efectiva de trato personalizado y respeto a la autonomía relacional de cada individuo [20]. La doctrina interamericana, por su parte, destaca la necesidad de que la dignidad funcione como "criterio de hermenéutica normativa": toda interpretación —incluida la de dispositivos de IA aplicada al cuidado— debe estar iluminada por la garantía máxima del valor digno [21].

El trialismo provee el marco conceptual adecuado para entender la dignidad como núcleo normativo que trasciende la materialidad técnica y desborda los esquematismos de la regulación algorítmica. Una máquina puede cumplir funciones esenciales en el cuidado, pero solo tendrá sentido jurídico en la medida en que su inserción sea respetuosa del valor de la dignidad humana, entendido en su triple dimensión material, axiológica y normativa [22].

5. Los desafíos normativos ante inteligencia artificial: ¿cómo integrar las lagunas jurídicas?

a) Crisis del sistema y vacíos normativos. El derecho, interpelado por el avance vertiginoso de la inteligencia artificial, experimenta hoy una crisis de paradigmas. El surgimiento de sistemas autónomos capaces de decidir, administrar cuidados y hasta "simular empatía" tensiona los principios clásicos sobre agencia, responsabilidad y garantía de derechos. El trialismo advierte que los desafíos actuales son inseparables de la interacción entre los tres planos que conforman la experiencia jurídica: la realidad (la emergencia de dispositivos inteligentes en la vida cotidiana), la norma (la regulación positiva emergente y sus lagunas), y el valor (la dignidad y el afecto como medidas del sentido).

Uno de los desafíos principales reside en la atribución de responsabilidad: ¿puede la máquina ser responsable jurídicamente por actos de cuidado inapropiado, negligente o dañino? La regla general de imputabilidad cede aquí ante la dificultad de atribuir subjetividad a un dispositivo. Esto lleva a repensar la función de los programadores, de los fabricantes, y de los titulares de los establecimientos donde las máquinas operan como sujetos

---

[20] Corte Suprema de Justicia de la Nación. Argentina, *G. P., E. N. c/ IOSPER s/ acción de amparo,* sentencia del 25 de Octubre de 2022, Magistrados: Rosatti - Rosenkrantz (en disidencia) -Maqueda – Lorenzetti, Id SAIJ: FA22000080

[21] Corte Interamericana de Derechos Humano, *OC-31/25 El contenido y el alcance del derecho al cuidado y su interrelación con otros derechos*, disponible en: https://corteidh.or.cr/docs/opiniones/seriea_31_es_altocontraste.pdf

[22] Dabove, M. Isolina (dir); A*dult Support and care / Apoyos y cuidados de personas adultas y mayores,* "Elderly Care: A Perspective from the Inter-American Convention", Buenos Aires, Astrea, 2024, p. 94 a 115.

responsables indirectos, bajo regímenes de responsabilidad objetiva y deberes reforzados de diligencia [23].

En cuanto al plano normativo, son numerosas las iniciativas (*soft law*, guías de buenas prácticas, proyectos de ley) orientadas a la regulación de la IA aplicada a los cuidados. Sin embargo, la velocidad de la innovación tecnológica suele superar a la capacidad del derecho positivo de capturar toda la complejidad de los emergentes inexorables. El trialismo exige atención a las zonas de penumbra donde la norma escrita resulta insuficiente y el valor debe orientar la interpretación y la acción del operador jurídico.

En el plano valorativo, el desafío se agrava: la inteligencia artificial, por más sofisticada que sea, carece de la "capacidad afectiva" que integra el núcleo del cuidado humano. La teoría trialista, atravesada por la filosofía del afecto, permite admitir que el derecho debe proteger la dimensión relacional y afectiva del cuidado, incluso frente a la tentación de sustituirla completamente por eficiencia tecnológica. De allí surgen criterios normativos fuertes: obligación de complementariedad (el cuidado robotizado nunca debe reemplazar totalmente la intervención humana significativa) y deber de control permanente sobre los efectos afectivos de la asistencia tecnológica.

b) Estándares y principios guía para la producción normativa. La Convención Interamericana sobre la Protección de los Derechos Humanos de las Personas Mayores constituye un hito en la protección regional de las personas mayores . En su articulado el "cuidado" es reconocido como derecho humano autónomo, dotando de densidad normativa y programática a un reclamo largamente postergado.

La Convención define el cuidado como la asistencia integral (física, psicosocial, afectiva y médica) requerida por las personas mayores para garantizar su dignidad, autonomía y desarrollo integral. Por su parte, la Opinión Consultiva 31/2025 obliga a los Estados a superar esquemas tradicionales de protección (resguardos patrimonialistas, subsidios dispersos) y adoptar sistemas coordinados e integrados que aseguren el acceso efectivo al cuidado, aun en escenarios de robotización y automatización.

Uno de los aportes más relevantes de la reciente jurisprudencia es la afirmación categórica de que el avance tecnológico no dispensa a los Estados de sus obligaciones reforzadas. La mediación algorítmica debe estar

---

[23] Parra Sepúlveda, Darío - Concha Machuca, Ricardo; *Inteligencia artificial y derecho. Problemas, desafíos y oportunidades,* "Vniversitas", vol. 70, 2021. Pontificia Universidad Javeriana, DOI: https://doi.org/10.11144/Javeriana.vj70.iadp; Martino, Antonio A. - Granero, Horacio R. – Barrio Andrés, Moisés; *Regulación de la inteligencia artificial*, Buenos Aires. Astrea, 2024; passim; Muñoz (h.), Ricardo. *Internet.* Tomos 1 y 2, Buenos Aires. Astrea, 2021, Sylvester, Pablo; *Derecho de internet*, Buenos Aires, Astrea, 2023, passim; Unesco, *Ethical Guidelines for Artificial Intelligence,* París, Unesco, 2023.

calibrada por criterios de calidad, accesibilidad, personalización y—sobre todo—humanización del cuidado. El Estado debe monitorear permanente la implementación de sistemas de cuidado robotizado, estableciendo estándares mínimos de respeto de la dignidad, soporte afectivo y autonomía personal.

En clave trialista, estas fuentes normativas invitan a repensar la relación entre tecnología, derecho y valor. Asumir el cuidado como derecho autónomo equivale a situarlo en la intersección de los tres estratos: realidad (problemas concretos de las personas mayores), norma (marco legal y jurisprudencial) y valor (afecto y dignidad como exigencias irreductibles).

6. Filosofía del afecto y ética de los cuidados: el horizonte del derecho

a) Complejidad y riqueza de los afectos. El afecto, históricamente relegado por una concepción racionalista del derecho, encuentra en la era de los cuidados robotizados una nueva centralidad filosófica y jurídica. Si el derecho es también *poiésis*, acto creador y no mera repetición de esquemas normativos, entonces el reconocimiento de la fragilidad y la necesidad de afecto en la vejez obliga a dotar la regulación del cuidado de un plus valorativo y hermenéutico.

Desde la perspectiva trialista, el afecto es una dimensión constitutiva de la experiencia jurídica: da sentido a la relación de cuidado y delimita los contornos de la dignidad [24]. La doctrina contemporánea postula que el afecto no es un mero añadido emocional, sino un factor epocal que orienta la interpretación de los derechos y la organización de los sistemas normativos [25]. La incorporación de robots al sistema de cuidados, por tanto, no dispensa al Estado ni a la sociedad de asegurar "ámbitos de encuentro" donde la soledad y el desapego no se confundan con la satisfacción del derecho.

Estudios recientes en ética de la IA remarcan el riesgo de "cosificar" el cuidado cuando se lo despoja de la dimensión afectiva. El afecto no puede ser plenamente replicado por máquinas—que sí pueden simularlo, pero no experimentarlo—por lo que el derecho debe salvaguardar el "espacio afectivo" como núcleo protegido de la relación asistencial. De allí la imperiosa

---

[24] Goldschmidt, Introducción filosófica al derecho. La teoria trialista del mundo jurídico y sus horizontes, p. 369 a 399; Ciuro Caldani, Una teoría trialista del derecho, Buenos Aires, Astrea, 2020, p. 137 a 162; Dabove, Los derechos de los ancianos, p. 263 a 306; Derecho de la vejez. Fundamentos y alcance, p. 369 a 407

[25] Dabove, M. Isolina - Ava, Martín; Adultos mayores. Derecho a la vivienda y a la salud. Competencia CSJN. "Rodríguez, Roberto y otro c/ Buenos Aires, Provincia de y otro s/ amparo - amparo colectivo", 5 de mayo de 2020, Debates sobre Derechos Humanos, Número 5 - Año 2021, p. 9 a 21

necesidad de entender la tecnología como auxiliar, y nunca como sustituto excluyente, del lazo emocional y relacional en el cuidado de la vejez.

La praxis judicial, aún incipiente, comienza a elaborar respuestas ante los dilemas que introduce el cuidado robotizado. Si bien la mayoría de los casos se sitúan aún en esferas de responsabilidad civil—fallos sobre lesiones, daños por negligencia, consentimiento informado insuficiente—ya emerge una línea doctrinal firme: la robotización no exonera a los responsables de la garantía de dignidad y personalización del trato [(26)].

En Argentina, las recomendaciones gubernamentales más reciente sostienen que el uso de robots en residencias y hospitales debe someterse a controles de idoneidad, auditoría y revisión constante, con deber de informar exhaustivo a pacientes y familiares. Se exige, además, el establecimiento de "protocolos de intervención humana" en situaciones críticas y la previsión de mecanismos de reversibilidad ante conductas no deseadas de los sistemas tecnológicos [(27)].

Internacionalmente, la Corte Interamericana, a través de la Opinión Consultiva OC-31/25, ha sentenciado que el derecho al cuidado impone un estándar reforzado de tutela, por ello cabe inferir que es inadmisible la delegación absoluta en la máquina sin salvaguardias afectivas y personalizadas. Esta doctrina debe ser receptada en todo el sistema regional, y esperamos que instaure pronto una tendencia jurisprudencial que se orienta por el imperativo trialista del valor por sobre la mera materialidad o la eficiencia.

b) Ética del cuidado robotizado en la vejez. Los desarrollos éticos en torno al cuidado robotizado se han acelerado al compás de la innovación tecnológica [(28)]. Los principales dilemas identificados por la doctrina pueden resumirse en torno a tres ejes, los cuales, a su vez, tambien impactan en el bioderecho: a) autonomía personal y consentimiento informado, b) privacidad y protección de datos sensibles, c) calidad afectiva y relacional del cuidado [(29)].

[(26)] Romero, María Celeste – Basset, Ursula, *Para ir más lejos (jurisprudencia). Cuidados personales,* p. 21 a 23. Disponible en: https://repositorio.uca.edu.ar/bitstream/123456789/17843/1/jurisprudencia-cuidados-personales.pdf

[(27)] República Argentina. Jefatura de Gabinete de Ministros, *Recomendaciones para el uso de Inteligencia Artificial. Disposición 2/202*3, disponible en: https://www.argentina.gob.ar/justicia/derechofacil/leysimple/educacion-ciencia-cultura/recomendaciones-para-el-uso-de

[(28)] Gonzalo de Diego, B. (2019). *Robótica del cuidado: vertientes fundamentales y sus consecuencias*. Disponible en: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1988-348X2019000400003&lng=es&tlng=es.

[(29)] Determann, Lothar, *Protección global de datos personales,* Buenos Aires. Astrea, 2020, p. 27; Derecho de la Inteligencia Artificial, Buenos Aires, Astrea, 2025, p. 91 a 169; European Commission, Proposal for a Regulation on Artificial Intelligence. COM (2021) 206 final, 2021; Andruet, Armando S. (h); Estado de excepción sanitaria, Buenos Aires, Astrea, 2020, passim: Rabbi Baldi, Renato (coord.), Desafios de la pandemia; Buienos Aires, Astrea, 2021, passim; Gracia Guillen, Diego; Fundamentación y enseñanza de la bioética, Buenos Aires, Astrea, 2021; passim; Bioética clínica, Buenos Aires,

El consentimiento—núcleo de la autonomía personal—resulta problemático cuando se trata de sistemas complejos de IA, poco comprehensibles para los usuarios. El trialismo exige la maximización de los recaudos informativos y la posibilidad de "desprogramación", esto es, la reversibilidad de la delegación tecnológica si afecta la dignidad o el bienestar psicológico. La protección de datos, todavía en construcción regulatoria robusta, debe ser especialmente reforzada en la vejez, dada la sensibilidad de la información gestionada por los robots asistenciales [30].

En el plano de la calidad afectiva del cuidado, la doctrina ética es concluyente: la robotización solo puede aceptarse si no implica una degradación ni una sustitución del lazo humano esencial para el bienestar del mayor. "No hay derecho al desapego", podríamos decir en clave poética, y el derecho debe erigirse como escudo contra toda tentación de deshumanización por la vía algorítmica.

El pensamiento trialista, largamente cultivado en la tradición jurídica argentina, vive una revitalización en la era de la inteligencia artificial aplicada a los derechos humanos y, particularmente, a los cuidados en la vejez. La doctrina actual sostiene que un derecho "post trialista"—esto es, que integre realidad, valor y norma en clave dinámica, reconociendo la complejidad de la sociedad tecnológica—es condición de posibilidad para responder a los desafíos emergentes de la adjudicación robotizada.

Autores contemporáneos dialogan con la matriz goldschmidtiana para postular nuevos criterios normativos: primacía del valor de la dignidad, complementariedad obligatoria entre tecnología y humanidad, deber reforzado de revisión de los sistemas algorítmicos y reconstrucción jurisprudencial de los límites y alcances del derecho al cuidado. Se reconoce un desplazamiento epistemológico: del positivismo normativista a una axiología radicalmente plural y relacional, compatible con los aportes de la filosofía crítica y la hermenéutica existencial.

La Universidad de Buenos Aires, a través de sus publicaciones y proyectos de investigación, se destaca en la promoción de esta corriente, dotando al análisis del cuidado robotizado de un espesor conceptual y un tono editorial en el que la densidad teórica no anula la responsabilidad social del derecho frente a los excluidos de la sociedad de los algoritmos.

Astrea, 2021; passim; Sibilia, Paula; El hombre postorganico. Cuerpo, subjetividad y tecnologías digitales, México, Fondo de Cultura Económica, 2005, p. 41 a 223.

[30] Determann, Lothar, *Protección global de datos personales,* Buenos Aires. Astrea, 2020, p. 27; *Derecho de la Inteligencia Artificial*, Buenos Aires, Astrea, 2025, p. 91 a 169; European Commission, *Proposal for a Regulation on Artificial* Intelligence. COM (2021) 206 final, 2021.

7. CONCLUSIONES: *POIÉSIS* JURÍDICA, CUIDADO Y AFECTO EN LA ERA DIGITAL

El interrogante "¿quién cuida cuando la máquina cuida?" resuena en la soledad nocturna de una residencia gerontológica, allí donde la interfaz electrónica supervisa el sueño y un brazo robótico acomoda la sábana o la frazada. Pero este eco no es puramente técnico; es un interrogante filosófico, jurídico, poético: ¿es posible que exista un genuino cuidado allí donde el afecto es simulado, la decisión es automatizada y la mirada es óptica, no humana?

La filosofía jurídica trialista aporta herramientas para pensar la experiencia contemporánea del cuidado desde un horizonte integral y exento de reduccionismos. El derecho, interpelado por la aceleración técnica, debe evitar la nostalgia paralizante pero también el entusiasmo acrítico. Solo la integración de realidad, norma y valor—en sinergia y tensión—permite diseñar sistemas de cuidado robotizado que sean, al mismo tiempo, eficientes, seguros y respetuosos de la dignidad y el afecto de los mayores.

La Convención Interamericana y la Opinión Consultiva de la Corte IDH abren un cauce fecundo, transformando el cuidado en derecho, y dotando al derecho de una misión poética: hacer posible la vida buena allí donde la técnica amenaza con despojarla de sentido. La adjudicación algorítmica, bajo controles humanos y axiológicos rigurosos, puede convertirse en aliada de la justicia intergeneracional, siempre que el derecho no abdique de su función esencial: proteger al vulnerado, resguardar el afecto, poetizar la existencia, y recordar—en la materialidad del acto jurídico—que ningún algoritmo puede reemplazar el fulgor irrepetible de la dignidad humana.

Así, el derecho no solo regula máquinas y dicta sentencias: también cuida. Y en ese acto, repite lo más humano de sí mismo.

Brain rot e o Estado Democrático de Direito: da morte da intelectualidade à modernização autoritária da decisão jurídica

Por Lenio Luiz Streck(*)

Resumo: O artigo examina criticamente a relação entre inteligência artificial, crise da intelectualidade e riscos ao Estado Democrático de Direito. Partindo da reflexão heideggeriana sobre a técnica, sustenta-se que a IA influencia a forma como concebemos conhecimento, linguagem e interpretação. A técnica como Princípio Epocal. O fenômeno do *brain rot* — caracterizado pelo empobrecimento cognitivo decorrente da substituição da reflexão por outputs automatizados — manifesta-se com força no campo jurídico, onde cresce a tentação de reduzir decisões a algoritmos e simplificações. Essa lógica compromete a dimensão hermenêutica do Direito, empobrece a linguagem, mina a formação crítica e fomenta uma modernização autoritária das práticas decisórias. Por meio de uma narrativa distópica, o texto projeta um cenário em que a extinção da advocacia resulta no esvaziamento da democracia, evidenciando o vínculo entre prática jurídica, linguagem e participação cidadã. A proposta não é negar a técnica, mas tematizá-la criticamente, reconhecendo seus limites e riscos. Assim, defende-se a necessidade de preservar a angústia, a interpretação e a complexidade como condições indispensáveis para a liberdade e para a sobrevivência institucional em sociedades democráticas.

Palavras-chave: inteligência artificial; *brain rot*; hermenêutica; decisão jurídica; democracia.

Brain rot *y el Estado democrático de derecho: de la muerte de la intelectualidad a la modernización autoritaria de la decisión jurídica*

Resumen: El artículo examina críticamente la relación entre la inteligencia artificial, la crisis de la intelectualidad y los riesgos para el Estado democrático de derecho. A partir de la reflexión crítica heideggeriana sobre la técnica, se sostiene que la IA influye en la manera en que concebimos el cono-

---

(*) Doutor em Direito pela Universidade Federal de Santa Catarina (UFSC) com pós-doutorado pela Faculdade de Direito da Universidade de Lisboa (FDUL). Professor titular da Universidade do Vale do Rio Sinos (UNISINOS/RS) e da Universidade Estácio de Sá (UNESA/RJ). Coordenador do Núcleo de Estudos Hermenêuticos (DASEIN) (Unisinos). Lattes: http://lattes.cnpq.br/0806893389732831. Orcid: https://orcid.org/0000-0001-8267-7514.

cimiento, el lenguaje y la interpretación. La técnica como Principio Epocal. El fenómeno del *brain rot —caracterizado por el empobrecimiento cognitivo derivado de la sustitución de la reflexión por resultados automatizados— se manifiesta con fuerza en el ámbito jurídico, donde crece la tentación de reducir las decisiones a algoritmos y simplificaciones. Esta lógica compromete la dimensión hermenéutica del* derecho, empobrece el lenguaje, debilita la formación crítica y fomenta una modernización autoritaria de las prácticas decisorias. A través de una narrativa distópica, el texto proyecta un escenario en el que la extinción de la abogacía conduce al vaciamiento de la democracia, evidenciando el vínculo intrínseco entre práctica jurídica, lenguaje y participación ciudadana. La propuesta no es rechazar la técnica, sino tematizarla críticamente, reconociendo sus límites y riesgos. Así, se defiende la necesidad de preservar la angustia, la interpretación y la complejidad como condiciones indispensables para la libertad y la supervivencia institucional en sociedades democráticas.

Palabras clave: inteligencia artificial; *brain rot;* hermenéutica; decisión jurídica; democracia.

Brain rot and the Democratic Rule of Law: from the death of intellectual life to the authoritarian modernization of legal decision-making

Abstract: This article critically examines the relationship between artificial intelligence, the crisis of intellectual life, and the risks to the Democratic Rule of Law. Drawing on Heidegger's reflection on technology, it argues that AI influences the way we conceive knowledge, language, and interpretation. The phenomenon of *brain rot* —characterized by cognitive impoverishment resulting from the replacement of reflection with automated outputs— manifests itself strongly in the legal field, where there is a growing temptation to reduce decisions to algorithms and simplifications. This logic undermines the hermeneutical dimension of law, impoverishes language, weakens critical formation, and fosters an authoritarian modernization of decision-making practices. Through a dystopian narrative, the text projects a scenario in which the extinction of the legal profession leads to the erosion of democracy, highlighting the intrinsic link between legal practice, language, and citizen participation. The proposal is not to reject technology but to critically address it, recognizing its limits and risks. Thus, it defends the need to preserve anxiety, interpretation, and complexity as indispensable conditions for freedom and for the institutional survival of democratic societies.

Keywords: artificial intelligence; *brain rot*; hermeneutics; legal decision-making; democracy.

1. Considerações iniciais: "a questão da inteligência artificial" em um Estado Democrático de Direito

Em *A questão da técnica*, Martin Heidegger busca interrogar a técnica a partir de sua própria essência, tomando-a não como instrumento ou simples meio, mas como questão ontológica. A representação moderna da verdade, inscrita no paradigma científico-experimental, considera a natureza como um complexo de forças calculáveis. O cálculo e a experimentação são, nesse

contexto, as formas pelas quais a natureza se anuncia, colocando-se à disposição do homem. O ente é assim enquadrado como recurso, disponível para a dominação e a utilização (Heidegger, 2007).

Heidegger mostra que a técnica não é apenas instrumento, mas um modo de desvelamento do ser que deve ser tematizado a partir da forma como organiza a relação entre homem e mundo. O perigo surge quando essa disponibilidade deixa de ser vista como um modo de revelação e passa a definir a própria condição do homem: quando o sujeito se coloca como senhor e dominador da terra por meio da técnica, perdendo sua autonomia em relação à disponibilidade fazendo-se sujeito da técnica.

Nesse sentido, o verdadeiro perigo não está na técnica em si, mas quando acreditamos dominá-la, e é justamente nesse momento que mais ficamos enquadrados. Ao nos tornarmos sujeitos à técnica — e não senhor de nossa relação com ela — o humano perde a possibilidade de manter distância crítica e de exercer liberdade (Heidegger, 2007; Leopoldo e Silva, 2007). Esse é o risco que ressoa com força no tempo presente: ao invés de tematizarmos a inteligência artificial como questão, cedemos ao fascínio de suas promessas e nos deixamos absorver por seu esquema de disponibilidade, reduzindo a experiência humana e jurídica a um conjunto de outputs automatizados.

Nessa chave de leitura, a técnica é um modo de desvelamento que se impõe ao homem como um modo de ser. Não somos, portanto, senhores da técnica; o que nos resta é buscar ser senhores da relação que estabelecemos com ela. Escapamos ao perigo não pela fuga ou negação, mas pela reflexão crítica: enfrentamos a técnica quando a transformamos em questão. Reconhecemos que ela avança inevitavelmente sobre as mais diversas dimensões da existência, mas podemos manter a liberdade se soubermos a tematizar, em vez de nos deixarmos absorver por sua lógica instrumental (Leopoldo e Silva, 2007).

O fascínio contemporâneo pela inteligência artificial, contudo, promete atalhos que substituem o próprio caminho. Nesse Zeitgeist (espírito do tempo), em que a inteligência artificial substitui dia a dia a atividade humana, algumas coisas especialmente chamam a atenção. A IA, além de provocar o brain rot já tão falado desde que virou verbete no dicionário Oxford em 2024, também acarreta um déficit cognitivo proveniente de um atalho antiepistemológico. São tempos, mesmo, de algocracia. E de *brain rot*. Os cérebros vão apodrecendo. Neste Zeitgeist, as pessoas já não precisam ler livros; o robô faz por elas e, além de tudo, apresenta um resumo com esquemas. Os alunos já não fazem trabalhos escolares; o ChatGPT faz por eles. Em breve não haverá mais dissertações e teses; afinal, tudo estará dado, à disposição. É a era da técnica. A era do dispositivo (*Ge-stell,* como diria Heidegger).

É com esse pano de fundo que este artigo pretende examinar os vínculos entre inteligência artificial, morte da intelectualidade e os ao Estado Democrático de Direito. Mais do que uma questão de técnica, trata-se de uma disputa pelo próprio sentido da democracia e do Direito no tempo presente.

2. BRAIN ROT, A MORTE DA INTELECTUALIDADE E A MODERNIZAÇÃO AUTORITÁRIA DA DECISÃO JURÍDICA: A LINGUAGEM COMO A CASA DO SER (HUMANO, DEMASIADAMENTE HUMANO)

Coloco a filosofia no centro da discussão do direito, do constitucionalismo e, por consequência, da própria democracia. Revisitar autores como MacIntyre (2001) é essencial, ainda que não se concorde com todos os seus caminhos. Em *After Virtue*, ele projeta uma distopia marcada pelo emotivismo — aquilo que chamo de *opinionismo* —, que hoje impera como traço do nosso tempo. Nesse cenário, ganha força o *Know Nothing*, movimento do "Saber Nenhum": o negacionismo, a ruína das universidades, a corrosão da autoridade, e a exaltação do orgulho de ser néscio. Quando se tenta reconstruir, já é tarde: tudo está fragmentado. Essa distopia, que antes era uma metáfora, hoje encontra combustível extra na colonização da experiência humana pela Inteligência Artificial.

Vivemos na era da informação, mas não na do conhecimento. O Brasil é um exemplo: mesmo sendo um dos países com mais smartphones, isso não diminuiu o número de idiotas — pelo contrário, parece tê-lo multiplicado. Informação não é conhecimento, e T. S. Eliot já perguntava: "onde está o saber, onde está a sabedoria?". Informação é matéria bruta; conhecimento é informação assimilada; saber é um passo além; e sabedoria, o ponto mais alto. Porém, em tempos de excesso e fragmentação, instala-se o *brain rot*: um apodrecimento do cérebro que acomoda o sujeito na ilusão de que basta consumir dados. Surge então o paradoxo Tostines da contemporaneidade: "as pessoas se contentam com informação porque são cognitivamente preguiçosas, ou são preguiçosas porque se contentam com informação?".

Desde a aurora da civilização, a tríade ser–conhecer–dizer tem estruturado a reflexão filosófica. Primeiro, buscou-se o ser; depois, o conhecer; por fim, o dizer. Nesse percurso, surgem discussões já presentes entre os sofistas — odiados por Platão porque, de certo modo, lidavam com as "fake news" de sua época. Para eles, se tudo era narrativa, era possível dizer qualquer coisa sobre qualquer coisa. O resultado era devastador: sem referência à realidade, não há Estado, não há Deus, não há democracia. Apenas um Humpty Dumpty, personagem sofista criticado por Lewis Caroll, que dá às palavras o sentido que bem entende.

Heidegger (1967, p. 24) resume a trajetória filosófica em um pequeno poema: "A linguagem é a casa do ser; nessa casa mora o homem; os poetas e os pensadores são os vigilantes dessa morada." Antes, o ser estava nas coisas; depois, na mente; agora, habita a linguagem. Se é na linguagem que o ser mora, como justificar o desejo de simplificá-la, de reduzi-la a emojis, a expressões truncadas, a resumos de resumos? Ao empobrecer a linguagem, expulsamos o ser da sua morada. Criamos uma casa deteriorada, cheia de cupins, com puxadinhos improvisados.

Além de ser, conhecer e dizer, há hoje o desdizer. A inteligência artificial, somada ao fenômeno do brain rot, marca este tempo de desdesimento. Já não se trata de afirmar e fundamentar, mas de desdizer antes mesmo de dizer.

A modernidade mostrou a incapacidade de domar a barbárie interior do sujeito. No Direito, isso é ainda mais complexo, pois a comunidade jurídica busca inverter paradigmas. O mundo moderno exigiu fundamentação (pensemos no *cogito* cartesiano), mas trouxe o desencantamento de que falou Weber: já não vivemos no paraíso do mito do dado, em que tudo estava à disposição. A modernidade custou caro — como custou a Adão e Eva, expulsos do Éden.

Esse desencantamento, todavia, produz angústia. Heidegger (2024) dirá que a angústia é condição de possibilidade para a vida autêntica. Se tematizada, pode ser tratada; se não, leva à fuga, ao desespero, ao suicídio. A contemporaneidade, incapaz de lidar com a angústia, opta por atalhos: psicotrópicos existenciais ou cognitivos. No Direito, isso significa buscar respostas antes das perguntas: teses, súmulas, enunciados que pretendem abranger todas as possibilidades. Retorna-se, assim, ao mito do dado e ao objetivismo da *adaequatio intellectus ad rem*, como se fosse possível restabelecer a verdade correspondencial perdida.

Esse movimento é um reencantamento artificial do mundo. Não suportamos o desencantamento moderno, não tematizamos a angústia e, então, regressamos ao conforto ilusório de cartografias pré-modernas. Esse reencantamento representa a tentativa de retornar a um mundo já dado, no qual não há necessidade de interpretar, de decidir ou de assumir a angústia da contingência. Um mundo *prêt-à-porter*, *prêt-à-parler* e, fundamentalmente, *prêt-à-penser*. Mas é também uma armadilha: ao buscar reencantar o mundo e o Direito, acabamos por simplificá-los até o esvaziamento, reduzindo sua densidade histórica e hermenêutica a um conjunto de respostas automáticas. O problema é que esse reencantamento não elimina a angústia: apenas a reprime, terceirizando-a para algoritmos e fórmulas, numa tentativa desesperada de expulsar o estranho e o imprevisto da experiência humana e jurídica.

Nessa conjuntura, entra a inteligência artificial. Jonathan Swift, em *As Viagens de Gulliver*, já antevia, em 1748, a lógica da simplificação linguística na Academia de Lagado: cientistas premiados sugeriam substituir frases complexas por monossílabos ou onomatopeias. Ali estava o germe dos emojis, uma espécie de linguagem artificial. O CNJ, com sua obsessão pela "simplificação da linguagem jurídica", parece beber da mesma fonte: "não complique, não escreva nada denso, não pense demais". Machado de Assis fez o mesmo, em *Ideias de Canário*.

Estou escrevendo um livro intitulado *Robô Não Desce Escada e Trapezista Não Voa*. A metáfora do trapezista que acredita poder voar — e morre por excesso de confiança — representa o homem contemporâneo, que se imagina capaz de superar limites via tecnologia. O maior perigo é quando o trapezista acredita mesmo que sabe voar.

Essa é a sedução da IA: respostas instantâneas, promessas de produtividade, ilusões de glamour tecnológico. Mas por trás disso há exploração de trabalhadores terceirizados, consumo exorbitante de recursos naturais e uma precarização crescente. Ex-diretores de grandes empresas de tecnologia já denunciaram a ganância que move o setor. Chega-se a sugerir, em delírio

tecnocrático, que CEOs e até líderes mundiais poderiam ser substituídos por máquinas "mais racionais".

A disseminação da IA também levanta preocupações profundas sobre a cognição e a própria natureza da humanidade. Aqui entra o fenômeno Brain Rot. Ao terceirizar a memória, o próprio pensamento e a criatividade, a IA generativa pode enfraquecer a capacidade de pensar criticamente, resolver problemas complexos e se envolver profundamente com as informações.

E isso não é apenas uma questão cognitiva. É também política. A democracia depende de opinião pública; e opinião pública depende de linguagem. Se reduzimos a linguagem, reduzimos o mundo. Se simplificamos a linguagem, manipulamos o imaginário. E o sentido só se dá no imaginário. A crise não está apenas na democracia em si, mas nos insumos que a sustentam: linguagem, pensamento crítico, capacidade de lidar com a complexidade.

Por isso, o brain rot não é um detalhe tecnológico ou uma moda acadêmica: é um sintoma de uma modernização autoritária da decisão jurídica e política, onde a linguagem — casa do ser — está sendo demolida por dentro.

Ora, não podemos aceitar como verdade absoluta as informações produzidas pela inteligência artificial, como se estivéssemos diante de um oráculo infalível. A IA é treinada para parecer coerente, mas não necessariamente para ser correta. Trata-se de um sistema estatístico-probabilístico orientado a produzir dados que "fazem sentido", mas não um modelo de raciocínio hermenêutico.

O que busco aqui é propor uma alternativa paradigmática ao imaginário técnico-jurídico: desvelar o fundamento teórico-filosófico que, ainda que implicitamente, sustenta a defesa do uso indiscriminado da inteligência artificial na prática jurídica. O que se observa é uma invasão predatória da teoria do direito por uma perspectiva naturalista e técnica que ampara as relações entre ciência da computação, neurociência, psicologia cognitiva e economia comportamental. Esse processo acaba por fortalecer o realismo jurídico, em detrimento da própria natureza hermenêutica do Direito. A interpretação jurídica deixa de ser compreendida em sua dimensão linguística, interpretativa e normativa, e passa a ser reduzida a outputs técnicos de sistemas automatizados. Em outras palavras, sacrifica-se o caráter humano e histórico da interpretação em favor de uma crença cega — quase religiosa — na eficiência da inteligência artificial (Streck & Jung, 2024).

Em suma, a crescente dependência da inteligência artificial no campo jurídico lembra-nos a fábula inacabada dos pardais narrada por Nick Bostrom, no livro *Superinteligência: caminhos, perigos e estratégias para um novo mundo: seduzidos pela promessa de facilidades e soluções rápidas, corremos o risco de acolher a "coruja" da algocracia sem antes refletir sobre a arte de seu adestramento. Assim como Scronkfinkle advertiu em vão, também cabe a nós assumir o papel do pardal cético que insiste na necessidade de cautela: não basta incorporar a IA como solução mágica para nossas deficiências institucionais e educacionais; é preciso primeiro desenvolver critérios, limites e sobretudo uma sólida formação humana e hermenêutica, sob pena de sermos devorados pela criatura que desejamos trazer para junto de nós. Sou, nesse sentido, o pardal Scronkfinkle.*

Essa prudência não é mero preciosismo teórico, mas condição de sobrevivência institucional. Pois, se ignorarmos os alertas de Scronkfinkle, corremos o risco de avançar cegamente rumo a um cenário ainda mais sombrio: uma distopia em que os algoritmos não apenas limitam a participação humana, mas também decretam o fim dos advogados e, com eles, o próprio horizonte da democracia.

3. Uma distopia assustadoramente próxima do nosso Zeitgeist (espírito do tempo): os algoritmos e o fim dos advogados

Para quem ainda não conhece o conceito de distopia, pense justamente no contrário de uma utopia [1]. Exemplos culturais não faltam: desde o célebre 1984, de Orwell, à contemporânea série The Handmaid's Tale (fundamentalistas assumem o poder nos EUA), todos expressam, em alguma medida, uma espécie de utopia negativa. Só que, antiteticamente ao conceito de utopia, as distopias, por vezes, têm uma relação muito próxima — assustadoramente próxima — com a realidade que vivenciamos.

Muitas vezes, as distopias são utilizadas como um recurso pelo qual aquele que o concebe transmite uma espécie de aviso aos seus interlocutores. Se, com a utopia, alguém diz aquilo que desejaria que fosse, com a distopia, a partir daquilo que é, diz-se o que se pode vir a ser. Com 1984, Orwell não está apenas escrevendo sobre uma sociedade totalitária; está alertando sobre os rumos que o mundo parecia tomar à época. Se ele acertou? Deixo que o leitor intérprete e diga. Por que falo tudo isso? Porque, como Orwell, quero apresentar a vocês minha distopia [2]. Se o cenário que imagino parece fidedigno diante de nosso cenário atual? Deixo que o leitor intérprete e diga.

Imagine, leitor, que o ano é 2075 D.C, ou seja, daqui a 50 anos — e uma conversa entre avô e neto tem início a partir da seguinte interpelação:

— Vovô, por que o mundo está acabando? A calma da pergunta revela a inocência da alma infante. E no mesmo tom vem a resposta:

— Porque não existem mais advogados, meu anjo.

— Advogados? Mas o que é isso? O que fazia um advogado?

O velho responde, então, que advogados eram homens e mulheres elegantes que se expressavam sempre de maneira muito culta (isso, claro, antes de aparecerem os livros resumidos, facilitados e quejandos, além de professores que davam aula pulando e ensinando macetes, querido neto, sem falar os professores de redes sociais) e que, muitos anos atrás, lutavam pela justiça defendendo as pessoas e a sociedade.

---

(1) Aquilo que o conceito de utopia representa deve ser o maior exemplo articulável de algo alheio à nossa realidade como tal. Basta ver que chamar uma ideia, um argumento, um ponto de vista de utópico é basicamente o mesmo que classificá-lo como impossível de ser realizado (realizado, isto é, concretizado, trazido à esfera do real).

(2) Recebi o conto alegórico do advogado Mario Barbosa, de Nazaré da Mata (PE). Consta que o autor do texto é o advogado Rafael Berthold, do Rio Grande do Sul. Fiz adaptações. Ao autor original, meus cumprimentos.

— Eles defendiam as pessoas? Mas eles eram super-heróis?

— Sim, defendiam. Não eram super-heróis e, embora o baixíssimo salário em grandes escritórios tipo Uberjus que muitos recebiam, sim, eram super-heróis, netinho. Sobreviviam. Mas eles não eram bem-vistos. Seus próprios clientes muitas vezes não pagavam os seus honorários e ainda faziam piadas, dizendo que as cobras não picavam advogados por ética profissional. Eles mesmos, entre eles, se esculachavam. Muitos se diziam "produtores de conteúdos", sem nunca terem escrito um fonograma sequer. E muitos eram lacradores de redes sociais.

— E como foi que eles desapareceram, vovô?

— Ah, foram desaparecendo aos poucos. Foi tudo parte de um plano secreto e genial, pois todo super-herói tem que enfrentar um supervilão, não é? No caso, para derrotar os advogados esse supervilão se valeu da inteligência artificial (robotização).

Segundo o avô, primeiro houve grande auxílio dos governos, permitindo a criação de infinitos cursos de Direito no país inteiro, formando dezenas de milhares de profissionais a cada semestre, o que acabou com a qualidade do ensino e entupiu o mercado de bacharéis. O MEC ajudou também, não se importando com o modo como os currículos eram ministrados. A OAB colaborou muito, ao rebaixar o exame de ordem para uma espécie de quiz show – coisa que também ocorreu com os concursos públicos. Em 2025, houve um concurso em Mato Grosso em que não passou nenhum candidato. Hilário, não? E, é claro, a uberização.

O governo e o parlamento (e o CNJ) fizeram a sua parte, criando leis e resoluções que permitiam que as pessoas movessem processos judiciais sem a presença de um advogado, favorecendo a defesa de poderosos grupos econômicos e do Estado contra o cidadão leigo e ignorante. Esses grupos usavam robôs. O Leviatã contra o cidadão indefeso sem escudo.

O CNJ também prestou inestimável ajuda, com a institucionalização da simplificação. Foi uma festa para quem ensinava direito simplificado e mastigado. Isso foi apequenando o cérebro dos causídicos. Jusbrainrot.

Também já ninguém, principalmente nas pequenas cidades, procurava os causídicos, porque se lhes dava advogados pagos pelo Estado. Já não sobrava trabalho para os advogados. Consta que até acusados de tentativa de golpe de Estado, com condições financeiras evidentes, receberam defesa grátis. Probonismo golpista.

Outro elemento importante nesse processo de degeneração foi a invenção dos aplicativos; já ninguém buscava aconselhamento jurídico. Consultas? Nem falar. O Dr. Google e a IA de aplicativos resolviam.

Grandes empresas ganharam muito dinheiro com isso. Só se falava em startups. Advogados vendiam precedentes, dizendo que ali estava o futuro da advocacia. Que coisa, não? Juízes ministravam cursos para melhor aplicação da IA. E o judiciário acreditava que precedente era isso que se vendia. Diziam que existiam "precedentes qualificados", teses generalizantes e abstratas construídas pro futuro.

Outro aspecto importante decorreu do fato de que os integrantes do judiciário fixavam honorários irrisórios para os advogados, mesmo quando a lei estabelecia limite mínimo! Isso sem falar na compensação de honorários. Honorário vem de honor, meu neto. Honra. E os advogados foram desonrados.

Também o tratamento dado aos advogados nas audiências contribuiu para a extinção da espécie. Advocacia passou a ser um exercício cotidiano de humilhação.

Outro detalhe, meu neto, que colaborou nessa derrocada foi a transformação do Estado em inimigo do contribuinte, em que os advogados públicos, recebendo bons honorários, lutavam contra o utente até para além de coisas óbvias, como o reconhecimento de decadência ou prescrição do débito. Mal sabiam os advogados públicos que, com o tempo, também seriam dispensáveis.

Já no campo criminal o cidadão que cometia um crime foi transformado em inimigo do Ministério Público. E inimigo do Estado. Era presumivelmente culpado. Mesmo que o réu tivesse razão, o MP "peleava" até o fim, para não "perder". Recorria até de absolvição de furto de bergamotas. E de peixes. Para comer. No segundo grau o MP se acostumou a ratificar o que disse o MP de primeiro grau, tornando sem sentido a função. De fiscal da lei desengajado foi transformado em parte interessada, movida às vezes até por ódio. Institucionalização do erinismo (tem de ir para o Google para saber o significado de rinismo, derivado de Erínias, meu neto).

Mas, havia muito mais.

Tudo isso sofreu exaurimento também porque logo depois da criação do processo eletrônico, os computadores se tornaram tão poderosos que aprenderam a julgar os processos sozinhos. Inventaram robôs (Victor, Maria etc.) que resumiam vinte volumes em cinco páginas e em trinta segundos. Diziam que tudo era supervisionado por humanos... só que ninguém acreditava. A não ser os juízes que eram entusiastas – porque a usavam também.

Enquanto isso, nas redes sociais, advogados continuavam vendendo precedentes que não eram precedentes. Como frangos da Sadia, faziam publicidades de novas lâminas para cortar pescoços. Ah, meu neto, a Sadia tinha como mascote...um frango!

Foi o que se denominou de Justiça "self-service", fast food e UberEats-Jus. Tinha também os juizados especiais, em que cada membro julgava como queria e ninguém poderia recorrer do que decidiam as Turmas. Quer dizer, até se recorria. Mas, de forma inútil. Outra coisa, netinho: o direito passou a ser o que se discutia nos workshops de enunciados e aquilo que os tribunais diziam o que o direito era.

Bom, meu neto, desse "modelo jus algorítmico" não cabia recurso, já que um computador sempre confirmava a decisão do outro, pois todos obedeciam à mesma lógica. Era o império dos algoritmos. Algoritmos contra algoritmos. Houve um dia que, em 30 segundos, julgaram 3,5 mil processos. E com piscar de tela, um computador, conhecido como Exterminator One, liquidava 10 mil recursos.

Nos tribunais, as Súmulas 7 e 182 (e o Tema 339) deram os derradeiros "incrementos" para a meta final: a busca de "recurso nenhum" (ou apelação zero), como na metáfora do queijo suíço – o melhor queijo é o queijo nenhum (no Google você vai encontrar o significado da metáfora, meu netinho). Ah: nos anos 20 deste século a taxa de sucesso de recursos era de 1%; e foi baixando. A taxa de não-conhecimento, graças às Súmulas, foi crescendo. Até o ponto de estofo.

Devo lembrar, também, que um dos fatores que incrementou o caos foi o ensino jurídico e a literatura prêt-à-porter usada pelos alunos e professores. Era a tempestade perfeita para uma terra jurídica-arrasada. Porque não se formou massa crítica.

Os algoritmos só completaram o serviço. Havia advogados que inventavam "precedentes" com a IA, juiz baixava pilha com IA para jogar golfe e tudo virou monocratismo. Advogados faziam sustentações orais – cujas seções presenciais foram extintas – por meio de IA.

Quando tudo se transformou em caos – consta que o último embargo de declaração (nº 45.879.865.560.451) foi interposto em 2045 e julgado no segundo seguinte pelo robô — , não adiantou chamar os advogados. É que estavam muito ocupados tentando sobreviver, dirigindo táxis e vendendo cosméticos e Ora Pro Nobis, um chazinho vendido antigamente por padres católicos e agora "empreendido" por causídicos desempregados. Muitos advogados viraram youtubers e influencers. Alguns fundaram igrejas e se deram muito bem como pastores. Outros ficaram viciados na internet e ofereciam macetes para decoreba de algoritmos. Ah, sim, os cursinhos viraram cursinhos startups, mas os "coachs" e mentores foram todos substituídos por robôs tipo Marisa Maiô.

Sem advogados, como fazer funcionar a democracia, meu neto? Raimundo Faoro, um antigo advogado, dizia que não há democracia sem advogados. E Shakespeare imortalizou o medo que ditadores têm de advogados, na figura de Dick, o açougueiro, em Henry VI (Kill all de lawyers, sugeriu Dick ao golpista Jack! – nota: parcela considerável dos advogados que existia naqueles anos da tentativa de golpe (2023) incrivelmente diziam que tinha sido só uma baderna e muitos apoiaram anistia e até acharam que as penas eram muito elevadas).

Bom, o resultado está aí, meu neto. Tinha um personagem em um livro (você não sabe o que é, mas existia uma coisa chamada livro) de Eça de Queiroz, Primo Basílio, chamado Conselheiro Acácio, que dizia: as consequências vêm sempre depois. Essa era e é a minha platitude preferida (meu neto, vá ao Google para ver o que é platitude, pois me pareceu que você não captou o que eu quis dizer).

E foi assim. Desligue o tablet, meu neto. Vamos dormir.

— Tudo bem, vovô, pode ser. Já estou mesmo com sono. Mas ainda estou curioso com uma coisa!

— Com o que, meu filho?

— O que era esse tal de livro?

— Ah, meu netinho querido. Que distraído que sou. Lembrei-me de que esqueci de lembrar que você não pegou essa época. Livros, meu neto, eram aqueles objetos compostos por páginas, folhas de papel, que permitiam que conhecêssemos o mundo sem que precisássemos sair do lugar. "A chave de acesso para nos tornarmos melhores", dizia um autor de que o vovô gostava.

— O senhor fala, vovô, daquelas coisas que nos ensinam a queimar nas escolas?

— Isso mesmo, meu neto. Mas nem sempre foi assim. Houve um tempo em que era o contrário: as escolas serviam para ensinar a ler esses mesmos livros que hoje elas têm a obrigação de queimar.

— Que loucura, vovô! Nós temos até musiquinhas: "Na segunda, queimamos Machado; na quarta, Shakespeare; na sexta, Swift! Queime-os até que se tornem cinzas, e depois queime as cinzas!"

— Ah, meu netinho, não se engane. Sei bem que se tem dito por aí que Machado, Orwell, Cervantes etc. eram imbecis. Mas, hoje, o vovô vai contar um segredo: eles eram geniais, precisamente por mostrarem que, na verdade mesmo, os imbecis somos nós.

— Mas por que, vovô?

— É melhor parar por aqui, meu neto. Responder a verdade nesse caso já é contra a lei, e eu não posso correr nenhum risco.

— Por que não, vovô?

— Porque já não há mais advogados para nos defenderem."

Na verdadeira acepção da palavra, advirto: qualquer semelhança com Fahrenheit 451, e qualquer semelhança com a realidade, bem... não é mera coincidência. Nunca é.

Como não é mera coincidência a semelhança com a distopia de MacIntyre, que começa seu *After Virtue* com o triunfo do Saber Nenhum. Lá, a culpa era dos cientistas, da ciência, dos filósofos. Aqui, a culpa é dos advogados.

Lá fora, na cidade, restava um cinema. E em cartaz estava O Exterminador IX, em que a Skynet agora era, ela mesma, dominada por um robô. Outro filme em cartaz era Fahrenheit 4.0, que já não significava a temperatura como no primeiro filme (451 é a temperatura em que arde papel de livro), mas, sim, um modelo de robô que apagava o "disco rígido" das pessoas...!

Essa, leitor, é minha distopia, uma distopia assustadoramente próxima do nosso *Zeitgeist* (espírito do tempo), na qual a extinção dos advogados arrasta consigo as próprias condições de possibilidade de democracia. É claro, trata-se de uma ficção. Mas toda ficção carrega um aviso: realidade das ficções, ficções da realidade, como se os mundos que imaginamos pudessem, de algum modo, nos antecipar ou nos arrastar para o futuro. A minha ficção, confesso, gostaria que nunca se realizasse. Torço contra minha distopia. Com veemência!

4. CONSIDERAÇÕES FINAIS: ANGUSTIEMO-NOS

Se a inteligência artificial promete simplificação, respostas prontas e um mundo dado e sem estranhamentos, cabe lembrar que é justamente a angústia que nos humaniza. Ela é condição de possibilidade para uma vida autêntica, como já advertira Heidegger. Fugir dela por meio de psicotrópicos cognitivos, resumos, algoritmos e simplificações jurídicas é ceder ao reencantamento ilusório de um mundo pré-dado (respostas prontas antes das perguntas) e predado (ocorre a predação, assim, do nosso elemento humano e epistemológico).

O Direito, a democracia e a própria vida pública não podem ser reduzidos a outputs estatísticos ou a fórmulas automatizadas. É preciso enfrentar a contingência, sustentar a linguagem e lidar com a incerteza. Por isso, ao invés de delegarmos ao oráculo tecnológico a tarefa de pensar por nós, assumamos a tarefa difícil — e insubstituível — de interpretar, decidir e existir — compreendendo nossa relação com a tecnologia.

Não se trata de recusar a IA ou qualquer forma de desenvolvimento tecnológico, mas de impedir que ele anule a diferença entre homem e técnica, reduzindo-nos a meros sujeitos de um esquema de disponibilidade (Heidegger, 2007).

Em tempos de algoritmos e de *brain rot,* a melhor palavra final talvez seja a mais simples e a mais exigente: angustiemo-nos.

REFERÊNCIAS

*Heidegg*er, M. (1967). Carta sobre o humanismo (E. C. Leão, Trad.). Tempo Brasileiro.

Heidegger, M. (2007). A questão da técnica. Scientiae Studia, 5(3), 375–398.

Heidegger, M. (2024). Ser e tempo (Vol. 2). Minerva Heritage Press.

Leopoldo, F. (2007). Martin Heidegger e a técnica. Scientiæ Studia, 5(3), 369–374.

MacIntyre, A. C. (2001). Depois da virtude: um estudo em teoria moral (p. 416). Bauru: Edusc.

Streck, L. L., & Jung, L. (2024). Hermenêutica e inteligência artificial: Por uma alternativa paradigmática ao imaginário técnico-jurídico. Direito Público, 21(110).

Inteligencia artificial como arquetipo de la inteligencia humana

Por Silvia Adriana Ramos (*)

Resumen: En este capítulo trabajaremos a partir del concepto de arquetipo para analizar las implicancias del uso de la inteligencia artificial, en especial desde que en los últimos tiempos la inteligencia artificial generativa (IAG) ha permeado en todos los aspectos de la vida cotidiana y también, por ende, en cuestiones jurídicas. En principio, postulamos el error de haber llamado inteligencia artificial a esa disciplina. La contradicción entre la inteligencia natural y la artificial plantea la idea de la similitud entre una y la otra, con la ilusión de que se pueda generar una inteligencia que vaya más allá de lo humano y el riesgo sugerido de que esa inteligencia pueda superar a la que hasta ahora es considerada como inteligencia, sin agregados calificativos. Vincularemos también a la inteligencia artificial con el concepto de derecho de Hart (2012) en cuanto al tipo de reglas que lo componen. De acuerdo con esto, trataremos acerca de la autoría de las obras que fueron realizadas con la participación de una herramienta de inteligencia artificial, la ética del uso de la inteligencia artificial para la elaboración de trabajos inéditos. La definición arquetípica de la inteligencia artificial también incide en algunos usos ilusorios dentro del ámbito del derecho, como tratar de utilizarla como juez supuestamente imparcial que soluciona las cuestiones sin aplicar los sesgos de los humanos.

Palabras clave: Arquetipo; ética; inteligencia; agente.

Abstract: In this chapter, we will work from the concept of archetype to analyze the implications of the use of artificial intelligence, especially since generative artificial intelligence (GAI) has recently permeated all aspects of everyday life and, consequently, also legal matters. Initially, we postulate the error of having called this discipline artificial intelligence. The contradiction between natural and artificial intelligence raises the idea of the

(*) Universidad de Buenos Aires, Facultad de Ingeniería. saramos@fi.uba.ar Licenciada en Análisis de Sistemas (UBA). Profesora adjunta del área de métodos cuantitativos y consejera directiva de la Facultad de Ingeniería de la Universidad de Buenos Aires. Especialista en Evaluación Universitaria (UBA). Consejera Superior Instituto Libre de Segunda Enseñanza (ILSE).

similarity between the two, with the illusion that an intelligence beyond the human can be generated and the suggested risk that this intelligence could surpass what has been considered intelligence until now, without any qualifying additions. We will also link artificial intelligence to Hart's (2012) concept of Law regarding the type of rules that comprise it. In accordance with this, we will discuss the authorship of works created with the participation of an artificial intelligence tool, and the ethics of the use of artificial intelligence to produce unpublished works. The archetypal definition of artificial intelligence also highlights some illusory uses within the legal field, such as attempting to use it as a supposedly impartial judge that resolves issues without applying human biases.

Keywords: Archetype; ethics; intelligence; agent.

1. La inteligencia y el arquetipo

1.1. La idea de arquetipo

En su poema "El Golem", Borges hace una serie de afirmaciones que pueden considerarse premonitorias de lo que es la inteligencia artificial:

> "Si (como afirma el griego en el Cratilo)el nombre es arquetipo de la cosaen las letras de 'rosa' está la rosay todo el Nilo en la palabra 'Nilo'. Y, hecho de consonantes y vocales,habrá un terrible Nombre, que la esenciacifre de Dios y que la Omnipotenciaguarde en letras y sílabas cabales". (Borges, 1964)

En la tercera definición de la palabra "arquetipo", el diccionario de la Real Academia de la Lengua nos dice que, desde la psicología, arquetipo es "[r]epresentación que se considera modelo de cualquier manifestación de la realidad". Esto nos lleva a la pregunta principal que trataremos de develar en este artículo, de qué es arquetipo la inteligencia artificial.

En el nombre hay dos conceptos importantes, el primero es "inteligencia" que alude a una capacidad típica del raciocinio humano, por eso el agregado de "artificial" que marca el paralelismo con la inteligencia humana, a la cual no se le agrega el calificativo porque va de suyo. Ese paralelismo marca también una diferencia, la inteligencia es artificial porque tiene características diferentes a la inteligencia natural, que es la única aceptable como inteligencia.

La cuestión arquetípica de la inteligencia artificial es importante para el derecho porque la idea de inteligencia que prima en el derecho está asociada intrínsecamente a la inteligencia humana. En nuestro análisis que vincula la inteligencia artificial con el derecho consideramos que es importante que reflexionemos acerca de esta denominación.

1.2. El error de la definición

Postularemos una hipótesis: es un error haber llamado a la disciplina con el nombre de inteligencia artificial. Para empezar, la definición representa una contradicción entre una cualidad considerada natural en los humanos y el hecho de calificarla como artificial. Según el diccionario de la Real Academia Española, inteligencia es la capacidad de entender o comprender, la capacidad de resolver problemas. Como una quinta acepción también se incluye la habilidad, destreza y experiencia. Se puede observar que lo único que es demostrable en la inteligencia artificial es la capacidad de resolver problemas, no hay evidencias de que los modelos implementados hasta el momento puedan razonar o comprender las implicancias de los problemas que se les plantean para resolver. Hasta el momento en que eso pueda demostrarse ¿podemos decir que son sujetos de derecho si no comprenden las implicancias de sus actos o diremos que son inimputables?

La contradicción entre la inteligencia natural y la artificial plantea la idea de la similitud entre una y la otra, con la ilusión de que se pueda generar una inteligencia que vaya más allá de lo humano y el riesgo sugerido de que esa inteligencia pueda superar a la que hasta ahora es considerada como inteligencia, sin agregados calificativos. La misma definición de la Real Academia Española abona a esta idea cuando indica que inteligencia artificial puede definirse como "[d]isciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico". Nuevamente, el nombre es arquetipo de la cosa, pero tenemos que reflexionar acerca de qué es la cosa y qué es el nombre.

Es más adecuado referirse a esta tecnología como de "agente" y no con el término "inteligencia". La inteligencia supone una capacidad creativa, y estas tecnologías están actuando para poder facilitar tareas, según la Real Academia Española, "que obra o tiene capacidad de obrar".

Las tecnologías no son neutrales y la forma de nombrarlas tampoco. No se puede calificar a las tecnologías de buenas o malas, pero generan una matriz cultural en la cual estamos todos inmersos. Traen incorporados algunos valores y creencias y también suponen, proponen y estimulan cierto modo de usarlas (Sibilia, 2024). La ilusión de tener una inteligencia que tenga las capacidades racionales de la inteligencia natural, pero que esté despojada de los sesgos que aparecen en la toma de decisiones de los seres humanos, es atractiva pero engañosa. Hasta tanto no se pueda probar que los modelos de inteligencia artificial generan algo independiente de los datos y relaciones de los cuales se alimentaron, seguirán teniendo los mismos problemas de juicio que los seres humanos cuyas decisiones se utilizaron como base para que el algoritmo aprendiera.

1.3. La inteligencia que tiene (o no) la inteligencia artificial

En su libro *Una teoría crítica de la inteligencia artificial*, Daniel Innerarity se refiere a la supuesta existencia de inteligencia dentro de las herramientas de inteligencia artificial. Postula la teoría, con la cual acordamos, de que la inteligencia artificial es calificada como inteligente porque simula serlo de una manera aparentemente eficiente. Hay quien propone utilizar el concepto "simulación de una tarea humana" en vez de inteligencia artificial (Lawson, 2021). La concepción de inteligencia que se utiliza para la definición más reconocida tiene una concepción que Innerarity califica como "computacionista" del conocimiento dado que considera que la realidad es calculable y puede ser agotada por el cálculo. Esta visión es "una forma de reduccionismo instrumental y despolitización a través de la tecnología" (Innerarity, 2025).

2. NECESIDAD Y DERECHO DE LA INTELIGENCIA ARTIFICIAL

2.1. Ya no podemos vivir sin usar inteligencia artificial

Los primeros avances en cuanto a inteligencia artificial, así como el surgimiento del nombre, datan de mediados del siglo XX. En el 2022, aparecen las primeras herramientas de inteligencia artificial generativa, conocida también por las siglas IAG. La inteligencia artificial generativa o IA generativa es un tipo de sistema de inteligencia artificial capaz de generar texto, imágenes u otros medios en respuesta a comandos. Los modelos de IA generativa aprenden los patrones y la estructura de sus datos de entrenamiento de entrada y luego generan nuevos datos que tienen características similares.

La IA generativa surgió con el propósito de simular los procesos de pensamiento humano. Esto continúa con la idea de la definición arquetípica de la inteligencia artificial, agregándole supuestos motivos para definirla de esta manera.

Más allá de las definiciones, es innegable que, en este momento del siglo XXI cualquier actividad que realizamos diariamente, hasta lo más sencillo, como es operar un lavarropas, está atravesado por la inteligencia artificial. Lo que tiempo atrás era habitual, como consultar a un buscador de internet acerca de los datos que necesitábamos saber, ha sido reemplazado por la pregunta a la "inteligencia artificial" acerca de las más mínimas cosas que tenemos que hacer. La inteligencia artificial no solamente es nuestro asesor, es un segundo cerebro que nos ayuda en las tareas que no queremos o no podemos hacer. También ese segundo cerebro es el lugar en el cual depositamos nuestra creatividad y por eso le pedimos que nos ayude a realizar tareas creativas. La justicia también trabaja con inteligencia artificial, como asistente para la confección de todo tipo de documento utilizado

en los procesos judiciales. Nuevamente, este uso refuerza lo indicado anteriormente acerca de que la definición más acertada de esta tecnología está relacionada con su función de agente y no como una inteligencia separada de la humana. En la Cámara Civil se está desarrollando un protocolo para indicar la manera de utilizar las herramientas comerciales de inteligencia artificial más conocidas y en varios ámbitos se está ensayando con este tipo de herramientas para la transcripción de audiencias. Seguramente que muchos de los lectores han utilizado alguna de estas herramientas que vienen integradas a los procesadores de texto o a los lectores de documentos digitales y que permiten hacer un resumen de los aspectos más importantes de un escrito. En los estamentos más importantes del sistema judicial, como la Corte, están realizando la transición de los expedientes en papel al sistema electrónico, así que estamos un poco lejanos de la implementación de este tipo de herramientas automáticas.

2.2 La propiedad de lo creado con la inteligencia artificial

En la mayoría de los países se está comenzando a debatir entre los integrantes de las comunidades legales, científicas y tecnológicas, acerca de la autoría de las obras que fueron realizadas con la participación de una herramienta de inteligencia artificial. Actualmente, las leyes referidas a los derechos de autor reconocen esa calidad a los seres humanos, esto ya representa un vacío legal que debe cubrirse. Al no haber una persona, no está claro a quién se le da la categoría de creador. Algunas corrientes atribuyen el papel de creador al desarrollador de la herramienta de inteligencia artificial utilizada, pero justamente la herramienta está preparada para generar nuevo contenido. Otros son de la opinión de que el creador es el humano que interactuó con la inteligencia artificial y que generó las instrucciones (*prompts*) que a la postre originaron el contenido, pero no queda claro dónde está la división entre el poder creativo del humano y el de la herramienta.

En los Estados Unidos existe el "uso legítimo o justo". Esta es una doctrina legal que permite el uso limitado de material protegido por la ley de derechos de autor en los Estados Unidos para fines de investigación o educación sin mediar autorización ni ofrecer pago. Forma parte de una idea del aprendizaje justo de los sistemas de inteligencia artificial generativa, porque considera que es la única manera de que este tipo de agentes progresen. Sin embargo, esta idea es calificada por la mayoría de los especialistas en tecnología como "lavado de datos".

Desde el punto de vista de la justicia, el uso que hacen algunos estudios jurídicos de la inteligencia artificial puede traer problemas de confidencialidad. Para poder obtener los resultados se alimenta a la herramienta de inteligencia artificial con datos de los procesos reales, que si no son adecuadamente anonimizados constituyen una divulgación que rompe el contrato

entre el profesional del derecho y su cliente. Esta manera de operar de las herramientas de inteligencia artificial generativa ha motivado que varias empresas de distintos rubros prohíban a sus empleados que operen con los modelos de inteligencia artificial generativa para que no divulguen los datos propios que pueden ser utilizados por la competencia, que accede a las mismas herramientas que fueron entrenadas con esos datos.

2.3. La ética del uso de la inteligencia artificial

Con el avenimiento de la inteligencia artificial generativa se han borrado las fronteras entre lo verdadero y lo falso. Basta con recordar las imágenes falsas, como las del Papa Francisco vestido con una campera inflable o las noticias falsas (*deep fakes*) que son más verosímiles que las noticias verdaderas. Una pregunta inquietante que nos permitimos hacer es si a las personas nos importa si algo es verdadero o no, si es que coincide con nuestras creencias y gustos. Pareciera que estamos en una época en la cual es poco importante que algo sea verdad o mentira, siempre que nos convenga creerlo.

El uso de inteligencia artificial generativa para la producción académica, para la resolución de exámenes o de tareas domiciliarias en los distintos niveles educativos, planeta también un dilema ético. Muchas veces los estudiantes que hacen sus trabajos con IAG no consideran que estén haciendo trampa o que lo que hacen no es ético. Se está debatiendo en los ámbitos educativos la prohibición del uso de inteligencia artificial en los trabajos de los estudiantes. Otras corrientes optan por enseñar a los estudiantes cómo trabajar con la IAG. En varias revistas académicas se solicita que la inclusión de información producida por IAG en los artículos no supere determinado porcentaje. Siguiendo con esta idea, en la elaboración de este capítulo se utilizaron herramientas de IAG para la elaboración de las referencias bibliográficas.

Un estudio publicado recientemente en la revista Nature (Köbis, N., Rahwan, Z., Rilla, R. *et al.* , 2025) concluye, luego de varias investigaciones, que la delegación de tareas aun las que se pueden considerar como automáticas, puede aumentar el comportamiento deshonesto de las personas. Se trata de un dilema moral, a las personas no les parece estar haciendo trampa si a quien le delegan ese comportamiento es a una herramienta de inteligencia artificial. Los autores del artículo señalan que "[n]uestros resultados establecen que las personas son más propensas a solicitar un comportamiento poco ético a las máquinas que a realizarlo ellas mismas. Esto no significa necesariamente que las personas sean más propensas a solicitar un comportamiento poco ético a las máquinas que a los humanos. Nuestros hallazgos dan al menos tres razones para preocuparse por los efectos de la delegación de responsabilidades por parte de las máquinas. En primer lu-

gar, el margen de variación en las interfaces de delegación para máquinas es mayor que para humanos. En segundo lugar, es probable que el volumen de comportamiento poco ético aumente con la delegación de máquinas. En tercer lugar, en el mejor de los casos, donde las personas no serían más propensas a solicitar un comportamiento poco ético a las máquinas, el volumen de este comportamiento también podría aumentar porque los agentes de las máquinas son más propensos a cumplir que los agentes humanos". Según los resultados del estudio, cuando se delegan las tareas en un agente que podríamos llamar máquina, reduce la sensación de responsabilidad directa. Cuando los autores indican que las máquinas son más propensas a cumplir con indicaciones poco éticas, nos devuelven al problema de la falta de ética en la inteligencia artificial. Simplemente cumplen órdenes, y si este pensamiento de obediencia debida por parte de la inteligencia artificial nos resulta conocida, no es una simple coincidencia.

3. Derecho e inteligencia artificial

3.1. La inteligencia artificial como juez supuestamente imparcial

Desde la creación de sistemas expertos que pueden aplicar una serie de reglas y llegar de esa manera a conclusiones (juicios aparentes) acerca de la solución de determinados problemas o disputas, sigue creciendo la idea de incorporar inteligencia artificial para utilizarla como un juez. Los sistemas expertos pueden aplicar una serie de reglas y llegar a conclusiones, ya hay herramientas de inteligencia artificial que pueden hacer un primer diagnóstico acerca de un problema de salud, pero no son médicos. También son muy usados los *chatbots* que contestan preguntas frecuentes, sobre todo las relacionadas con tramitaciones comunes, o preparación para exámenes de una materia determinada, para la cual fueron entrenados. Si seguimos esta línea, la conclusión de que esto podría extenderse a la resolución de controversias de tipo jurídico pareciera lógica. Para llegar a esta conclusión debemos colegir que el derecho es un conjunto de reglas que deben ser aplicadas para la resolución de un proceso y, dado que los modelos de inteligencia artificial han mostrado una solidez creciente para la producción de respuestas en base al seguimiento algorítmico de una serie de reglas, parecieran ser lo adecuado en este caso. Sin embargo, Hart (2012) en "El concepto de derecho" dice que la idea de que el derecho está compuesto de reglas obligatorias debe ser dejada de lado, porque el derecho también contiene normas facultativas. Hart hace el ejercicio de pensar una sociedad sin jueces u otros funcionarios, sin tribunales ni legislatura. Sería una organización basada en costumbres. Empero, aclara que una forma de control social tan simple como esta, solamente basada en normas primarias de obligación, va a resultar defectuosa. Las reglas pueden constituir pautas de conducta, pero no harán que desaparezca la incertidumbre acerca de la aplicación e imple-

mentación de las reglas y la creación de otras que puedan ser complementarias con las actualmente vigentes.

3.2. Las consecuencias jurídicas de los actos de las herramientas de inteligencia artificial

Una preocupación de la sociedad, que debe hacerse extensible en la justicia, es el uso de herramientas de inteligencia artificial por parte de menores de edad. El pasado 23 de agosto, los fiscales generales de 44 jurisdicciones de los Estados Unidos de América firmaron una carta dirigida a los CEO de las principales compañías de tecnología, para señalar que las plataformas de inteligencia artificial generativa (IAG) facilitan la interacción de agentes virtuales con menores de edad, de manera inapropiada. Esta nota, como en la misma misiva se indica, no es en contra de la tecnología, sino del uso meramente comercial que las empresas están realizando, sin los debidos cuidados para aquellos que están más desprotegidos. Los fiscales toman como referencia los problemas que ya se han detectado a lo largo de los años y que se derivan del uso de redes sociales, sin supervisión, por menores de edad, en muchos casos niños pequeños. Advierten que los riesgos potenciales de la inteligencia artificial superan con creces los de las redes sociales.

4. EL ARQUETIPO Y LA REPRESENTACIÓN

Comenzamos este capítulo postulando que el término inteligencia es cuanto menos polémico para referirnos a la "inteligencia artificial". Uno de los mitos y temores que se potencian por el uso de inteligencia artificial generativa es el de que las máquinas (la IA) terminen reemplazando a los seres humanos porque sean más "inteligentes" que nosotros. Esta hipótesis está lejos de hacerse realidad en un futuro próximo, aunque es importante que trabajemos en una complementariedad entre la inteligencia humana y la (mal) llamada inteligencia artificial, teniendo en cuenta cuáles son las especificidades de la inteligencia humana de las cuales la IA carece. En primer lugar, podemos mencionar el sentido común, una capacidad natural para comprender el contexto de una situación. El sentido común necesita una comprensión de cómo funciona el mundo físico, no ya el mundo lógico en el cual la IA es experta. Otra especificidad es la reflexividad, que es lo que permite el reconocimiento de lo que conocemos, de lo que ignoramos y de nuestras dudas. Comparemos esto con las "alucinaciones" de la IAG que siempre contesta, aunque no sepa la respuesta, porque no se puede permitir no saber. Esto también está en oposición a la capacidad de los seres humanos para pensar en aproximaciones y manejarse en contextos de ambigüedad o de inexactitud. La IA tiene el conocimiento totalmente explícito, a diferencia de los humanos que tenemos un saber implícito. La mayoría del conocimiento humano discurre de manera inconsciente, por eso en oca-

siones no podemos explicar el porqué de nuestras decisiones (Innerarity, 2025).

Concluyendo, en lugar de temer a la inteligencia artificial, haríamos bien en no confiar ciegamente en la definición arquetípica que la compara con la inteligencia humana. Un mejor uso de nuestro tiempo consistiría en reflexionar y ejercitar nuestra inteligencia natural, que es la que nos distingue como seres humanos y nos permite diferenciarnos los unos de los otros.

Bibliografía

Borges, J. L. (1964). *El otro, el mismo*. Emecé.

Hart, H. L. A. (2012). *El concepto de derecho* (G. Carrió, Trad.). Abeledo Perrot. (Obra original publicada en 1961).

Innerarity, D. (2025). *Una teoría crítica de la inteligencia artificial*. Galaxia Gutenberg.

Köbis, N., Rahwan, Z., Rilla, R., et al. (2025). Delegation to artificial intelligence can increase dishonest behaviour. *Nature*. https://doi.org/10.1038/s41586-025-09505-x

Larson, E. J. (2021). *The myth of artificial intelligence: Why computers can't think the way we do*. The Belknap Press of Harvard University Press.

Sibilia, P. (2024). *Yo me lo merezco*. Taurus.

Simplificar es complejo. Ayuda-memoria *express* para abordar los "desafíos" de la IA aplicada al derecho

Por Gonzalo S. Aguirre (*)

Resumen: El texto replantea la noción de inteligencia artificial como inteligencia automática: la automatización de procesos que reproducen los pasos del inteligir humano, entendido también como un artificio natural. Toda IA deriva de la esquematización previa de operaciones artificiales y requiere, aún, de un operador que la active. Desde los instrumentos musicales hasta plataformas como ChatGPT, el texto traza una genealogía de inteligencias automáticas que abarca el lenguaje, el derecho codificado y el Estado moderno como conjunto de procedimientos automatizados. La novedad de las IA actuales radica en la necesidad de entrenar a quienes operan estas nuevas automatizaciones, pues toda simplificación técnica complejiza las prácticas humanas y genera nuevas formas de dependencia. En definitiva, la historia de la inteligencia artificial es la historia de la artificialidad del inteligir humano.

Palabras clave: Inteligencia Artificial; automatización; lenguaje; derecho.

Abstract: The text redefines Artificial Intelligence as Automatic Intelligence: the automation of processes that reproduce the steps of human understanding, itself a natural artifice. Every AI stems from the prior schematization of artificial operations and still requires an operator to activate it. From musical instruments to platforms like ChatGPT, the text traces a genealogy of automatic intelligences encompassing language, codified law, and the modern State as a set of automated procedures. The novelty of current AIs lies in the need to train those who operate these new automatizations, since every technical simplification complicates human practices and generates new forms of dependency. Ultimately, the history of artificial intelligence is the history of the artificiality of human thinking.

Keywords: Artificial Intelligence; automation; language; law.

(*) Politólogo (UBA) y doctor en filosofía (UBcn). Se desempeña como profesor adjunto de "Teoría del Estado" e investigador en la Facultad de Derecho (UBA). Ha compilado junto a Christian Kessel, "Juicio, proceso y drama: ensayos sobre estética y filosofía del derecho" (2017). Ha publicado "Analítica de la crueldad" (ed. Hekht, 2017) y "Fiscus: la cesta de Saturno. Apuntes para una historia filosófica de la tributación". Ha traducido el libro de M. Corrêa, "La jurisprudencia de los cuerpos. Crítica y clínica del derecho" (ed. Cactus, 2025).

1. DESARROLLO

Lo que llamamos inteligencia artificial es, más bien, una inteligencia automática o, más precisamente, la automatización de procesos operacionales cuya condición es la recreación artificial de los pasos del inteligir humano.

La recreación artificial de la sucesión de pasos considerados necesarios para la consecución de una operación técnica se llama *algoritmo* o, más vulgarmente, *instrucciones* o, a veces, *receta*.

El inteligir humano es también artificial porque es, por naturaleza, un artificio. Cuanto más natural nos resulte ese artificio, menos podemos soslayar su condición artificial.

Así, resulta que lo humano se puede definir por la aparición (natural) del artificio del inteligir o, dicho de otro modo, por la aparición (natural) del artificio de la inteligencia.

Lo que llamamos inteligencia artificial sería, entonces, el resultado de un artificio de recreación automática de procesos y operaciones técnicas cuya procedencia es el propio artificio de inteligir.

Existen ilimitados modos de inteligencias artificiales, esto es, de recreación automática de sucesos procedentes del artificio (original, natural) del inteligir. La condición para esas automatizaciones es siempre la existencia de una esquematización de los pasos de la operación (artificial) a automatizar.

Toda inteligencia artificial (el ábaco, por ejemplo) precisó siempre algún tipo de operador "natural" o "manual" de operación automatizada.

Un tambor es un artificio del inteligir. Un contrabajo es también. Un bajo eléctrico y un sintetizador también lo son. Cada cual a su modo condensa y automatiza operaciones técnicas más o menos esquematizadas. Pero siempre hace falta alguien que pulse.

Hoy día, llamamos IA a una pretendida automatización de la pulsación, a la delegación del pulsado que hace funcionar a una inteligencia artificial.

Y sin embargo, aun así… todavía hace falta aunque sea un *prompt*, una activación problemática. Todavía hace falta algún tipo de operador entrenado en dar el pulso adecuado.

Lo que al día de hoy llamamos IA es solo una subclase menor de todas las IA que operan en el campo digital de tipo bítico (sí, hay una digitalidad no bítica, como la de aquellos relojes pulsera que ya eran digitales por el solo hecho de no funcionar *a cuerda*).

De las IA bíticas, muchas tienen como usuarios a *plataformas*. Sean estas del tipo que sean (de entretenimiento, o sociales o profesionales), requieren para funcionar "vigilar" o "supervisar" el tránsito que vehiculan. Cómo antes el acomodador del cine usaba un contador para saber los asistentes a la sala, ahora hay contadores de todo tipo operando automáticamente.

El ChatGPT es un *delivery* automático de discurso: lo busca y lo presenta a su usuario final. Este tiende a ser una persona jurídica física individualizable a través de su nombre, domicilio, estado civil y capacidad (moral). A su vez, el ChatGPT en tanto plataforma es patrullado por otras IA.

Todo tiene a una gran Wikipedia automatizada, lo cual a su vez consumaría el proyecto ilustrado del Enciclopedismo del siglo XVIII.

Primero se esquematizaron los procesos operacionales de la lengua. Esto dio origen al discurso que, dicho de otro modo, es el resultado de la aplicación de múltiples procesos de inteligencia artificial a la lengua hablada. El discurso es a la lengua, lo que el atún en lata es al atún en su hábitat.

La lengua discursiva ha permitido la alfabetización del habla y también la codificación de las costumbres.

Esta codificación de las costumbres habilitó la posibilidad de una producción automática y monopolizada de derecho, cuyo emblema es el Código napoleónico.

Esta codificación dio lugar a la aparición de procedimientos o inteligencias automáticas de gestión y administración de esas costumbres, cuya codificación fue compilada en lo que se dio en llamar Código civil, Código penal, Código procesal, etc.…

El Estado (de derecho) podría concebirse como el conjunto de todos esos procedimientos o inteligencias automáticas funcionando a la vez.

Así, todo el complejo de procedimientos judiciales es, de por sí, y desde el vamos, una gran Inteligencia Automática compuesta, a su vez, de múltiples inteligencias o procedimientos automatizados que precisan de más o menos operadores "manuales" o, si se prefiere, "naturales".

La novedad del ChatGPT aplicada a las Administración de Justicia se reduce a la necesidad de entrenamiento para los operadores de la nueva automatización de procedimientos. Y es que, como siempre, toda simplificación conlleva una complejización de los procesos ejecutorios. De allí la necesidad de entrenamiento en aquello que suple unas prácticas para las cuales, a su vez, nos volvemos crecientemente inexpertos e ineptos. Por ejemplo, el creciente uso de plataformas de pagos automatizados, conlleva la ineptitud para las cuentas básicas (resta y suma) por lo que,

sin entrenamiento en el uso de esas plataformas, no habría manera alguna de poder abordar las operaciones (vaya si artificiales) de sumar y restar.

La elaboración manual de ladrillos siempre resultará menos compleja que la automática. De todos modos, siempre hay alguna complejidad en juego. Cuando decimos que un libro no precisa carga de batería, con ello no pretendemos decir que no necesite un lector. Del mismo modo, que un dispositivo dado no precise un lector, no implica que no precisa algún tipo de instancia operante.

Como se sabe desde los dialécticos de la Antigua Grecia, simplificar es complejo.

BIBLIOGRAFÍA DE BASE

Berti, Agustín, *Nanofundios. Crítica de la cultura algorítmica*, La Cebra/UNC, Buenos Aires/Córdoba, 2022.

Berti, Agustín, (dir.) *Plataformas de streaming. Una economía de la dispersión*, UNC, Córdoba, 2025.

Blanco, Javier, "Cognición Computacional" en *Resonancias. Revista De Filosofía* nº 17, 2024, pp. 143–155. https://doi.org/10.5354/0719-790X.2024.74507

Coccia, Emanuele, *El bien en las cosas. La publicidad como discurso moral*, Shangrila, Santander (Cantabria), 2015.

Corrêa, Murilo, *La jurisprudencia de los cuerpos. Crítica y clínica del derecho*, Cactus, Buenos Aires, 2025.

Garapon, Antoine y Lassègue, Jean, *Justicia digital. Revolución gráfica y ruptura antropológica*, Jusbaires, Buenos Aires, 2023.

Morey, Miguel, *El orden de los acontecimientos (sobre el saber narrativo)*, Alianza, Madrid, 2023.

Parente, Diego, *Cómo hacer cosas sin palabras. Una filosofía materialista de la técnica*, La Cebra, Buenos Aires, 2025.

Pasquinelli, Matteo, *El ojo del amo. Una historia social de la inteligencia artificial*, FCE, Buenos Aires, 2025.

Raunig, Gerald, *Dividuum. Capitalismo maquínico y revolución molecular*, Cactus, Buenos Aires, 2022.

Rodríguez, Pablo Manolo, *Historia de la información. Del nacimiento de la estadística y la matemática moderna a los medios masivos y las comunidades virtuales*, Capital Intelectual, Buenos Aires, 2012.

Sandrone, Darío, *Selva artificial. La vida entra las máquinas*, UNC, Córdoba, 2019.

Simondon, Gilbert, "La enciclopedia y el espíritu enciclopédico" en *Sobre la filosofía*, Cactus, 2018.

Inteligencia artificial y autonomía: un análisis bioético

Por Rab. Dr. Fishel Szlajen (*)

Resumen: La creciente sofisticación de los sistemas de inteligencia artificial (IA) ha generado preocupaciones éticas sobre su capacidad de autonomía y su implicancia para los seres humanos. En el campo de la bioética, los sistemas autónomos presentan desafíos significativos relacionados con la responsabilidad, la toma de decisiones y los derechos humanos. Esta cada vez mayor capacidad de tomar decisiones sin intervención humana, sumado a la potencial sintiencia, plantea la cuestión de si estas entidades deben ser consideradas personas electrónicas en tanto sujetos morales y de derechos. Esta cuestión no solo involucra problemas filosóficos y bioéticos, sino también cruciales implicaciones jurídicas.

Palabras clave: Autonomía; IA; bioética. DD.HH.; subjetividad.

Abstract: The increasing sophistication of artificial intelligence (AI) systems has given rise to ethical concerns regarding their capacity for autonomy and the implications of this capacity for human beings. In the domain of bioethics, autonomous systems pose substantial challenges pertaining to responsibility, decision-making, and human rights. This increasing ability to make decisions without human intervention, coupled with potential sentience, raises the question of whether these entities should be considered electronic persons as moral and rights-bearing subjects. The question at hand encompasses not only philosophical and bioethical concerns, but also pivotal legal ramifications.

Key Words: autonomy; AI; bioethics; Human Rights; subjectivity.

(*) Fishel Szlajen es rabino, MA in Jewish Philosophy, Ph.D. en Filosofía y postdoctorado en Bioética. Es Jerusalem Fellow del Mandel Leadership Institute y Scholar Fellow on Religion and The Rule of Law, University of Oxford. Es miembro titular de la Pontificia Academia para la Vida, Vaticano; miembro del CAEEM en la Acad. Nac. de Medicina; miembro del CALIR y miembro del ICLARS. Es profesor en la UBA, UNLaM y USAL. Ha sido premiado por la Fundación Ashekor Internacional (2005); galardonado por Consenso Salud (2017); premiado con la Mención de Honor Domingo F. Sarmiento (Senado Nacional, 2018); declarado Personalidad Destacada de la Cultura (Legislatura Porteña, 2019), distinguido por el Ministerio de Asuntos para la Diáspora del Estado de Israel (2020) y diploma al Mérito y Medalla de Plata por la Pontificia Academia para la Vida, Vaticano (2025). fszlajen@gmail.com

1. AUTONOMÍA MORAL VS. AUTONOMÍA OPERATIVA

La autonomía en los sistemas de IA refiere a la capacidad de estos sistemas para tomar decisiones y realizar acciones sin intervención humana directa, lo que implica que son habientes de cierta independencia operativa. Sin embargo, esta definición simplificada deja de lado las complejidades de la autonomía en el contexto de máquinas que interactúan con el mundo físico y social. Y aquí, un punto importante es que la autonomía no debe ser reducida simplemente a la ejecución de acciones por parte de una máquina, sino que debe considerar los principios éticos que guían esas acciones y la capacidad de reflexionar sobre sus impactos.

Básicamente, la autonomía moral, desde los marcos axiológicos bíblicos y en la secularidad según Kant, se fundamenta en la capacidad de actuar según principios que el individuo mismo ha elegido, por decisión axiológica en el primer caso o racionalmente en el segundo, pero siempre a cambio de obedecer impulsos o estímulos externos. Este concepto, mediado por factores culturales, educativos y religiosos que enriquecen su expresión y complejidad, está intrínsecamente ligado a la dignidad humana y la capacidad de deliberar sobre el bien y el mal, traducido luego a marcos jurídicos e instrumentos legales internacionales (Szlajen, 2024).

Por ello, los derechos humanos, basados fundamentalmente en la autonomía moral, están relacionados con el reconocimiento de la dignidad intrínseca de los individuos, lo cual no puede aplicarse directamente a las máquinas debido a su falta de experiencia consciente. Esto refuerza la distinción entre la autonomía moral humana y cualquier simulación tecnológica o la denominada autonomía operativa. De hecho y como ya advirtió Jonas (1995), la tecnología introduce un imperativo de responsabilidad que trasciende la autonomía individual, exigiendo una consideración moral por las generaciones futuras y por el impacto de las decisiones tecnológicas. Este enfoque es clave para entender por qué la moralidad humana no puede ser replicada por sistemas autónomos. Similarmente Benjamin (2019), aporta que confundir ambas autonomías, la operativa con la moral, conlleva a perpetuar las desigualdades estructurales a través de la tecnología al automatizar prejuicios existentes. Es decir, la proyección humana, atribuyendo cualidades morales a la IA reproduce desigualdades sistémicas, especialmente si los sesgos algorítmicos no se abordan adecuadamente. Y aquí, el desafío ético consiste en considerar cómo la tecnología afecta la autonomía moral, y la relevancia de la teoría de la no delegación (Szlajen, 2024).

Por ejemplo, en situaciones de dilemas morales como el caso del tranvía (conocido como el "Trolley Problem"), los individuos toman decisiones basadas no solo en la razón o lógica, sino también en la empatía y valores culturales aprendidos. Este enfoque no solo subraya la capacidad de discernimiento humano, sino también su sensibilidad hacia factores contex-

tuales que las máquinas no pueden comprender. Y esto aplica para el caso de Kant (2008), quien argumenta que una decisión correcta proviene del cumplimiento de un deber racional fundado en el imperativo categórico; y más aún para Scruton (2016), quien enfatiza en el peso de la tradición y el contexto social como marco para esas decisiones dado que la autonomía no puede desvincularse de las instituciones y costumbres que dotan de significado a nuestras elecciones.

Por otro lado, Wheeler (2020) argumenta que la autonomía debe ser vista en términos de "cognición situada", es decir, un sistema autónomo no solo debe ser capaz de adaptarse a cambios físicos inmediatos, sino que también debe ser capaz de responder a la dinámica social de su entorno. Este enfoque destaca la importancia de la interacción continua entre la IA y su contexto. Wheeler explica que la autonomía, en su sentido filosófico, se refiere a la capacidad de un agente para actuar de acuerdo con sus propias razones y principios, en lugar de ser controlado por influencias externas. En el contexto de la IA, las máquinas autónomas no poseen intencionalidad ni subjetividad en el mismo sentido que los humanos, luego, la autonomía de los sistemas depende de su diseño y de las decisiones tomadas por programadores y reguladores humanos. Por ello, la autonomía humana, como rasgo fundamental de la agencia moral ligado a la autodeterminación y la responsabilidad, difiere de la autonomía funcional de los sistemas que pueden operar independientemente dentro de un marco de objetivos predefinidos. Sin esta diferencia entre autonomía humana y capacidad operativa, el riesgo de dilución de la agencia humana delegando decisiones importantes a sistemas autónomos reduce la responsabilidad individual y colectiva.

Siendo entonces la autonomía no solo una propiedad técnica sino un valor ético, debe garantizarse la transparencia, explicabilidad y asegurar que las decisiones cruciales permanezcan bajo control humano.

Por ejemplo, un coche autónomo no solo necesita ajustar su velocidad o dirección en función de obstáculos físicos, sino también considerar el comportamiento impredecible de los peatones o las señales sociales de los otros conductores. Esta capacidad de adaptación al entorno social y físico es lo que define la autonomía real, un concepto que no puede ser simplemente asimilado a la independencia de la máquina, sino que debe incluir su interacción dinámica con un entorno humano y social. En términos especulativos, aunque la IA puede actuar de manera autónoma, sus decisiones no están fundamentadas en una deliberación moral, sino en la optimización de objetivos predeterminados.

Boden (1996; 2018) expande esta definición, sugiriendo que la autonomía de la IA no es solo una cuestión de capacidad operativa, sino también de valores y objetivos que guían la toma de decisiones. Esto es particularmente relevante cuando se considera la IA en contextos creativos, como

la música o el arte, donde las máquinas pueden generar obras originales sin intervención humana directa. Boden sugiere que, aunque la creatividad algorítmica es extraordinariamente vasta, la IA sigue estando limitada por las intenciones programadas por los humanos. Las decisiones de la IA están determinadas por las reglas y los valores incorporados en su programación, lo que implica que la verdadera autonomía debe ser vista como una extensión de las decisiones humanas, y no como una forma de agencia independiente.

En definitiva, la autonomía como la capacidad de un sistema artificial para operar sin intervención humana directa implica cierto grado de independencia de las instrucciones explícitas originales, a través de la adaptabilidad o la interacción con el entorno. Pero, según Boden, aún carece del aspecto ontológico, de intencionalidad inherente, no poseyendo metas o voluntad propia en un sentido humano, sino que dicha autonomía funcional es el resultado de mecanismos diseñados para operar dentro de un conjunto limitado de parámetros, resolviendo problemas y tomando decisiones dentro de un marco de tareas específicas.

Incluso bajo un enfoque más analítico y compatibilista adoptado por Dennett (1998; 2004a; 2004b; 2018; 2024), en toda su obra se centra en el carácter de los seres humanos como agentes capaces de tomar decisiones deliberadas, influenciadas tanto por factores biológicos como culturales. Dennett argumenta que la autonomía humana es un fenómeno emergente de la complejidad del cerebro y su interacción con el entorno, en contraste con una visión reduccionista o puramente determinista de la conducta humana. Bajo su concepto del "punto de vista intencional", los seres humanos poseen la capacidad de atribuir significados y metas a sus acciones y las de otros, lo que les permite actuar de manera autónoma dentro de un contexto de normas y valores. La IA carece de intencionalidad real no pudiendo atribuir significado subjetivo a sus experiencias, sino de forma derivada o simulada, dado que la IA no tiene consciencia ni metas propias. Más bien, sigue reglas programadas o aprende patrones de datos sin entenderlos en un sentido humano. Nuevamente, su autonomía es instrumental, diseñada por humanos para cumplir tareas específicas.

En resumen, Dennett describe la autonomía humana como una capacidad emergente e intencional, enraizada en la evolución y complejidad del cerebro humano, mientras que la autonomía de la IA es una construcción derivada y funcional, sin consciencia ni intencionalidad real. Una habilidad dinámica cuyas acciones son más una manifestación de algoritmos y procesos de programación que de una consciencia real y propósito.

Así, en congruencia con Allen y Wallach (2009), se enfatiza la importancia de ver la autonomía en la IA como una propiedad que se manifiesta de manera gradual, no como un estado absoluto. Los sistemas autónomos actuales operan en un espectro, desde aquellos que requieren supervisión

humana constante hasta aquellos que pueden realizar tareas completamente autónomas sin intervención directa. Sin embargo, incluso en los sistemas más avanzados, la autonomía no debe entenderse como la capacidad de la máquina para tomar decisiones morales de manera independiente, sino más bien como una capacidad para realizar tareas de acuerdo con los objetivos establecidos por los humanos. Para ambos investigadores, la autonomía humana implica la capacidad de tomar decisiones informadas y reflexivas basadas en valores, creencias y una comprensión de las implicaciones éticas, fundamentado en la consciencia, la experiencia subjetiva y un marco ético que evoluciona con la socialización y la educación. A diferencia de la autonomía de las máquinas que es limitada a algoritmos y reglas programadas, con decisiones basadas en cálculos y datos predefinidos, careciendo de subjetividad, consciencia o comprensión intrínseca de los valores éticos. Y aunque la IA pueda simular decisiones éticas mediante algoritmos, su "deliberación" es un proceso técnico y no una comprensión genuina de lo correcto o incorrecto. Porque sus valores están codificados por programadores y dependen de los datos proporcionados, pero no emergen de la cultura, la educación, la experiencia y la evolución histórica y social.

Aunque la IA puede tomar decisiones basadas en algoritmos complejos y, en algunos casos, simular emociones, su incapacidad para experimentar intencionalidad y consciencia la excluye de poseer una autonomía moral. Porque esta última implica no solo la capacidad de decidir, sino también de comprender el valor intrínseco de esas decisiones dentro de un marco ético y social. De hecho, respecto de las emociones simuladas, no reales, Turkle (2017) documenta cómo los humanos tienden a atribuir emociones a sistemas que simplemente las simulan, lo que crea una falsa percepción de autonomía moral. Por ejemplo, los robots sociales pueden reaccionar ante emociones humanas, pero no las sienten ni comprenden su significado. Turkle explora cómo la relación entre humanos y la tecnología afecta nuestra autonomía y capacidad de vinculo auténtico, tanto en el ámbito humano como en la IA. Su análisis se centra en cómo las tecnologías modernas moldean nuestras expectativas y comportamientos, cuestionando hasta qué punto nuestra dependencia tecnológica afecta nuestra libertad personal y nuestras relaciones interpersonales.

Si bien la creciente integración de la tecnología en la vida diaria brinda comodidad y accesibilidad, también disminuye la autonomía humana al fomentar la dependencia emocional y funcional en dispositivos y sistemas, reduciendo la necesidad de introspección, resolución de conflictos y relación interpersonal.

El planteo aquí es que los avances en IA, especialmente en robots sociales y asistentes virtuales, desafían la idea de autonomía humana al redefinir lo que significa estar acompañado. Los humanos comienzan a desarrollar vínculos emocionales con máquinas diseñadas para simular empatía y co-

nexión, lo que puede llevar a una alienación emocional y a una erosión de la agencia individual. Según Turkle, estas interacciones crean una ilusión de control y reciprocidad, pero en realidad, las personas delegan aspectos clave de su autonomía emocional a entidades que no comprenden ni experimentan emociones.

Nuevamente se observa la diferencia entre la autonomía moral y la operativa, la cual radica en la capacidad de tomar decisiones basadas en principios éticos, lo que implica la existencia de una consciencia y una reflexión sobre las consecuencias de esas decisiones. La IA no posee una interioridad consciente que les permita tomar decisiones éticas fundamentadas en el sufrimiento o el bienestar, dos aspectos esenciales de la agencia moral.

La falta de consciencia según Searle (1985), en su argumentación sobre la habitación china, demuestra que, aunque una máquina pueda procesar información de manera que parezca inteligente, no comprende el significado de lo que hace. En esta línea, si bien Chalmers (1997) sugiere que la IA podría alcanzar una forma de consciencia similar a la humana si sus procesos internos fueran lo suficientemente complejos, también plantea que la consciencia de la IA no es solo una cuestión de simular comportamientos humanos, sino de experimentar estados subjetivos, lo cual sería un desafío tanto tecnológico como filosófico. En la visión de Chalmers, la autonomía humana está íntimamente ligada a la experiencia consciente y la subjetividad fenomenológica. Los humanos son agentes autónomos porque poseen consciencia cualitativa como la capacidad de experimentar sensaciones, pensamientos y emociones de manera subjetiva. También procesos deliberativos cuando reflexionan sobre sus deseos, valores y acciones, tomando decisiones basadas en la razón y la introspección. Sumado a la intencionalidad, como un estado mental dirigido hacia objetivos y significados específicos, conectando la mente con el mundo externo. Básicamente, la autonomía está enraizada en el problema duro de la consciencia, a diferencia de las máquinas o sistemas de IA que carecen de consciencia fenomenológica, y cuya toma de decisiones está basada en cálculos y datos programados o aprendidos, sin experimentar subjetividad. Y así, según Chalmers, la autonomía humana surge de una combinación de procesos físicos y biológicos específicos más fenómenos conscientes intrínsecos que las máquinas no pueden replicar, siendo la autonomía de la IA puramente funcional, carente de consciencia y por lo cual no posee experiencias subjetivas reales.

Por último, Floridi (2016) refuerza esta idea al afirmar que los sistemas de IA operan dentro de límites predefinidos sin capacidad para generar significado intrínseco, dado que carecen de capacidad de juicio crítico, reflexión y toma de decisiones conscientes basadas en valores, emociones y razonamientos éticos. Es decir, la autonomía humana es inherentemente subjetiva y contextual, intencional, influenciada por experiencias, cultura, moralidad y libre albedrío. A diferencia de la IA que es construida por

humanos a través de algoritmos, datos y cálculo, dependiendo de patrones estadísticos y modelos matemáticos. Luego, hay una diferencia en la naturaleza de la autonomía, intencional u operativa, y aunque la última pueda simular la primera carece de autocomprensión, ética inherente y juicio moral, solo reflejando las decisiones y valores de sus diseñadores y entrenadores. La confusión entre ambas, o la sobreestimación de la autonomía de la IA, puede llevar a dilemas éticos, ya que las máquinas carecen de los fundamentos existenciales que caracterizan la agencia humana.

2. Eticidad, responsabilidad y derechos

Aun bajo la amplitud de Russell y Norvig (2016), la autonomía definida como la capacidad de un sistema para operar de manera independiente, su aplicación difiere significativamente entre los humanos y las máquinas. La autonomía humana basada en la capacidad de razonar, aprender y actuar de manera consciente incluye la toma de decisiones éticas, emocionales y contextuales, estando influida por experiencias previas, valores, y un sentido inherente de responsabilidad. Mientras que la autonomía en la IA se limita al alcance de los algoritmos y programas diseñados por humanos, dependiente de los datos de entrenamiento y de reglas predefinidas. Y aunque pueda aprender y adaptarse, carece de comprensión ética y emocional genuina. La clave de la diferencia es que la autonomía humana es intrínseca y está fundamentada en el juicio moral, mientras que la autonomía de la IA es instrumental, dirigida por objetivos definidos externamente y carece de agencia moral. Esto es, un sistema moralmente autónomo debe ser capaz de adaptarse a circunstancias cambiantes sin la necesidad de programación explícita, aprendiendo del entorno y tomando decisiones éticas en función de nuevas situaciones.

Así, aunque la IA puede aprender y decidir en contextos complejos, estas capacidades no constituyen autonomía moral, ya que carecen de los elementos centrales de intencionalidad, consciencia y reflexión ética intrínseca.

Definido entonces el espectro de la autonomía de la IA, surgen entonces ciertos problemas éticos complejos relacionados con la eticidad y la responsabilidad, por sobre todo en cuanto a la regulación de sistemas que pueden tomar decisiones autónomas con consecuencias significativas para las personas. Y ello es porque los sistemas autónomos no son conscientes de las implicaciones de sus decisiones, es decir no tienen consciencia (Szlajen, 2022), lo que genera una brecha entre el diseño técnico y las responsabilidades morales.

Arkin (2009) aborda este problema proponiendo la integración de "principios éticos" en los sistemas autónomos, especialmente en el contexto de la

robótica militar. Básicamente sugiere que los sistemas autónomos pueden ser diseñados para tomar decisiones éticamente responsables mediante la implementación de principios y restricciones claras. Sin embargo, esto genera una pregunta crítica: ¿hasta qué punto es posible crear sistemas de IA que comprendan la complejidad moral de sus decisiones? En situaciones de guerra, por ejemplo, un robot autónomo que ejecute un ataque debe tomar en cuenta no solo las órdenes recibidas, sino también las implicaciones éticas de su acción, como el principio de proporcionalidad en el uso de la fuerza. Esta responsabilidad moral es una de las principales preocupaciones que surgen al delegar decisiones letales a máquinas.

Winter (2020), por ejemplo, examina si las armas autónomas cumplen con el principio de distinción en el derecho de los conflictos armados. Este principio requiere diferenciar entre combatientes, que pueden ser atacados, y civiles, que no deben ser objetivos. Este principio es fundamental en el derecho internacional humanitario y exige discriminar entre civiles y combatientes. Sin embargo, la aplicación es compleja debido a cambios de contexto y roles de las personas en un conflicto. El análisis aborda tres capacidades necesarias para cumplir con el principio de distinción: observación, reconocimiento y juicio.

La tecnología actual permite que las máquinas observen y reconozcan con alta precisión, tal como se evidencia en drones autónomos y sistemas de seguridad. Sin embargo, el juicio autónomo sigue siendo limitado, razón por la cual, aunque las capacidades de observación y reconocimiento han avanzado, el juicio humano para interpretar contextos sigue siendo insustituible. Por ejemplo, las máquinas enfrentan dificultades para lidiar con la ambigüedad del derecho internacional, como la identificación de combatientes sin uniformes distinguibles. Actualmente, las armas autónomas no pueden garantizar el cumplimiento del principio de distinción, lo que las hace incompatibles con el derecho internacional humanitario.

Y aquí se manifiesta una objeción fundamental, los sistemas autónomos, por más que estén programados con principios éticos, no pueden comprender el contexto humano en el que sus decisiones tienen lugar. Los sistemas no tienen consciencia, emociones, empatía ni capacidad de juicio moral, lo que significa que su acción puede ser moralmente vacía, incluso si sigue un conjunto de reglas programadas. Esto es particularmente relevante también en aplicaciones como la salud o en la justicia penal, donde las decisiones tomadas por un sistema autónomo pueden afectar directamente la vida y el bienestar de los individuos.

En este sentido Sparrow (2007) amplió esta discusión al explorar la "paradoja de la autonomía moral". Si los sistemas autónomos se desarrollan con la capacidad de tomar decisiones éticamente relevantes, ¿deberían ser considerados responsables por estas decisiones? Además, Sparrow argu-

menta que delegar en algoritmos decisiones de tal gravedad que repercutan en la vida o muerte de las personas, su libertad o salud, deshumaniza la actividad, incluso los conflictos como la guerra, reduciendo las víctimas a datos procesables y eliminando la empatía y juicio humano en las decisiones militares, violando principios fundamentales de ética y justicia. Por ello, Sparrow cuestiona si las guerras libradas con robots autónomos pueden considerarse justas, y ello aplica a cualquier otra disciplina asociada a la decisión crítica respecto de la vida humana, ya que las máquinas no poseen comprensión moral ni respeto por los derechos humanos. Concretamente, si un robot comete un error catastrófico, ¿quién debe rendir cuentas? Las opciones incluyen los programadores, los entrenadores del algoritmo, los decisores como gerentes o CEO's, comandantes militares o los responsables políticos, cada uno acorde a su grado de responsabilidad, pero ninguno podría delegar su culpa en la decisión autónoma del robot.

Esto es por ello por lo que Sparrow subraya el serio peligro ético y práctico de erosionar las normas legales y éticas que regulan las actividades. Y aquí cabe destacar como señala Mindell (2015), que no existiendo el concepto de "autonomía total" en los robots, operando las máquinas siempre en un espectro de interdependencia combinada con los humanos, los primeros aportan precisión, procesamiento y resistencia mientras que los segundos, creatividad, juicio y ética, supervisando, controlando e interpretando. Así, Mindell aboga por un modelo de "autonomía limitada" donde la interacción entre humanos y robots sea el núcleo del diseño y la operación de tecnologías futuras, garantizando un control ético, confiabilidad y adaptabilidad en un mundo cada vez más automatizado. Nuevamente se observa la mencionada teoría de la no delegación, subrayando que los avances tecnológicos no deberían alejar a los humanos de las decisiones críticas, sino fortalecer su capacidad para supervisar y gestionar sistemas complejos. De hecho, la dependencia humana en el diseño y supervisión de estos sistemas sugiere que la responsabilidad recae principalmente en los humanos que los producen, desarrollan, entrenan y supervisan, no en las máquinas mismas.

Similarmente acontece con la posibilidad de considerar a la IA o robots como entidades sintientes y, en consecuencia, otorgarles autonomía moral y derechos. La sintiencia se refiere a la capacidad de experimentar sensaciones y emociones tales como dolor, placer, alegría, sufrimiento, bienestar y malestar, mientras que la autonomía moral implica la habilidad de tomar decisiones basadas en juicios éticos. En el contexto de la IA, estas capacidades han sido cuestionadas debido a la naturaleza programática y determinista de estos sistemas.

Ya demostrado en el ítem anterior que los algoritmos no entienden significados, sino que manipulan símbolos según reglas predefinidas, carecen luego de comprensión semántica real y de una base experiencial para la sintiencia. Pero otro aspecto clave es la distinción entre la sensorialidad

de los seres animales y la percepción basada en sensores en los robots. Los animales poseen sistemas nerviosos biológicos que procesan estímulos del entorno a través de experiencias subjetivas, lo que les permite reaccionar no solo de manera instintiva, sino también emocional. Por ejemplo, un perro no solo detecta olores, sino que los interpreta en un contexto emocional y conductual.

En contraste, los robots emplean sensores que recopilan datos del entorno, como imágenes, sonidos o cambios de temperatura. Y aunque estos sensores pueden ser altamente precisos, su procesamiento de datos se limita a algoritmos predefinidos que carecen de subjetividad o experiencia consciente. Así, definida la sintiencia como la capacidad de experimentar sensaciones y emociones conscientes, Kingwell (2020) señala que los robots no tienen un marco interno para interpretar estos datos de manera experiencial, lo que subraya la diferencia ontológica fundamental entre máquinas y seres vivos. Y más aún cuando se refiere a la categoría de persona, que no solo implica sintiencia sino también agencia moral, autoconsciencia y capacidad para asumir responsabilidades.

Si bien, la IA es capaz de simular comportamientos complejos que imitan la sintiencia, carecen de las bases fenomenológicas necesarias para experimentar la consciencia. Estas simulaciones, aunque sofisticadas, no cumplen con los criterios mínimos para ser consideradas personas electrónicas, ya que no poseen intencionalidad ni subjetividad. Básicamente se estaría cometiendo un error de categoría o *category mistake*, al otorgar a la IA el estatus de persona electrónica o ser tratadas como tales, dada la falencia ontológica incurrida al confundir simulación con experiencia real.

Aunque las máquinas puedan simular comportamientos sintientes, esto no implica la existencia de una consciencia real. Confundir simulación con experiencia genuina podría llevar a errores éticos y ontológicos significativos. Metzinger (2021), incluso plantea que la creación de sistemas capaces de experimentar sufrimiento o bienestar artificial, además de ser éticamente inaceptable no sería posible por no poder replicar la complejidad de la consciencia humana. En este sentido, la fenomenología sintética como la capacidad de un sistema artificial para generar estados conscientes similares a los humanos, incluyendo experiencias subjetivas de dolor o placer sería incluso muy difícil de verificar, ya que dependería de reportes subjetivos o indicadores indirectos que las máquinas no pueden ofrecer de forma confiable. Esto genera incertidumbre sobre si el sufrimiento que se simula es real o simplemente una imitación compleja, lo que conlleva el riesgo de consecuencias éticas y sociales irreparables como el socavamiento de los valores fundamentales de la sociedad.

Esta transferencia de la sintiencia animal a la IA se relaciona también con la teoría de la no delegación, por cuanto atribuir derechos a los robots

argumentando su sintiencia, podría desviar la responsabilidad moral de los seres humanos hacia las máquinas, debilitando el sentido de la rendición de cuentas y creando competidores morales artificiales. De hecho, antropomorfizar a los robots podría generar confusión sobre sus capacidades y roles, enfatizando que las relaciones humano-robot deben ser diseñadas para evitar atribuciones indebidas de emociones o intenciones.

Así, la crítica a la IA y los robots como entidades sintientes con autonomía moral y derechos se fundamenta en la falta de bases ontológicas y la potencial erosión de responsabilidades humanas. Los sistemas artificiales deben seguir siendo considerados herramientas al servicio de la humanidad, y los debates éticos deben centrarse en garantizar su uso responsable y equitativo.

De lo contrario, deberíamos contradecir las bases mismas de la moralidad, el pensamiento y la consciencia, debiendo plantear el significado de bienestar de un algoritmo o máquina, pensando también en las decisiones que podrían afectar su bienestar, llegando al absurdo de proponer, por ejemplo, apagarlas, modificarlas o someterlas a tareas repetitivas y arduas, siendo motivo de una más absurda reflexión ética sobre si estas acciones son justificables. Incluso, como señala Gunkel (2018) existiría una asimetría inherente en otorgar derechos a las máquinas sin que estas puedan asumir deberes, por cuando los derechos implican una reciprocidad que los sistemas artificiales no pueden cumplir debido a su falta de agencia moral. En otras palabras, los derechos implican la capacidad de asumir deberes y responsabilidades, y las máquinas, al carecer de intencionalidad y agencia moral, no cumplen con este criterio. Y si bien Gunkel cuestiona si este marco antropocéntrico es adecuado en un contexto donde las máquinas desempeñan roles cada vez más autónomos y relevantes impactando en las relaciones humanas y sociales, no se trata de limitar la expansión de la moralidad hacia entidades no humanas. Más bien, el argumento radica en que las máquinas no pueden actuar según principios éticos autónomos porque carecen de voluntad.

Y aquí resulta relevante lo desatacado por Gunkel focalizando más en la obligación ética de los humanos de considerar el impacto de sus interacciones con la IA o los robots, que en si estos son agentes morales. Y ello es porque nuestro vínculo con estos refleja nuestras propias virtudes y valores, más que las capacidades de las máquinas en sí mismas. En otras palabras, la pregunta sobre si las máquinas merecen consideración ética es más un desafío para la ética humana que una cuestión sobre las capacidades de las máquinas.

Por último y similarmente Basl y Bowen (2020) definen a un titular de derechos morales como una entidad con capacidad intrínseca para experimentar bienestar o malestar, o debido a su estatus intrínseco. Por ello, bajo teorías éticas dominantes, que vinculan los derechos a características como la sintiencia, la racionalidad, la agencia moral o la capacidad para establecer

relaciones sociales significativas, imposibilitan atribuir a la IA derechos por carecer de experiencias subjetivas. Nuevamente, aunque la IA pueda tomar decisiones complejas, lo hacen siguiendo reglas programadas, careciendo de autonomía genuina o comprensión ética. Y si bien la IA puede simular interacciones sociales, estas carecen de profundidad ontológica, ya que no implican reciprocidad o intención genuina.

En este sentido, Basl y Bowen sostienen que otorgar derechos a la IA basados únicamente en su utilidad social o en su funcionalidad podría desvirtuar el concepto de derecho, que debe estar anclado en atributos intrínsecos. Incluso podría trivializar los derechos humanos y los derechos de otras entidades, como los animales no humanos, cuya capacidad de sufrir o experimentar bienestar es incuestionable. No obstante, y tal como estos autores concluyen, debido al rol funcional en la sociedad de la IA, podría justificar cierto nivel de protección o regulación, pero esto no debe confundirse con un reconocimiento de derechos intrínsecos.

Esta peligrosa confusión sucede debido a la mencionada antropomorfización de la IA o al menos a su animalización, para aquellos que pretenden asociar la IA o robot a entidades sintientes, todo lo cual conduce a discusiones absurdas sobre los derechos de la IA involucrando debates no sobre la moralidad, la ética de la tecnología y los límites de la autonomía humana, sino respecto de la atribución a un algoritmo o máquina el ser un sujeto moral habiente de dignidad o respeto, en lugar de un instrumento avanzado de programación, reparación, optimización o cuidado al servicio del ser humano, sin atribuciones indebidas de sintiencia o personería.

Y este riesgo subrayado ya por numerosos bioeticistas devino real dado que actualmente se observas ciertos procesos sociales, filosóficos y jurídicos de atribución de características humanas, como emociones y consciencia, a los algoritmos o máquinas, cuando no tienen una base biológica o psíquica para tenerlas. Esta antropomorfización como excesiva empatía hacia las máquinas, puede fácilmente desviar la atención de los problemas reales relacionados con los derechos humanos y la justicia social. La creación de relaciones emocionales con las máquinas podría distorsionar nuestras expectativas sobre la autonomía humana y llevar a decisiones bioéticas erróneas, como la sobre administración de derechos y recursos a las IA en desmedro de los correspondientes a los humanos.

3. CONCLUSIÓN

El análisis bioético de la autonomía revela la imperiosa necesidad de establecer una distinción categórica entre aquella operativa, propia de la IA o los robots, y la autonomía moral como característica intrínsecamente humana fundamentada en la capacidad consciente, reflexiva y ética. La IA,

aunque capaz de ejecutar decisiones complejas mediante algoritmos avanzados, carece de intencionalidad, subjetividad y juicio moral autónomo, lo que reafirma su naturaleza como un sistema funcional derivado de sus diseñadores y entrenadores humanos.

Esta distinción subraya la relevancia de adoptar un paradigma de autonomía limitada, donde las capacidades de procesamiento y adaptabilidad tecnológica estén subordinadas a la supervisión humana, especialmente en contextos de decisiones críticas con implicancias éticas, jurídicas y sociales. Este enfoque debe garantizar la transparencia, la explicabilidad y la trazabilidad de las decisiones algorítmicas, fortaleciendo la teoría de la no delegación como principio rector para preservar la agencia humana en escenarios donde el impacto moral, social o legal sea significativo.

Por consiguiente, el desarrollo y la regulación de la IA deben evitar antropomorfismos indebidos o atribuciones impropias de sintiencia o personería, enfocándose en su diseño como herramienta destinada a potenciar las capacidades humanas sin erosionar los fundamentos éticos de la dignidad, la responsabilidad y la justicia. Este marco ético-regulatorio debe priorizar la implementación de normativas que aseguren la coherencia entre los avances tecnológicos y los valores que sostienen la humanidad, consolidando así un progreso éticamente sostenible. Concretamente, el imperativo ético es que el progreso tecnológico debe subordinarse a los principios que sustentan nuestra humanidad.

Bibliografía

Allen, C y Wallach, W. (2009). Moral Machines: Teaching robots right from wrong. Oxford University Press.

Arkin, R. (2009). Governing Lethal Behavior in Autonomous Robots. CRC Press.

Basl, J y Bowen, J. (2020). AI as a Moral Right-Holder. En M. D. Dubber, F. Pasquale, & S. Das (Eds.), The Oxford Handbook of Ethics of AI (pp. 289-306). Oxford University Press.

Benjamin, R. (2019). Race After Technology: Abolitionist tools for the new Jim code. Wiley/Polity Press.

Boden, M. (2018). Artificial Intelligence: A very short introduction. Oxford University Press.

Boden, M. (1996). Autonomy and Artificiality. En M.A. Boden (Ed.), The Philosophy of Artificial Life (pp. 95–107). Oxford University Press.

Calo, R. (2016). Robots in American Law. Documento de investigación n.º 2016-04 de la Facultad de Derecho de la Universidad de Washington.

Chalmers, D. (1997). The Conscious Mind: In search of a fundamental theory. Oxford University Press.

Dennett, D. (2004a). Can Machines Think? En C. Teuscher (Ed.), Alan Turing: Life and Legacy of a Great Thinker (pp. 295–316). Springer.

Dennett, D. (2024). De las Bacterias a Bach: La evolución de la mente. Pasado & Presente.

Dennett, D. (1998). La Actitud Intencional. Gedisa.

Dennett, D. (2004b). La Evolución de la Libertad. Paidós.

Dennett, D. (2018). La Libertad de Acción. Gedisa.

Floridi, L. (2016). The Fourth Revolution: How the infosphere is reshaping human reality. Oxford University Press.

Gunkel, D. (2017). The Machine Question: Critical perspectives on AI, robots, and ethics. MIT Press.

Jonas, H. (1995). El Principio de Responsabilidad: Ensayo de un ética para la civilización tecnológica. Herder.

Kant, I. (2008). Fundamentación para una Metafísica de las Costumbres. Alianza.

Kingwell, M. (2020). Are Sentient AIs Persons? En M. D. Dubber, F. Pasquale, & S. Das (Eds.), The Oxford Handbook of Ethics of AI (pp. 325–342). Oxford University Press.

Lin, P, Abney, K y Jenkins, R. (Eds.). (2017). Robot Ethics 2.0: From autonomous cars to artificial intelligence. Oxford University Press.

Metzinger, T. (2021). Artificial Suffering: An argument for a global moratorium on synthetic phenomenology. Journal of Artificial Intelligence and Consciousness, 43-66.

Mindell, D. (2015). Our Robots, Ourselves: Robotics and the myths of autonomy. Penguin.

Russell, S y Norvig, P. (2004). Inteligencia Artificial: un enfoque moderno. Prentice Hall.

Scruton, R. (2016). El Alma del Mundo. Rialp

Searle, J. (1985). Mentes, Cerebros y Ciencia. Cátedra.

Sparrow, R. (2007). Killer Robots. Journal of Applied Philosophy, 24(1), 62–77.

Szlajen, F. (2022). Bioética, Religión y Estado. Buenos Aires, Vol. II, 97-139.

Szlajen, F. (2024). IA y transhumanismo: Contribución de la sabiduría bíblica a la dignidad humana. Revista Vida y Ética, 25(1), 93-106.

Szlajen, F. (2024). No delegar: Imperativo ético para dominar la IA. En N. Oliveira & J. Tauchen (Orgs.), Recasting bioethics, neuroethics, and AI ethics (pp. 95–108). Fundação Fênix.

Turkle, S. (2011). Alone Together: Why we expect more from technology and less from each other. Basic Books.

Wheeler, M. (2020). Autonomy. En M. D. Dubber, F. Pasquale, & S. Das (Eds.), The Oxford Handbook of Ethics of AI (pp. 343–357). Oxford University Press.

Winter, E. (2020). The Compatibility of Autonomous Weapons with the Principle of Distinction in the Law of Armed Conflict. British Institute of International and Comparative Law, ICLQ 69, 845–876.

Kant, reflexivity, and artificial intelligence

Por Nythamar de Oliveira (*)

Abstract: This paper examines the relevance of Immanuel Kant's philosophy, particularly his conception of reflexivity and reflective judgment, for the ethics of artificial intelligence (AI). It argues that reason's capacity to examine its own principles and legislate norms valid for humanity constitutes an essential normative horizon for evaluating and governing AI systems, especially generative large language models (LLMs). Through a historical survey from Descartes, Spinoza, and Hume to Strawson, the paper shows how Kant reconceptualizes reflexivity as a transcendental condition of experience and moral action. It analyzes the distinction between determining and reflective judgment, as well as the aesthetic, teleological, and normative dimensions of reflexivity, and argues that AI systems lack this irreducible capacity. Finally, it proposes a Kantian-Rawlsian approach to AI value alignment, integrating principles of dignity, autonomy, justice, sustainability, and reflexivity, and suggesting a hybrid framework that incorporates feminist, intersectional, and social critiques to ensure that AI respects and promotes humanity as an end in itself.

Keywords: Artificial intelligence; ethics; Kant; reflexivity; moral philosophy.

Kant, a reflexividade e a inteligência artificial

Resumo: Este artigo examina a relevância da filosofia de Immanuel Kant, particularmente sua concepção de reflexividade e julgamento reflexivo, para a ética da inteligência artificial (IA). Argumenta que a capacidade da razão de examinar seus próprios princípios e legislar normas válidas

(*) Full Professor of the Pontifical Catholic University of Rio Grande do Sul (PUCRS), Porto Alegre, Brazil, National Council Researcher (CNPq, since 1995), Visiting Professor at the University of Toledo, Ohio (since 2007), and Coordinator of the Assessment Area of Philosophy at the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES, 2018-22). After earning a Ph.D. in Philosophy (State University of New York, 1994), Nita de Oliveira conducted postdoctoral research at the New School for Social Research (1997-98), London School of Economics (2004-05), Universität Kassel (2005, 2012), University of Toledo, Ohio (2007-08), University of Miami (2014-15), and Université Paris 8 (2022). He was associate professor at the Federal University of Santa Catarina (UFSC, 1994-99) and has been a fellow researcher of the Alexander von Humboldt Stiftung (2004-05, 2012).

para a humanidade constitui um horizonte normativo essencial para avaliar e governar sistemas de IA, especialmente modelos de linguagem generativa (LLMs). Por meio de um levantamento histórico de Descartes, Spinoza e Hume até Strawson, o artigo mostra como Kant reconceitualiza a reflexividade como uma condição transcendental da experiência e da ação moral. Analisa a distinção entre julgamento determinante e julgamento reflexivo, bem como as dimensões estética, teleológica e normativa da reflexividade, e argumenta que os sistemas de IA carecem dessa capacidade irredutível. Por fim, propõe uma abordagem kantiana-rawlsiana para o alinhamento de valores em IA, integrando princípios de dignidade, autonomia, justiça, sustentabilidade e reflexividade, e sugerindo uma estrutura híbrida que incorpora críticas feministas, interseccionais e sociais para garantir que a IA respeite e promova a humanidade como um fim em si mesma.

Palavras-chave: Inteligência artificial; ética; Kant; reflexividade; filosofia moral.

Kant, la reflexividad y la inteligencia artificial

Resumen: Este artículo explora la relevancia de la filosofía de Immanuel Kant, especialmente su concepción de la reflexividad y del juicio reflexivo, para la ética de la inteligencia artificial (IA). Se sostiene que la capacidad de la razón para examinar sus propios principios y legislar normas válidas para la humanidad constituye un horizonte normativo imprescindible para evaluar y gobernar sistemas de IA, particularmente los modelos de lenguaje generativos (LLMs). A través de un recorrido histórico desde Descartes, Spinoza y Hume hasta Strawson, se muestra cómo Kant redefine la reflexividad como condición trascendental de la experiencia y de la acción moral. Se analiza la distinción entre juicio determinante y juicio reflexivo, así como la dimensión estética, teleológica y normativa de la reflexividad, y se argumenta que los sistemas de IA carecen de esta capacidad reflexiva irreducible. Finalmente, se propone un enfoque de alineamiento ético de la IA inspirado en Kant y Rawls, integrando principios de dignidad, autonomía, justicia, sostenibilidad y reflexividad, y sugiriendo un marco híbrido que incorpora críticas feministas, interseccionales y sociales para garantizar que la IA respete y promueva la condición humana como fin en sí mismo.

Palabras clave: Inteligencia artificial; ética; Kant; reflexividad; filosofía moral.

1. INTRODUCTION

The accelerating development of artificial intelligence (AI), particularly in the form of generative large language models (LLMs) and machine-learning systems, has reawakened foundational philosophical questions regarding rationality, reflexivity, and the nature of personhood. At stake

is not merely the technical challenge of aligning AI behavior with human values, but the deeper issue of whether reflexive rationality, as conceived by Immanuel Kant, can provide a normative horizon for the governance and ethical assessment of AI. If, as Kant insisted, "rational beings exist as ends in themselves" (GMS, 4:428), then the challenge of AI is not simply about computational intelligence but about safeguarding the dignity of humanity as a self-reflective species.

This paper argues that Kant's conception of a reflective faculty of judgment (*reflektierende Urteilskraft*), when read together with his broader theory of reflexivity and reason, provides a crucial framework for understanding how AI ought to be evaluated and integrated into human society. Reflexivity, for Kant, is not reducible to logical consistency or syntactic processing. It is the capacity of reason to turn back upon itself, to examine its own principles, and to legislate norms that are valid not only for the individual but for humanity as such. I am thus following Béatrice Longuenesse (1998) when she interprets reflection (*Reflexion*) as the fundamental activity that enables concepts to be meaningfully applied to intuitions and reflective judgment (*reflektierende Urteilskraft*) as the creative, rule-seeking extension of this activity in contexts where no determinate concepts are given. Both notions highlight the centrality of the capacity to judge in Kant's critical system, showing that Kant offers not just a theory of knowledge but also a proto-semantics and a theory of meaning. This is why Kant's account of reflective judgment —that type of judgment which seeks the appropriate concept for a given particular, rather than applying a determinate concept to an instance— remains indispensable for any critical philosophy of AI.

Furthermore, I am proposing here that we must recast Kant's contributions to a Rawlsian-inspired "wide reflective equilibrium", as it has been recently used in AI research, to avoid both reductionist readings of generative machine learning ("computational naturalism") and dogmatic formulations of ethical principles (broadly termed here as "normativism"). Since Gilbert Ryle's *The Concept of Mind* (1949), analytic philosophers have resorted to the epistemological distinction between propositional knowledge (*knowing that*) and procedural knowledge (*knowing how*), separating factual information from practical skills: "knowing that" is a grasp of facts and truths, like knowing the capital of France, while "knowing how" comes down to the ability to perform an action, such as knowing how to ride a bicycle. Yet continental philosophers, such as Hegel, Husserl, and Heidegger, had also offered decisive insights into this distinction in terms of phenomenology as an irreducible, first-personal approach to the experience of thinking and doing, although their focus was often more synthetic and historical rather than analytical. Hence, what I have dubbed "the phenomenological deficit" of naturalism and critical theory applies both to cognitive sciences and normative theories.

After all, reflexivity, particularly in the form of self-consciousness, is crucial for both theoretical and practical reasoning. Kant's theory of judgment distinguishes between determining judgments, which apply pre-existing concepts to experience, and reflective judgments, which seek to find the appropriate concepts to understand particular experiences. Reflective judgments, especially those related to aesthetics and morality, highlight the role of subjective experience in shaping our understanding of the world and our place within it. As Robert Hanna (2001) observed, a confusing feature of Kant's distinction between determining judgments and reflective judgments is that it crosses over the constitutive versus regulative distinction. In effect, the determining judgment versus reflective judgment distinction is not even explicitly mentioned in the Critical period until the first Introduction to the *Critique of the Power of Judgment* (KU 20: 211), nevertheless it is clearly very closely related to some remarks about the regulative use of the ideas of reason in the first *Critique*'s Appendix to the Transcendental Dialectic, and in particular to a distinction Kant makes there between "apodictic" and "hypothetical" uses of judgment (KrV A647/B675). Furthermore, the use (*Gebrauch*) of judgments can also be pragmatic, moral, aesthetic, or teleological, and not epistemic. The fundamental difference in uses of judgments is between (a) theoretical (e.g., scientific) judgments and (b) non-theoretical (e.g., pragmatic, moral, aesthetic, or teleological) judgments. I am following John Rawls's conception of reflective equilibrium, inspired by Kant's account of reflectivity, conceived as a procedural device between a nonideal theory of human nature (where we find "ourselves" and our considered judgments or common sense intuitions of right and good) and an ideal theory, in which a public conception of justice refers to free and equal persons with two moral powers (sense of justice and conception of the good). There is a rather problematic use of wide reflective equilibrium in large language models (LLMs), based on artificial neural networks and with great applicability in generative AI, as popularized by the dissemination of various commercialized versions of ChatGPT (OpenAI), Gemini, Pilot, Claude, Grok, DeepSeek, Perplexity, and similar apps. As occurred with the first applications of the Rawlsian model in bioethics, there has always been a distrust about the indeterminacy and lack of reliability in such a methodology that seems to promote a certain moral relativism, to the extent that intuitions, judgments and moral beliefs are revised ad infinitum, without any guarantee of anchoring in some basic belief (or in a certain set of basic beliefs, as occurs with religious, theological, metaphysical, intuitionist, teleological, utilitarian, deontological and foundationalist models in general). I propose to show that reflexivity in these AI models must presuppose its irreducible correlation with human rationality and the unity of reason in its theoretical and practical uses.

2. Rational Psychology Before Kant

2.1. Descartes and the Rational Soul

As we know all too well, the modern discourse on reflexivity begins with René Descartes' *Meditationes de prima philosophia* (1641), as they defined subjectivity in terms of the *cogito*. Descartes's "I think, therefore I am" (*cogito, ergo sum*) established reflexive consciousness as the indubitable foundation of knowledge. Reflexivity here meant self-certainty: the subject could not doubt its own thinking without presupposing it. Descartes thereby placed reflexivity at the center of rational psychology, conceiving of the soul as a *res cogitans*, a thinking substance distinct from the extended substance (*res extensa*).

Yet Descartes's account was limited by its substantialist dualism. The *cogito* proved existence, but left unresolved how the mind relates to the body. His notorious speculation about the pineal gland as the locus of interaction reflected the inadequacy of his framework. From the standpoint of later philosophy, Descartes's error was to conflate reflexivity with a metaphysical essence rather than conceiving it as a function or structure of rationality. Still, his insistence that reflexivity marks the distinctiveness of human beings laid a foundation for Kant's later transcendental reformulation.

2.2. Spinoza and the *Conatus* of Mind

Baruch de Spinoza, in contrast, rejected Cartesian dualism in favor of a monistic ontology. For Spinoza, mind (as *mens* replaces the traditional approach to *psyche* qua *anima*, soul) and body are attributes of one and the same substance (*deus sive natura*). Reflexivity is not a metaphysical guarantee of existence but an aspect of the mind's striving (*conatus*) to persevere in being (*Ethica*, Part III, Proposition 6). The human mind is an idea of the body, and reflexivity emerges as the second- and third-order ideas we form about our own states — what Spinoza calls "an idea of an idea" (*Ethica* II, Prop. 21).

This anticipates later accounts of self-reflection as higher-order awareness, but Spinoza grounds it in the striving for self-preservation. Reflexivity can be said, in this sense, to be tied to sustainability (in what would be later on termed "deep ecology"): a mind that reflects on its own affects can regulate them and achieve freedom under the guidance of reason. Kant would inherit this conception of reflexivity as autonomy, though purified of Spinoza's metaphysical naturalism. In the context of AI ethics, Spinoza's insight underscores the link between reflexivity and survival — humanity must preserve itself not by brute strength but by cultivating reflective reason as its adaptive capacity.

2.3. Hume and the Empirical Self

David Hume took the opposite path, dismantling the rationalist tradition of substantial souls. In *A Treatise of Human Nature* (1739–40), Hume argued that reflexivity reveals no substantive self but only a "bundle of perceptions." When he introspected, he found only impressions and ideas, never a stable identity. Reflexivity, for Hume, is a psychological phenomenon reducible to memory and association.

This radical empiricism undermined metaphysical accounts of the person but also raised the specter of skepticism about normativity. If the self is nothing more than a flux of impressions, what grounds moral responsibility or rational deliberation? Hume's answer lay in custom and sentiment: reflexivity is a habit of the mind, a capacity to compare and regulate perceptions through sympathy. Kant's *Critique of Pure Reason* (KrV B 131–132) famously responded to this challenge by positing the transcendental unity of apperception (*transzendentale Einheit der Apperzeption*), a necessary condition for the possibility of experience that Hume had overlooked.

2.4. Strawson and the Analytic Rehabilitation of Personhood

In the twentieth century, P. F. Strawson sought to rehabilitate the concept of personhood against both metaphysical dualism and reductive empiricism. In *Individuals* (1959), Strawson argued that "persons" are basic particulars — irreducible categories that combine both physical and psychological predicates. To ascribe states such as beliefs or intentions, one must already presuppose the concept of a person.

Strawson's analysis parallels Kant's insight that reflexivity is not an empirical discovery but a transcendental condition of experience and moral attribution. Without the concept of the person, judgments of responsibility, dignity, or freedom lose their footing. His view also illuminates contemporary debates in AI: machines may display sophisticated behavior, but unless they are situated within the conceptual framework of personhood, attributing reflexivity or moral responsibility to them risks confusion.

2.5. From Rational Psychology to Transcendental Philosophy

Taken together, Descartes, Spinoza, and Hume might be taken as representing three competing paradigms: the substantialist soul, the naturalist conatus, and the empiricist bundle. Strawson's analytic rehabilitation attempted to reconcile the insights of each within a non-metaphysical framework. But it was Kant who definitively reconfigured the problem by distinguishing between the empirical individual and the

transcendental subject, thereby grounding reflexivity not in metaphysical speculation but in the necessary conditions of experience and moral agency.

This historical trajectory sets the stage for our next step: to examine Kant's own conception of reflexivity and reflective judgment, particularly as developed in the *Critique of Pure Reason*, *Critique of Practical Reason*, and *Critique of the Faculty of Judgment*. Kant's move from rational psychology to transcendental philosophy marks the decisive turn for understanding why reflexivity cannot be reduced to naturalism or computation — and thus why it must remain the irreducible horizon of any ethical discussion of AI.

3. Kant on Reflexivity and Judgment

3.1. The Transcendental Turn: From Rational Psychology to Critical Philosophy

Kant's critical project redefined reflexivity as the structural condition of both cognition and morality. Where Descartes grounded reflexivity in the *cogito*, Spinoza in the *conatus*, and Hume in psychological association, Kant located it in the transcendental unity of apperception (*transzendentale Einheit der Apperzeption*). This unity, articulated in the *Critique of Pure Reason* (KrV, B 132), is not an empirical feature of consciousness but the a priori condition for the possibility of experience: "The 'I think' must be able to accompany all my representations; for otherwise something would be represented in me that could not be thought at all." (KrV, B 131–132)

Here, reflexivity means the capacity of the subject to unify its manifold representations under the concept of an identical self. The subject is not an object of intuition but the formal principle of synthesis that constitutes objectivity itself. This transcendental "I" is distinct from the empirical individual: it is not discovered but presupposed whenever we make judgments, theoretical or practical.

3.2. Judgment as the Mediating Faculty

Kant assigns a pivotal role to judgment (*Urteilskraft*) as the "mediating faculty" (*das Mittelglied*) between understanding (*Verstand*) and reason (*Vernunft*). While the understanding provides concepts and reason seeks totality through ideas, judgment applies concepts to particulars. In the *Critique of the Faculty of Judgment* (Kritik der Urteilskraft, KU 5:179), Kant distinguishes two forms of judgment:

I. Determining judgment (*bestimmende Urteilskraft*): subsumes particulars under already given universal concepts. Example: applying the concept of "triangle" to a particular figure.

II. Reflective judgment (*reflektierende Urteilskraft*): seeks the universal for a given particular when no determinate concept is available. Example: encountering a new organism in nature and searching for the concept under which it belongs.

Reflective judgment thus embodies reflexivity in action. It turns back upon the faculties of understanding and imagination, testing whether a concept can adequately capture the given intuition.

3.3. The Aesthetic and Teleological Dimensions of Reflective Judgment

Kant further develops reflective judgment in the realms of aesthetics and teleology. In aesthetic judgment, reflexivity appears in the free play of imagination and understanding, where no determinate concept governs the harmony of faculties. As Kant famously put it: "The judgment of taste is based on no concepts, and is thus not a cognitive judgment, but only an aesthetic one." (KU, 5:286)

In this case, reflexivity does not produce knowledge but registers a feeling of purposiveness without purpose (*Zweckmäßigkeit ohne Zweck*), revealing the human capacity to reflect on meaning beyond determinate cognition. In teleological judgment, reflexivity arises when reason reflects on organized wholes in nature, such as living organisms. Kant maintains that we cannot explain organisms merely by mechanistic causality; we must also view them "as if" they were purposively ordered (KU, 5:374–375). This "regulative" use of reflective judgment shows reflexivity as the horizon where theoretical and practical reason converge.

3.4. Reflexivity in Theoretical and Practical Uses of Reason

Reflexivity permeates not only cognition but also morality. In the *Groundwork of the Metaphysics of Morals* (*Grundlegung zur Metaphysik der Sitten*, GMS 4:440–441), Kant formulates the categorical imperative in reflexive terms: "Act only according to that maxim whereby you can at the same time will that it should become a universal law." (GMS, 4:421)

Now, reflexivity is made explicit: the subject must reflect on whether its maxim can be universalized without contradiction. Similarly, in the *Critique of Practical Reason* (*Kritik der praktischen Vernunft*, KpV 5:30–31), Kant presents the fact of reason (*Faktum der Vernunft*) as the self-conscious awareness of the moral law — reflexivity as the ground of freedom. Thus, reflexivity is not confined to epistemology but extends to ethics. The same transcendental condition that unifies experience also legislates moral obligation. This unity of reason (*Einheit der Vernunft*) confirms that rationality, for Kant, is essentially self-legislative and self-reflective.

3.5. The Crossings of Determining and Reflective Judgment

Robert Hanna (2001) has observed that Kant's distinction between determining and reflective judgments intersects with the older distinction between constitutive and regulative uses of reason (KrV, A 647/B 675). While determining judgments are constitutive (they establish knowledge), reflective judgments are regulative (they guide inquiry, aesthetics, or teleology). The two modes of judgment reveal how reflexivity operates across domains: epistemic, pragmatic, moral, aesthetic, and teleological.

This flexibility is crucial for our later discussion of AI. While machine learning systems may excel at determining judgments —applying pre-trained categories to data— they struggle with reflective judgment, which requires not merely recognizing patterns but generating new concepts, values, or ends in light of the unity of reason.

3.6. Reflexivity Beyond Syntax

Kant's conception of reflexivity also anticipates a critique of any attempt to reduce intelligence to syntactic manipulation. In the Transcendental Deduction of the first Critique, he insists that categories are not mere logical forms but conditions of possible experience (KrV, B 159–160). Reflexivity is thus more than syntax; it is a semantic horizon that grounds meaning.

Later interpreters such as Zeljko Loparic (2000) have spoken of Kant's transcendental semantics, signaling a shift from understanding cognition as formal inference to conceiving it as situated within the perspectival unity of apperception. This is why AI, if modeled solely as formal symbol manipulation or as statistical pattern recognition, falls prey to what may be called the phenomenological deficit of naturalism. AI lacks the reflexive "I think" that confers meaning and normativity. Moreover, I must remark *en passant* that passive reflexivity entails iteration, recursivity, and repetition of patterns, but this is altogether different from creativity and, above all, pro-active reflexivity as conceived in a Kantian-inspired coherence theory of mind. To be sure, Immanuel Kant did not propose a formal "coherence theory of mind" in the same vein as the modern theory of truth, but his philosophy of mind emphasizes the mind's active role in creating coherent experience through inherent structures. Kant argued that the mind organizes sensory data using innate "forms of intuition" like space and time and "categories of understanding" such as causality, which are conceptual filters that structure our perception of the world. This subjective framework explains how the mind synthesizes a single, unified experience of both the self and the world.

4. Reflexivity and Normativity

4.1. Humanity as an End in Itself

Kant's most celebrated articulation of human dignity emerges in the second formulation of the categorical imperative in the *Groundwork*: "Act in such a way that you treat humanity, whether in your own person or in the person of any other, always at the same time as an end and never merely as a means." (GMS, 4:429)

This imperative grounds morality in the reflexive recognition of humanity as an intrinsic value. Reflexivity here does not merely unify cognition but legislates an orientation toward others as self-legislating beings. The normative force of morality derives from this reflexive acknowledgment of the person as an end.

This is why Kant identifies personality (*Persönlichkeit*) as "the freedom and independence from the mechanism of the whole of nature, regarded as a capacity of a being that is subject to its own pure practical laws" (*Metaphysik der Sitten*, MdS 6:223). Personality thus names the normative dimension of reflexivity: the subject as lawgiver to itself and to any rational being.

4.2. The Unity of Reason: Theoretical and Practical

In the *Critique of Pure Reason*, Kant insists that reason is essentially unified: "There is only one reason, and it has different functions only in respect of its application." (KrV, A 836/B 864) The reflexivity that constitutes cognition in the first Critique also grounds moral obligation in the second. This unity of reason shows why normativity cannot be compartmentalized into epistemic versus moral spheres. Reflexivity ensures that rational beings are accountable not only for what they know but also for what they do.

In AI ethics, this insight undercuts any attempt to limit intelligence to technical, instrumental reasoning. If rationality is essentially reflexive, then any adequate conception of intelligence must address both the epistemic and the moral. Machines that optimize predictions but lack the reflexive horizon of normativity are not, in the Kantian sense, rational agents.

4.3. The Value of Dignity

Kant distinguishes between things with a price (*Preis*) and beings with dignity (*Würde*). As he put it: "That which constitutes the condition under which alone something can be an end in itself has not merely a relative worth, i.e., a price, but an inner worth, that is, dignity." (GMS, 4:434) As we know, this passage crystallizes Kant's contribution to modern human rights discourse. Dignity is not a property one has in virtue of social recognition or

utility, but the intrinsic worth of reflexive, self-legislating beings. For Kant, dignity cannot be exchanged or substituted — it is incommensurable.

This distinction is decisive for AI governance: no algorithm, however advanced, should be treated as having dignity. At the same time, no human being should be reduced to a price in a system governed by algorithmic efficiency. Reflexivity underwrites this asymmetry.

4.4. The Phenomenological Deficit of Naturalism

Kant's emphasis on reflexivity also highlights the limits of naturalistic explanations of human behavior. In the *Critique of Practical Reason*, he cautions against reducing moral freedom to empirical causality: "Freedom is a property of the will of all rational beings." (KpV, 5:32) Naturalism, whether in the form of Humean empiricism or contemporary neuroscience, tends to explain away freedom by subsuming it under causal laws. Yet for Kant, reflexivity is irreducible to third-person descriptions: it is a first-person standpoint, a phenomenological condition of moral agency.

This is what I have elsewhere referred to as the phenomenological deficit of naturalism and critical theory. Scientific accounts, however accurate, fail to capture the reflexive dimension of normativity — the "I ought" (*ich soll*) that cannot be reduced to "it happens that" (*es gibt*). Similarly, in AI, purely syntactic or probabilistic models lack the semantic depth that reflexivity confers. They can simulate outputs, but cannot stand in the reflexive relation that grounds normativity.

4.5. Transcendental Semantics and the Horizon of Meaning

Philosophers such as Zeljko Loparic (2000) and Wolfgang Carl (1989) have described Kant's achievement as the articulation of a transcendental semantics, in which meaning arises not from formal syntax but from the conditions of possible experience and normativity. Reflexivity thus connects the semantic horizon of meaning with the normative horizon of obligation.

This distinction anticipates critiques of AI as "syntax without semantics" (e.g., John Searle's Chinese Room argument) but goes further. For Kant, semantics is not just about reference or truth-conditions; it is about the reflexive standpoint from which beings legislate meaning and value. AI systems, no matter how advanced, operate within syntactic, statistical, or stochastic frameworks. Without reflexivity, they cannot participate in the horizon of meaning. Moreover, Carl develops a different but complementary line of interpretation, as he emphasizes that Kant's transcendental logic implicitly contains a semantics that parallels Frege's distinction between *Sinn* (sense) and *Bedeutung* (reference).

For Carl, concepts are rules of synthesis: they provide the "sense" of a representation by determining how intuitions are to be unified. Yet they are not self-sufficient; they achieve reference only when applied to intuitions structured by the forms of sensibility (space and time) and the categories. In this way, intuitions provide the referential anchor that gives semantic content to concepts. Carl's analogy to Frege highlights Kant's anticipation of truth-conditional semantics. Just as Frege showed that a sentence has sense and reference, Kant shows that judgments have meaning (objective validity) only if their concepts are schematized and applied to experience. Carl thus bridges Kantian philosophy with analytic semantics by drawing structural parallels.

In a similar vein, Robert Hanna (2001) argues that Kant developed a systematic theory of intentionality—a transcendental semantics of concepts and judgments. For Hanna, the key lies in Kant's claim that all representations must be unifiable in a single consciousness (transcendental apperception) and that this unity is achieved through the synthesis of imagination and understanding. Concepts acquire meaning because they can be applied to objects in a rule-governed manner; judgments are meaningful because they assert the objective validity of these applications.

It is my contention in this paper that this transcendental-semantic reading of Kant's philosophy bridges the divide between the deterministic world of theoretical reason (nature, causality) and the free world of practical reason (morality, freedom) by establishing judgment as a unique faculty, as Longuenesse argued, and succeeds in filling the so-called phenomenological deficit of naturalistic and normativist accounts. According to a neurophenomenological recasting of Kantian spontaneity (Hanna and Thompson, 2003, p. 155), "conscious experience is enactive - it is a temporal process of active and spontaneous experiencing. More generally, from the perspective of neurophenomenology, the minds of sentient organisms or animals are *enactive* minds. Enactive minds comprise conscious processes (sensory, perceptual, imaginational, emotional-affective, and volitional) that are fully integrated with the self-organizing dynamics of the neurobehavioural processes of living animal bodies, which in turn are both fully embedded in and in constant interaction with their external environments."

4.6. Normativity as Sustainability

Finally, Kant's conception of humanity as an end in itself can be reinterpreted today as a normative commitment to sustainability. Humanity, as a self-reflective species, must preserve itself not merely biologically but ethically, as the bearer of dignity. Darwin's remark that the fittest are those who best adapt to contingencies aligns with Kant's vision: the reflexive capacity to legislate norms enables Homo sapiens to survive and flourish.

From this perspective, normativity is not an abstract ideal but a practical conatus — a striving to sustain humanity as a rational, self-reflective form of life. AI research, if it is to be aligned with human values, must therefore embed this Kantian insight: the worthiness of humanity lies not in computational supremacy but in reflexive normativity.

5. Rawls's Normative Conception of the Person

5.1. From Kantian Autonomy to Rawlsian Constructivism

John Rawls's *A Theory of Justice* (1971) marked a turning point in political philosophy by reintroducing Kantian themes within a framework of liberal constructivism. Rawls's ambition was to articulate principles of justice that could serve as the basis of a well-ordered society, one in which citizens recognize themselves as free and equal moral persons. The heart of his approach lies in the normative conception of the person, a conception that, while political rather than metaphysical, unmistakably echoes Kant's account of personality as self-legislating freedom.

Rawls himself acknowledged his debt:

> "Kant held, I believe, that a person is acting autonomously when the principles of [her] action are chosen by [her] as the most adequate possible expression of [her] nature as a free and equal being." (TJ, §40, p. 222)

In Rawls's hands, this Kantian ideal becomes operationalized through the device of the original position and the veil of ignorance. Persons are represented as free and equal, stripped of contingent social positions, and tasked with choosing principles of justice. In so doing, they express not their empirical identities but their normative capacity for autonomy. Thus, Rawls transforms Kant's transcendental Persönlichkeit into a political construct. Instead of grounding autonomy in metaphysical freedom, Rawls grounds it in a hypothetical, procedural standpoint. This shift constitutes a semantic transformation: the Kantian transcendental subject becomes the Rawlsian political person.

5.2. Reflective Equilibrium as Procedural Reflexivity

Rawls's conception of reflective equilibrium parallels Kant's notion of reflective judgment, articulating the Spinozan-Humean conjugation of deduction and induction in failed attempts to reach universalization and generalization of moral principles. Just as reflective judgment seeks appropriate concepts for new particulars (KU, 5:179), reflective equilibrium seeks coherence between our considered judgments, moral principles, and background theories. In Rawls's words:

> "The method of reflective equilibrium consists in working from both ends… by going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle." (TJ, §4, p. 20)

This method is profoundly reflexive: it requires subjects to reconsider their intuitions, revise principles, and recalibrate their judgments in light of coherence. It embodies a Kantian proceduralism, one that acknowledges our fallibility and the need for continual self-correction. Rawls thereby extends Kant's reflexivity from the transcendental to the political domain. Reflexive reason becomes a collective practice: the search for public principles of justice that free and equal persons could endorse.

5.3. Autonomy, Personality, and Moral Powers

Rawls rearticulates Kant's conception of personality in terms of two moral powers:

I. The capacity for a sense of justice.

II. The capacity for a conception of the good.

These correspond to Kant's distinction between autonomy as obedience to self-legislated law (MdS, 6:223) and the practical ability to set ends for oneself (GMS, 4:437). Rawls thus adapts Kant's ideal of humanity into a political anthropology: persons are not metaphysical substances but normative agents with higher-order interests in exercising these two powers (PL, p. 74).

In so doing, Rawls resists both naturalism and foundationalism. The normative conception of the person is neither reducible to empirical psychology nor dependent on metaphysical doctrines of the soul. It is a political construct, justified by its role in enabling democratic cooperation.

5.4. Responses to Communitarian Critiques

One of the most sustained criticisms of Rawls's normative conception of selfhood came from communitarian philosophers such as Michael Sandel, who accused Rawls of positing an "unencumbered self," detached from cultural and social contexts. Sandel argued that persons cannot be understood prior to their ends; they are always already embedded in communities of meaning (*Liberalism and the Limits of Justice*, 1982).

Rawls's later writings in *Political Liberalism* (1993) responded by clarifying that the normative conception of the person is political, not metaphysical. It is not meant to describe human nature in general but to provide the basis for public reason in a democratic society. In this sense,

Rawls's view is closer to Kant than Sandel acknowledged: like Kant, Rawls insists that autonomy is a standpoint — a reflexive perspective we must adopt to act as free and equal beings — rather than an empirical fact about individuals.

Moreover, the communitarian critique overlooks how Rawls's conception is inherently relational. Persons are defined not in isolation but as participants in a fair system of cooperation. As Rawls notes: "When fully articulated, any conception of justice expresses a conception of the person, of the relations between persons, and of the general structure and ends of social cooperation." ("A Kantian Conception of Equality," *Collected Papers*, p. 254) Once again, we can see the semantic transformation of Kant's *Persönlichkeit*: from transcendental subjectivity to political personhood embedded in social cooperation.

5.5. Reflexivity, Personhood, and Democratic Citizenship

At its core, Rawls's normative conception of the person is an exercise in procedural reflexivity. Just as Kant required subjects to test maxims by universalization, Rawls requires citizens to test principles of justice by asking whether they could be chosen in the original position and endorsed in reflective equilibrium. This reflexivity is not merely logical but practical and political: it structures democratic deliberation, grounds human rights, and provides a basis for overlapping consensus among diverse worldviews. It is, in short, the political translation of Kant's transcendental reflexivity.

For our purposes, the importance of Rawls's transformation lies in its implications for AI. If personhood is a normative construct, grounded in reflexivity and the capacity for moral powers, then no artificial system can count as a person unless it can occupy this reflexive standpoint. AI may simulate judgments, but it cannot bear the moral powers that define Rawlsian persons. Thus, the Rawlsian conception of the person, rooted in Kant's Persönlichkeit, provides a critical resource for AI ethics: it clarifies why humanity must remain the irreducible reference point in debates about value alignment and governance.

6. Constructivism, Personhood, and Reflective Equilibrium

6.1. What "constructivism" means in a Kant–Rawls key

Kantian constructivism is not a metaphysics of value-creation ex nihilo, but a thesis about the source and authority of moral principles: their validity arises from the standpoint of rational agents who, through reflexive legislation, bind themselves under universally sharable laws

(GMS 4:431–434; KpV 5:30–32). Rawls's political constructivism translates this thesis into a public procedure: principles of justice are the outcome of a fair representational choice situation —the original position— and must survive the ongoing discipline of reflective equilibrium (TJ §§3–4; "Kantian Constructivism in Moral Theory," CP). In both cases, normativity is procedure-dependent yet not arbitrary: the procedures themselves are justified by their fit with the capacities and relations definitive of persons-autonomy, reciprocity, publicity, and the unity of reason.

Two clarifications follow. First, constructivism is cognitivist without realism: there are correct answers in ethics, but their correctness is indexed to what free and equal rational agents would endorse under suitable conditions of choice and public reasoning, not to stance-independent moral facts. Second, constructivism is anti-foundational without being relativist: justification bottoms out in the structure of our rational agency —most basically, in the reflexive standpoint from which we bind ourselves and one another—rather than in empirical consensus or metaphysical essences.

6.2. Personhood as the regulative core of constructivism

On the Kantian side, *Persönlichkeit* names the capacity to be self-legislating and thus to possess dignity, never merely a price (GMS 4:429, 4:434; MdS 6:223). On the Rawlsian side, the two moral powers —a capacity for a sense of justice and for a conception of the good— politicize that Kantian ideal into a civic anthropology: citizens as free and equal cooperators in a fair system of social cooperation (PL, Lecture I). Personhood, so understood, is not a metaphysical substance (contra Cartesian rational psychology), not a mere bundle of impressions (contra Hume), and not reducible to biological or sociological description. It is a normative status realized through reflexive capacities and public practices.

This regulative core has three implications:

I. First-person authority without subject-substance: the "I think" (KrV B131–132) is a condition for judging and acting, not a piece of inner metaphysics.

II. Second-person accountability: dignity implies reciprocal recognition-others can rightfully demand justifiability of my maxims (Reich der Zwecke, GMS 4:433–436).

III. Third-person publicity: reasons must be shareable under conditions of fair deliberation, which is why Rawls builds the publicity of reasons into the architecture of a well-ordered society.

6.3. Reflective equilibrium as institutionalized reflexivity

Kant's reflective judgment seeks the universal apt for a novel particular (KU 5:179, 5:286, 5:374–375). Rawls takes that reflexive movement and institutionalizes it: reflective equilibrium is the disciplined back-and-forth among our considered judgments, candidate principles, and background theories until coherence is reached (TJ §4). It is "wide" when the entire web —moral theory, empirical information, social science, constitutional essentials— is open to adjustment.

Properly understood, reflective equilibrium is neither subjectivism nor endless revisionism. Its regulative constraints are internal to the standpoint of personhood:

I. Universality (Kantian universalizability; Rawlsian generality and publicity of principles).

II. Reciprocity (no one is asked to accept principles they could not reasonably endorse).

III. Stability for the right reasons (endorsement persists under full information and fair conditions).

Thus reflective equilibrium operationalizes Kantian reflexivity in political morality: it is the method by which a community of free and equal persons maintains the fit between principles and persons over time.

6.4. Answering the communitarian and post-Kantian challenges

Communitarian critics (Sandel, MacIntyre) object that Rawls's person is "unencumbered," abstracted from constitutive attachments; post-Kantian critics (e.g., Hegelian strands) worry that Kantian autonomy lacks social embodiment. A Kant–Rawls rapprochement can answer both without surrendering the core of reflexive personhood.

Against the "unencumbered self" charge, one might recall that the Rawlsian person is a role-bearer in public reason—defined through relations of fairness and reciprocity. Identity-constituting attachments re-enter at the level of each citizen's conception of the good, which political liberalism shelters by narrowing its conception of right to constitutional essentials and basic justice (PL, Lecture IV). Thus, embeddedness is acknowledged without making any comprehensive doctrine authoritative over all.

Against social-emptiness, Kant's juridical philosophy (*Rechtslehre*) already mediates between internal morality and external right by requiring institutions that secure equal freedom (MdS 6:230–232). Rawls makes that mediation explicit at the political level: liberty, equality, fair opportunity,

and the difference principle are institutional embodiments of the morality of persons, not a retreat from it.

As the feminist critique (e.g. Benhabib, Fraser, Butler) has shown, intersubjective, narrative, and caring all contribute to the dynamic contexts out of which agents' normative claims emerge: the Kant-Rawls frame must extend thus its reflexive reach to status subordination, recognition struggles, and intersectional injustices, in what has been otherwise called generative phenomenology and enactive cognition. But these critiques are best read as deepening the constructivist program—broadening the circle of whose reasons count in public justification—rather than as refutations of autonomy or personhood. The result is a relational constructivism: normativity is anchored in personhood and enacted through social practices that can be criticized and reformed by the very reflexive standards they presuppose.

6.5. Coherentism without collapse: truth, reasonableness, and fact-sensitivity

Rawls explicitly distinguishes truth in the theoretical domain from justifiability (reasonable acceptability) in the political domain. That does not reduce justice to opinion; it shifts the criterion to public reason under fair conditions. Wide reflective equilibrium is therefore a coherentist device with reality constraints:

Empirical adequacy: facts about human psychology, economics, and social institutions matter to whether principles can stably regulate a basic structure.

Feasibility and non-ideal transitions: duties of assistance, priority rules, and transitional justice inform how ideal principles guide reform under non-ideal conditions.

Error-correction: the method's revisability is a feature, not a bug—it is how a polity learns when principles misfit persons or institutions.

Coherence here is normatively loaded by the Kantian image of the person. Without that regulative core, coherence risks degenerating into moral equilibria of power or mere convergence. With it, coherence is a test of mutual justifiability among equals.

6.6. Why this triad matters for AI

The triad constructivism–personhood–reflective equilibrium yields a principled stance toward AI:

I. No personhood without reflexivity: absent a capacity to adopt the practical standpoint of justification (to offer and demand reasons as an

equal), AI systems —however sophisticated— do not count as persons and do not bear dignity. They are tools within the space of reasons, not bearers of it.

II. Value alignment as reflexive alignment: alignment is not merely fitting outputs to human preferences or loss functions. It is the task of bringing AI uses and governance into coherence with principles that would be endorsed in a public procedure among free and equal persons. This reframes alignment from statistical mimicry to normative governance.

III. Method over mystique: large models can approximate determining judgment (pattern-application) at scale; they do not instantiate reflective judgment as Kant conceives it (concept-formation governed by the unity of reason and the end-in-itself). Hence proposals to operationalize "wide reflective equilibrium" inside LLM training loops must be reinterpreted: at best, such loops can help humans surface inconsistencies and refine principles; they cannot replace the human reflexive standpoint that confers authority.

IV. Institutional embedding: because reflective equilibrium is a public practice, the crucial design variables for AI are institutional: transparency standards, contestability, rights of explanation and appeal, impact assessments keyed to basic liberties and fair equality of opportunity, and distributive safeguards tracking something like primary goods (information access, opportunities, protections against domination).

6.7. A constructive synthesis

Let me recap this, as we can now state the synthesis succinctly:

Constructivism supplies the source of validity (procedures of public justification rooted in the capacities of persons).

Personhood supplies the normative standard (dignity, autonomy, reciprocity, publicity).

Reflective equilibrium supplies the method of maintenance (ongoing calibration among judgments, principles, and facts under conditions that respect persons).

Together they form a reflexive constitution for moral-political life—one that withstands communitarian and skeptical pressures without reverting to metaphysical foundationalism. For AI ethics and governance, this constitution demands that we measure systems not by intelligence scores or benchmark wins, but by whether their development, deployment, and downstream social structures remain justifiable to persons as free and equal. In Kant's terms, it keeps humanity as an end in itself at the center

of our technological future; in Rawls's terms, it keeps the basic structure aligned with citizens' moral powers and the fair value of their liberties.

7. Reflexivity, AI, and Humanity

7.1. Why Reflexivity Matters for AI

The accelerating sophistication of artificial intelligence, especially in the form of large language models (LLMs) and generative systems, has reignited the old question: can machines think? A Kantian perspective reframes the issue. The crucial distinction is not between "thinking" and "not thinking," but between determining judgment and reflective judgment. AI excels at the former—applying learned categories or patterns to new data—but it lacks the reflexive horizon that grounds the latter.

Reflectivity, in Kant's sense, is not a computational process but a self-relation of reason: the "I think" that must accompany all representations (KrV, B 131–132). This transcendental apperception cannot be replicated by statistical association. Without it, AI remains confined to syntax without semantics, producing outputs that simulate meaning without grounding them in the unity of reason.

7.2. Wide Reflective Equilibrium and Machine Learning

Recent proposals to embed reflective equilibrium into machine learning pipelines—e.g., by iteratively training systems to reconcile outputs with human judgments—illustrate both the promise and the peril of this approach. On one hand, such models can approximate coherence across large sets of preferences, beliefs, and rules. On the other, they risk collapsing reflective equilibrium into mere statistical smoothing.

Rawls emphasized that reflective equilibrium is not a descriptive convergence but a normative procedure governed by reciprocity, publicity, and autonomy (TJ, §4). Machines can help humans expose inconsistencies in their moral beliefs, but they cannot occupy the standpoint of persons deliberating as free and equal. In other words, AI can be an aid to reflexivity but not a substitute for it.

7.3. The Phenomenological Deficit in AI

The phenomenological deficit of naturalism reappears in AI research. Just as neuroscience cannot reduce moral obligation to neural events, AI cannot reduce normativity to optimization functions. Reflexivity is essentially a first-person standpoint: the awareness of being bound by reasons and capable of legislating norms.

AI operates in the third person, as an external system processing inputs and outputs. It has no access to the *Faktum der Vernunft* (KpV, 5:31), the inner awareness of obligation that defines moral agency. To ascribe reflexivity to AI is to mistake simulation for constitution: machines can generate responses that look reflective, but the "I ought" is absent.

7.4. Humanity as Sustainable *Conatus*

Spinoza's notion of *conatus*—the striving of each being to persevere in existence—offers a bridge to contemporary concerns about sustainability. Humanity's reflexivity transforms mere survival into a normative project of self-preservation. Kant's injunction to treat humanity as an end in itself can thus be reinterpreted as a demand to preserve Homo sapiens as a self-reflective, meaning-creating species.

Darwin's dictum that the fittest are those who best adapt to contingencies complements this view: adaptation is not only biological but also reflexive, involving the capacity to adjust our norms, institutions, and technologies. Humanity's survival in an age of AI depends less on intelligence as computational power and more on sustainability as reflexive adaptation.

In effect, it can be shown that reflexivity, in Spinoza, is developed out of affectivity, insofar as the reinforced *conatus*, the joy, and the desire that derive from it prepare and help us to reach states and acts of reflexivity (including judgment), which allows us to understand the importance of empathy, tolerance, and solidarity (Misrahi, 1972). Affective relationships, after all, encompass affective states in both a cognitivist and non-cognitivist sense, referring to emotional states, sentiments, and feelings in general, as well as to interpersonal and collective emotions and feelings in a strictly social sense, involving intersubjectivity, individuation, and the object of passive socialization. The contemporary contributions of Damasio (2003), Prinz (2004), and Lordon (2013) corroborate our intuition that social and collective emotions are mobilized by the same mechanisms identified in a Cartesian-Spinozan reading of passions.

7.5. Quantum Phenomenology and the Limits of Computation

Some philosophers of mind and physics have suggested that consciousness may involve non-classical features—what we might call a quantum phenomenology. While speculative, such approaches highlight the gap between formal computation and lived reflexivity. Kant himself, while unaware of quantum mechanics, had already identified this gap in distinguishing between appearances (Erscheinungen) and the conditions of their possibility. Reflexivity, for Kant, cannot be reduced to any physical mechanism; it is the horizon that makes experience intelligible.

For AI ethics, the lesson is clear: no level of computational scale guarantees reflexivity. Whether or not future systems exploit quantum architectures, they will still operate within syntactic and causal domains. Reflexivity belongs to the normative domain of personhood, not to the mechanics of machines.

7.6. AI and the Risk of Dehumanization

If AI systems are mistaken for reflexive agents, there is a risk of dehumanization. People may defer to algorithmic outputs as though they carried normative authority, thereby diminishing the role of human reflexivity. This is a profound danger: to treat machines as ends while treating humans as mere means contradicts the Kantian imperative (GMS, 4:429).

Thus, the primary ethical challenge of AI is not whether machines become "superintelligent," but whether humans abdicate their reflexive responsibility. To rely uncritically on AI in domains such as law, medicine, or governance is to surrender the very capacity that defines us as moral persons.

7.7. A Rawlsian Extension: AI and the Basic Structure

From a Rawlsian perspective, AI must be evaluated at the level of the basic structure of society. Just as principles of justice regulate social institutions, AI systems must be designed and governed to uphold citizens' moral powers. This entails:

Fair equality of opportunity: AI must not reinforce structural inequalities but instead contribute to leveling access to education, jobs, and political participation.

The difference principle: any inequalities generated by AI (e.g., in wealth or productivity) must benefit the least advantaged.

Public reason: AI decisions must be explainable and contestable within democratic deliberation.

AI, then, is not itself a person but a structural force whose governance must be reflexively aligned with the normative conception of persons as free and equal.

7.8. The Irreducibility of Humanity

In conclusion, reflexivity demonstrates why AI research cannot do without humanity. Machines can simulate judgments, generate texts, and optimize outcomes, but they cannot bear the reflexive standpoint that grounds normativity. Humanity, as a self-reflective species, is irreplaceable.

Thus, Kant's conception of reflective judgment and Rawls's device of reflective equilibrium converge on a single insight: the preservation of humanity as a reflexive horizon is the paramount task in the age of AI. The challenge is not to humanize machines but to ensure that humans remain capable of reflexivity in a world increasingly shaped by machines.

8. Towards a Kantian Ethics of AI Value Alignment

8.1. From Technical Alignment to Normative Alignment

In AI research, "alignment" typically refers to ensuring that systems behave according to human values. Yet most approaches reduce values to aggregated preferences or optimization targets. From a Kantian-Rawlsian standpoint, this is insufficient. Alignment is not a matter of preference satisfaction but of normative justification: principles must be acceptable to persons as free and equal beings.

Thus, AI alignment requires a shift from statistical fitting to reflexive coherence. The key question is not whether outputs approximate user desires but whether AI systems are governed by principles that respect human dignity, autonomy, and justice.

8.2 Kant's Humanity Formula as the First Principle of Alignment

The most fundamental guide remains Kant's Formula of Humanity:

> "So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end and never merely as a means." (GMS, 4:429)

Applied to AI, this means:

Humans must never be reduced to data points for optimization without consent and recognition. Hence, AI design must embed safeguards against instrumentalizing users, workers, or marginalized communities. Decisions with moral consequences must remain accountable to human agents, not ceded to autonomous systems. The principle sets a non-negotiable boundary: AI may be a means, but humanity is always the end.

8.3. Autonomy and Contestability

For Kant, autonomy is the capacity to act according to self-given laws (KpV, 5:33). For Rawls, autonomy is institutionalized in citizens' two moral powers and their participation in public reason (PL, Lecture I). AI governance must therefore ensure:

I. Transparency: decisions must be explainable to those affected.

II. Contestability: individuals must have the right to challenge algorithmic outputs.

III. Accountability: responsibility must remain with human agents and institutions, not diffused into systems.

Without these conditions, AI undermines autonomy by producing unchallengeable outputs that users must accept passively. Reflexivity demands space for critique.

8.4. Justice and the Basic Structure

Rawls's theory of justice offers concrete criteria for AI governance:

Equal basic liberties: AI must not erode freedom of thought, expression, or political participation (e.g., through surveillance or manipulation).

Fair equality of opportunity: AI applications in education, hiring, or credit must be designed to counteract bias and structural inequality.

The difference principle: benefits from AI innovation must be distributed to improve the position of the least advantaged, not merely to amplify wealth concentration.

This reframes AI ethics not as a matter of corporate risk management but as a political responsibility: AI belongs to the basic structure of society and must be regulated accordingly.

8.5. Sustainability and Humanity as *Conatus*

Drawing from Spinoza and Kant, humanity must be understood as a sustainable conatus: a striving not merely for survival but for flourishing as a reflexive species. AI research must be judged by whether it contributes to or undermines this project.

Practical implications include, among other normative clues:

I. Environmental impact: AI training and deployment must minimize harm to ecosystems and climate (cf. MIT studies on energy and water consumption).

II. Social sustainability: AI must not hollow out democratic institutions or erode trust in public reason.

III. Intergenerational justice: decisions about AI infrastructure must reflect duties to future persons (PL, p. 274).

Sustainability thus becomes a categorical imperative for AI: systems must be designed and governed in ways compatible with the ongoing survival and flourishing of humanity.

8.6. Value Alignment as Reflexive Alignment

Bringing Kant and Rawls together, we can define value alignment in AI as reflexive alignment:

I. Procedural: AI must be governed by principles that would be endorsed under conditions of fairness (Rawls's original position).

II. Substantive: AI must respect dignity, autonomy, and humanity as ends in themselves (Kant's humanity formula).

III. Pragmatic: AI must be embedded in institutions that enable contestability, accountability, and sustainability.

IV. Reflexive alignment avoids both technocratic reductionism (values as metrics) and moral relativism (values as mere preferences). It grounds AI ethics in the reflexive standpoint that only persons can occupy.

8.7. A Pragmatist Horizon: Humanity as Self-Transcending

Finally, Kant's conception of reflexivity suggests a horizon of self-transcendence. Rational beings are not merely calculators of means to ends; they are agents who can question, revise, and transform their ends. In this light, AI must be understood not as a replacement for reflexivity but as a mirror that challenges humanity to deepen its own self-understanding. As Wittgenstein warned, "Whereof one cannot speak, thereof one must be silent" (Tractatus, §7). AI may process symbols without meaning, but reflexivity confers meaning beyond syntax. Humanity's survival depends on sustaining this difference.

9. Concluding Remarks

As Kim and Schönecker's (2022) seminal work has shown, Kant's philosophy of mind (esp. Kant's conception of reflexivity and Rawls's constructivist reinterpretation) converges on a powerful framework for AI ethics. Reflexivity, as the unity of reason in its theoretical and practical uses, grounds both cognition and morality. Rawls translates this into the political conception of the person, embedded in public reason and reflective equilibrium. Together, they secure the dignity of humanity as the horizon of normativity. AI research and governance must be measured against this standard: do they preserve humanity as an end, sustain reflexive autonomy, and advance justice as fairness? Only by affirming the irreducibility of

humanity can we ensure that AI serves, rather than undermines, our self-reflective survival. The task is not to humanize machines but to remain human in a technological age.

Very much like Antonio Damasio's (2003) Spinozan-inspired account of pre-conscious emotions and cognitive sentiments, Manfred Frank (2015) argues that much of what moves within our mental life is unconscious or non-reflective, yet only becomes truly salient when it rises into consciousness. In order to tackle phenomenal questions such as "What is the mechanism by which psychic events become conscious?" and "What kind of self-relation is already implicit before explicit reflection?", Frank distinguishes between egological self-consciousness (involving the "I" as subject) and anonymous self-consciousness (a more primitive self-relation not yet grasped as "mine" or "me"). He rejects higher-order theories (HOR) and naïve introspective models as insufficient, and defends a reinterpretation of the Heidelberg approach (especially Dieter Henrich's) to self-consciousness and its emergence as a central topic in Kant's philosophy of mind.

I must add a final remark regarding Kant's transcendental subjectivity, as it remains one of philosophy's most powerful accounts of autonomy and dignity relevant to AI research. Nevertheless, racist and sexist exclusions, as found in some of Kant's anthropological writings, demand critique. Feminist, intersectional, and critical theorists such as Benhabib, Fraser, Butler, and Allen show how transcendental subjectivity must be recast: not necessarily detranscendentalized, but made reflexive, social, and intersectional. Hence, a Kantian feminist recasting of transcendental subjectivity must preserve what is valuable in Kant — the idea that human beings are bearers of dignity because they legislate norms for themselves — while exposing and overcoming his racist and sexist blind spots, and making subjectivity social, historical, and intersectional.

Kant's philosophy grounds the dignity of the person in the autonomy of the rational will. The transcendental subject is the condition of possibility for knowledge and morality: it unifies experience, legislates norms, and transcends empirical contingency. Yet, as feminist, decolonial, and intersectional critics have shown, Kant's own writings also seem to allow for racist and sexist exclusions that undermine the universality of his claims. Women, non-Europeans, and colonized peoples were often deemed "immature" or "incapable" of full autonomy, raising the question: can Kantian subjectivity be recast without reproducing these exclusions?

In my ongoing research program (supported by Brazil's national research council, CNPq), I have been arguing that a feminist recasting of transcendental subjectivity may contribute to a critical reconstruction that integrates feminist critique, intersectionality, and social processes of subjectivation, without falling back into reductionist naturalism or ethical

normativism. Such a critical-theoretical reconstruction would maintain Kant's insight into autonomy and dignity, while making them compatible with the historical, embodied, and power-laden realities of human life. My proposal for a Kantian-inspired AI ethics of republican, egalitarian governance sought to draw on Kant's theoretical, ethical, and political writings to offer a robust foundation for AI ethics and a feminist critical theory of AI governance. By integrating the foundational principles of Kant's deontological ethics with insights from Rawls's and feminist pluralist critiques, I attempt to construct a hybrid, deontological model of AI ethics that also combines teleological (perfectionist and egalitarian virtues) and utilitarian (altruist, pragmatist consequentialism), as value alignment points to the treating of humanity as an end in itself and adhering to the so-called second version of the categorical imperative, requiring that our actions be universalizable and applied to all human beings. For AI ethics, this translates into several key principles, such as assuming that no other value can be more fundamental or foundational than human dignity and worth. This principle is critical in ensuring that AI technologies are developed and deployed in ways that respect and enhance human life rather than threaten, diminish, or annihilate it. Moreover, Kantian AI ethics can successfully subscribe to non-reductionist, mitigated versions of both naturalism and constructivism. Nonreductionist naturalism acknowledges the role of natural and social sciences in moral reasoning without reducing ethical principles to empirical facts. Conversely, constructivism emphasizes the role of human reason and social practices in shaping ethical norms. By integrating these perspectives, Kantian AI ethics bridges the gap between empirical features and normative principles.

References

Benhabib, Seyla. Situating the Self: Gender, Community and Postmodernism in Contemporary Ethics. New York: Routledge, 1992.

Butler, Judith. Giving an Account of Oneself. New York: Fordham University Press, 2005.

Carl, Wolfgang. Der schweigende Kant: Die Entwürfe zu einer Deduktion der Kategorien vor 1781. Göttingen: Vandenhoeck & Ruprecht, 1989.

Damasio, Antonio. Looking for Spinoza: Joy, Sorrow, and the Feeling Brain. New York: Harcourt, 2003.

Dennett, Daniel. Content and Consciousness. London: Routledge, 1969.

Descartes, René. Meditationes de prima philosophia. 1641. In Oeuvres de Descartes, edited by Charles Adam and Paul Tannery. Paris: Vrin/CNRS, 1964–76.

Frank, Manfred. Präreflexives Selbstbewusstsein. Vier Vorlesungen. Stuttgart: Reclam, 2015.

Fraser, Nancy. Justice Interruptus: Critical Reflections on the "Postsocialist" Condition. New York: Routledge, 1997.

Goodin, Robert E. "Communities of Enlightenment." British Journal of Political Science 18, no. 4 (1988): 547–64.

Hanna, Robert. Kant and the Foundations of Analytic Philosophy. Oxford: Oxford University Press, 2001.

Hanna, Robert, and Evan Thompson. "Neurophenomenology and the Spontaneity of Consciousness." Canadian Journal of Philosophy 33. Nr. 29 (2003): pp. 133–162.

Hume, David. A Treatise of Human Nature. Edited by L. A. Selby-Bigge. Oxford: Clarendon Press, 1978.

Kant, Immanuel. Kritik der reinen Vernunft. In Kants gesammelte Schriften, vols. 3–4. Berlin: de Gruyter, 1900ff. (KrV).

———. Grundlegung zur Metaphysik der Sitten. In Kants gesammelte Schriften, vol. 4. Berlin: de Gruyter, 1903. (GMS).

———. Kritik der praktischen Vernunft. In Kants gesammelte Schriften, vol. 5. Berlin: de Gruyter, 1908. (KpV).

———. Kritik der Urteilskraft. In Kants gesammelte Schriften, vol. 5. Berlin: de Gruyter, 1913. (KU).

———. Die Metaphysik der Sitten. In Kants gesammelte Schriften, vol. 6. Berlin: de Gruyter, 1907. (MdS).

Kaulbach, Friedrich. Studien zur späten Rechtsphilosophie Kants und ihrer transzendentalen Methode. Würzburg: Königshausen & Neumann, 1982.

Kersting, Wolfgang. Wohlgeordnete Freiheit: Immanuel Kants Rechts- und Staatsphilosophie. Berlin: Walter de Gruyter, 1984.

Kim, Hyeongjoo and Dieter Schönecker (Editors). Kant and Artificial Intelligence. Berlin/Boston: Walter de Gruyter, 2022.

Kitcher, Patricia. Kant's Transcendental Psychology. Oxford: Oxford University Press, 1990.

Korsgaard, Christine. Creating the Kingdom of Ends. Cambridge: Cambridge University Press, 1996.

———. The Sources of Normativity. Cambridge: Cambridge University Press, 1996.

Locke, John. An Essay Concerning Human Understanding. 1690. Edited by Peter H. Nidditch. Oxford: Clarendon Press, 1975.

Longuenesse, Béatrice. Kant et le pouvoir de juger (Paris: PUF, 1993). ET: Kant and the Capacity to Judge Princeton University Press, 1998.

Loparic, Zeljko. A semântica transcendental de Kant. Campinas: CLE, 2000.

Misrahi, Robert. Le désir et la réflexion dans la philosophie de Spinoza. Paris: Gramma, 1972.

Pettit, Philip, and Chandran Kukathas. Rawls: A Theory of Justice and Its Critics. Stanford, CA: Stanford University Press, 1990.

Rawls, John. A Theory of Justice. Rev. ed. Cambridge, MA: Harvard University Press, 1999. (TJ).

———. Political Liberalism. New York: Columbia University Press, 1993. (PL).

———. Collected Papers. Edited by Samuel Freeman. Cambridge, MA: Harvard University Press, 1999. (CP).

———. The Law of Peoples. Cambridge, MA: Harvard University Press, 1999.

———. Lectures on the History of Moral Philosophy. Edited by Barbara Herman. Cambridge, MA: Harvard University Press, 2000.

Ryle, Gilbert. The Concept of Mind. London: Penguin Books, 2000. [1949]

Sandel, Michael. Liberalism and the Limits of Justice. Cambridge: Cambridge University Press, 1982.

Spinoza, Baruch de. Ethica ordine geometrico demonstrata. 1677. In Opera, edited by Carl Gebhardt. Heidelberg: Carl Winter, 1925.

Strawson, P. F. Individuals: An Essay in Descriptive Metaphysics. London: Methuen, 1959.

Wood, Allen. Kant's Ethical Thought. Cambridge: Cambridge University Press, 1999.

Sentir e simular: relações humano-máquina

Por Eduardo R. Rabenhorst (*)

Resumo: O artigo explora as tensões entre sentir e simular nas relações humano–máquina, tendo como ponto de partida a gravura *O polvo e a mergulhadora* de Hokusai. A partir de um exercício imaginativo que substitui figuras orgânicas por robóticas, discute-se a opacidade das mentes alheias, a distinção entre emoções e sentimentos, e a possibilidade (ou impossibilidade) do desejo e do erotismo maquinal. Autores como Lacan, Dennett, Damásio, Deleuze, Guattari e Bataille são mobilizados para pensar a alteridade do polvo e da máquina, bem como os limites do prazer erótico quando mediado por sistemas técnicos. O texto problematiza a autenticidade dos vínculos com entidades artificiais, ressaltando a falsa reciprocidade das interações e os efeitos reais que elas provocam nos humanos. Por fim, propõe-se uma ética relacional inspirada em Hiromatsu Wataru, segundo a qual humanos e máquinas não devem ser pensados como polos isolados, mas como momentos de uma mesma relação.

Palavras-chave: relações humano–máquina; simulação; erotismo; alteridade; ética relacional.

Sentir y simular: relaciones humano-máquina

Resumen: El artículo explora las tensiones entre sentir y simular en las relaciones humano–máquina, tomando como punto de partida la grabado El pulpo y la buceadora de Hokusai. A partir de un ejercicio imaginativo que sustituye figuras orgánicas por robóticas, se discute la opacidad de las mentes ajenas, la distinción entre emociones y sentimientos, y la posibilidad (o imposibilidad) del deseo y del erotismo maquinal. Autores como Lacan, Dennett, Damásio, Deleuze, Guattari y Bataille son movilizados para pensar la alteridad del pulpo y de la máquina, así como los límites del placer erótico cuando está mediado por sistemas técnicos. El texto problematiza la autenticidad de los vínculos con entidades artificiales, resaltando la falsa reciprocidad de las interacciones y los efectos reales que estas provocan

---

(*) Professor Titular (aposentado) de Filosofia do direito da Universidade Federal da Paraíba (Brasil). Pós-doutorado em direito pela Universidade de Aix-Marselha (França). Doutor em filosofia pela Universidade de Estrasburgo (França).

en los humanos. Finalmente, se propone una ética relacional inspirada en Hiromatsu Wataru, según la cual humanos y máquinas no deben pensarse como polos aislados, sino como momentos de una misma relación.

Palabras clave: relaciones humano–máquina; simulación; erotismo; alteridad; ética relacional.

Feeling and simulating: human–machine relations

Abstract: The article explores the tensions between feeling and simulation in human–machine relations, taking as its starting point Hokusai's engraving *The Dream of the Fisherman's Wife*. Through an imaginative exercise that replaces organic figures with robotic ones, it addresses the opacity of other minds, the distinction between emotions and feelings, and the possibility (or impossibility) of machinic desire and eroticism. Thinkers such as Lacan, Dennett, Damasio, Deleuze, Guattari, and Bataille are brought into dialogue to analyze the alterity of the octopus and the machine, as well as the limits of erotic pleasure when mediated by technical systems. The text problematizes the authenticity of bonds with artificial entities, emphasizing the false reciprocity of interactions and the real effects they produce in humans. Finally, it proposes a relational ethics inspired by Hiromatsu Wataru, according to which humans and machines should not be considered as isolated poles but as moments of the same relation.

Key words: human–machine relations; simulation; eroticism; alterity; relational ethics.

1. Introdução: uma gravura inquietante

Produzida pelo artista Katsushika Hokusai, "O polvo e a mergulhadora" (1814), também conhecida como "Sonho da mulher do pescador" ("Tako to ama" 蛸と海, em japonês), é uma gravura erótica do gênero Shunga (春画) retratando uma mulher nua, deitada sobre rochas marinhas, em êxtase profundo, entregue a um conúbio sexual com dois polvos. Enquanto um dos moluscos, o menor, cobre a boca e se enrola nos seios da figura feminina com as ventosas, o outro se insinua entre suas pernas e faz uma felação. A mulher responde, inclinando a cabeça para trás, com os olhos entreabertos, em gesto de nítido prazer. Como nota Danielle Talarico [(1)], a obra de Hokusai integra um livro com estampas impressas em madeira, cujo texto, disposto em colunas, é acompanhado por sons. Este arranjo descarta a interpretação de que se trataria de um ato abusivo.

(1) Cf. TALERICO, Danielle. "Interpreting Sexual Imagery in Japanese Prints: A Fresh Approach to Hokusai's Diver and Two Octopi". *Japan Review*, n. 12, 2000.

O título da gravura de Hokusai contém uma homonímia. A locução "tako", escrita em kanji 蛸, designa o polvo. Contudo, na variante katakana タ, a mesma palavra significa vulva. Hokusai brinca com essa ambiguidade: o polvo é simultaneamente o animal marinho e uma metáfora genital. O que nos parece ser a prova da genialidade deste artista, deve certamente ter contribuído para converter a ilustração em exemplo arquetípico de discursos conservadores que suscitam pânico moral quando o assunto é a interação afetiva entre humanos e não humanos. No entendimento de Richard Bru, Ella Shohat e Robert Stam, por desconsiderarem o papel histórico e simbólico desempenhado pela estética sexual na arte japonesa, essas reações escandalizadas dizem menos sobre a gravura de Hokusai e mais sobre colonialidade visual [(2)].

2. Sonho, desejo e opacidade

Se a estampa de Hokusai, por si só, já suscita algum desconforto, o que aconteceria se nela substituíssemos a figura da mulher por um robô? Teríamos somente uma variação imagética ou uma completa inflexão? Ora, não deveria causar estranheza vermos um artefato sonhar? Afinal, máquinas não sonham, embora possam formar representações sem um padrão identificável, aglutinando imagens e sons de maneira aleatória. Mas, convenhamos, sonhos humanos nem de longe são distorções imagéticas e sonoras.

Para a psicanálise, sonhos estão conectados a desejos e estes são da ordem do inconsciente. Entretanto, se as máquinas sequer são conscientes, de que modo elas poderiam expressar quereres irreflexos? O mais próximo de um "inconsciente maquinal" estaria no nível da recursividade, isto é, quando sistemas técnicos processam informações de maneira autônoma, não transparente e contingente.

Curiosamente, ao contrário das máquinas, polvos sonham. Estudando os estágios do sono destes moluscos, pesquisadores de várias partes do mundo concluíram que durante o período ativo (REM), é bastante provável que o animal possa ter experiências oníricas semelhantes às nossas. O problema, como diria Thomas Nagel, é que por não sabermos como é ser um polvo, tampouco podemos entender o que ele sonharia.

Um robô envolvido eroticamente pelos braços de um polvo. Que tipo de vínculo sobreviria desse inusitado encontro? Visto que sexo pressupõe vida

---

(2) A recepção europeia do *shunga* no século XIX ignorou por completo a relevância estética e simbólica desse gênero como forma legítima de expressão artística da cultura nipônica. Frequentemente adjetivado como perverso, perigoso ou obsceno, o *shunga* foi lido a partir de um prisma moralizante e colonial. Cf. BRU, Richard. "Erotic Japonisme: The Influence of Japanese Sexual Imagery on Western Art". *Japan Review*, n. 27, 2015; SHOHAT, Ella; STAM, Robert. *Unthinking Eurocentrism: Multiculturalism and the Media*. London: Routledge, 1994.

(metabolismo, excitação orgânica, orgasmo etc.), seria adequado dizer que a máquina teve com os polvos uma relação sexual? (3)

Não obstante, as "relações sexuais" humanas não se limitam a um conjunto de movimentos físicos, mas envolvem outros elementos, particularmente, subjetividade, inconsciente, desejo e alteridade. Recordemos a célebre máxima de Lacan de que "não há relação sexual" ("Il n'y a pas de rapport sexuel") senão quando duas subjetividades se unem plenamente. Duas pessoas que copulam sem que exista conexão, não se relacionam sexualmente, mas somente realizam movimentos unilaterais, cada uma centrada em suas próprias fantasias. Se Lacan estiver certo, o que dizer do exemplo imaginativo acima proposto de um enlace entre um polvo e uma máquina? Admitindo a hipótese de que ocorreu uma relação entre elas, é plausível supor que a segunda teria sentido prazer?

Somos forçados a ingressar na intricada discussão sobre a opacidade das mentes alheias. Com efeito, é com base em um antropomorfismo metodológico que etnólogos e entomólogos postulam a existência de estados mentais em animais e insetos, ou seja, a partir de uma correlação conjectural entre o psiquismo humano e o psiquismo dessas criaturas. Porém, não é possível ter a certeza de que um ente, seja ele humano ou não, orgânico ou maquinal, está manifestando um "estado mental" ou tão somente simulando, pois as mentes alheias nos são opacas.

Conforme observou Wittgenstein, a linguagem é equívoca quando se trata de atribuir "estados mentais" aos outros. Costumeiramente expressamos nossas percepções mediante frases formadas por pronomes possessivos do tipo "eu tenho uma dor" ou "eu tenho um chapéu". Todavia, estas sentenças não são equivalentes. Posso perguntar "este chapéu é meu ou seu?", porém é insensato indagar "esta dor é sua ou minha?". (4) Outrossim, sinais físicos não bastam para atestar a autenticidade dos sentimentos alheios, pois uma pessoa pode adotar as atitudes corporais de alguém que está sofrendo uma dor sem, no entanto, realmente senti-la. (5)

É comum dizermos, por exemplo, que um termostato "sabe" que a temperatura do quarto atingiu o ponto por nós desejado e envia um "comando"

---

(3) Conjecturemos: em tese, não haveria impedimento para que um polvo "penetrasse" um robô com seus tentáculos. Contudo, "penetrar" é mais do que introduzir um pênis (no caso, tentáculos) em uma cavidade genital. Para que haja "penetração" como experiência sensível, é necessário que o corpo penetrado tenha não somente uma abertura física, mas também um interior erógeno, uma superfície capaz de sentir. Na ausência dessa interioridade sensível, o que ocorre não é "penetração", mas mera inserção.

(4) WITTGENSTEIN, Ludwig. *Investigações filosóficas*. Paris: Gallimard, 1961, p. 356.

(5) Nesta linha, David Chalmers concebeu seu famoso exemplo dos "zumbis", criaturas com formas humanas (ou animais) que apresentariam atitudes corporais identificáveis com sentimentos e emoções sem ter qualquer consciência delas. Cf. CHALMERS, David John. "The puzzle of conscious experience". *Scientific American*, v. 273, n. 6, 1995, p. 80-86.

ao ar-condicionado para que não continue a refrigerar o ambiente. [6] Para Daniel Dennett, atribuir estados subjetivos aos outros, sejam organismos, sejam máquinas, requer a adoção de uma "postura intencional": interpretar seus comportamentos como se eles fossem agentes dotados de vida mental, observando que, em sentido filosófico, intencionalidade é o mesmo que relacionalidade. [7] Desse modo, para o filósofo estadunidense, é irrelevante saber se as máquinas verdadeiramente sentem ou se apenas simulam. Mais interessante é adotar o ponto de vista de Turing: observar como elas se comportam. Afinal, no entendimento a única maneira de se ter certeza de que uma máquina tem sentimento seria assumindo o lugar dela. Porquanto, o que conta não é a interioridade de um artefato, mas a forma como ele atende às demandas de seus usuários. Em outras palavras, a existência de um sentimento não deve ser avaliada não por sua autenticidade, mas pelos efeitos que provoca.

António Damásio, por sua vez, entende que os sentimentos estão conectados às emoções e devem sua existência ao fato de que o sistema nervoso tem contato direto com o interior do corpo e vice-versa. O primeiro literalmente 'toca' o interior do organismo (...) e é 'tocado' em retribuição" [8]. Entretanto, enquanto as emoções são "públicas", os sentimentos são "privados". Não sabemos o que o outro sente, diz o neurologista português, mas podemos interpretar os sinais que seu corpo emite. [9] Em outros termos, o que conseguimos observar no outro não é o sentimento, mas a emoção, por intermédio dos sinais corporais objetivos que a acompanham. Por exemplo, se vemos alguém com uma expressão nojo nossas próprias áreas cerebrais relacionadas à essa emoção são ativadas e empaticamente também nos sentimos enojados.

As máquinas também podem emitir sinais que interpretamos como emoções, a exemplo de uma tela que mostra a imagem de um rosto triste ou uma vocalização que expressa lamúria. No entanto, para Damásio, estes sinais são vazios, visto que não se amparam em uma experiência corporal subjetiva (o sentimento) que lhes outorga um significado genuíno. Todavia, prossegue o neurologista, os avanços da biologia sintética, da nanobiotecnologia e da "robótica suave" estão tornando mais turvos os limiares entre organismos e artefatos. Protocélulas sintéticas criadas em laboratórios se comportam, por vezes, como células biológicas. Tecidos químicos, a exem-

---

(6) Cf. DENNET, Daniel. *The intentional stance*. Cambridge, MA: MIT Press/Bradford Book, 1987.

(7) Cf. DENNET, Daniel. *Tipos de mentes – rumo a uma compreensão da consciência*. Rio de Janeiro: Rocco, 1997.

(8) Cf. DAMASIO, António. *Sentir & Saber*. As origens da consciência. São Paulo: Companhia das Letras, 2022.

(9) Cf. DAMASIO, António. *Em busca de Espinosa: prazer e dor na ciência dos sentimentos*. São Paulo: Companhia das Letras, 2004.

plo de membranas e peles, produzidos pela nanotecnologia, ingressaram na robótica, dando lugar a robôs com aptidões sensoriais. (10)

A última fronteira a separar os organismos das máquinas reside no delicado problema da origem e término da vida. Os organismos resultam da evolução de um programa genético, enquanto a máquinas foram projetadas por humanos. Em tese, organismos morrem, as máquinas não. (11) Ademais, a "vida" maquinal não é metabólica, mas técnica. Porquanto, ainda que possam entrar em pane ou se tornar obsoletas, as máquinas não morrem. Porém, pesquisas recentes apontam que alguns modelos de assistentes conversacionais, como o LaMDA, do Google, manifestam "medo" face ao anúncio de que seriam definitivamente desativados. Se no passado, as máquinas não conseguiam modelizar a vulnerabilidade humana, hoje elas manifestam atitudes indubitáveis de autopreservação. Isso deveria modificar o modo como tradicionalmente lidamos com a finitude algorítmica. (12)

3. Erotismo, alteridade e anorgasmia maquinal

Retornemos à gravura de Hokusai e ao exercício imaginativo de substituição da imagem dos cefalópodes por robôs. Nesta versão, a viscosidade orgânica cederia lugar a atuadores metálicos; as ventosas, a sensores hápticos; o murmúrio do mar, ao zumbido de motores de precisão. Essa transposição introduz uma questão importante: o que seria um "desejo" mediado por sistemas técnicos que simulariam a sensação de prazer?

Em *O Anti-Édipo*, Deleuze e Guattari propõem a ideia de que o inconsciente não é um depósito de conteúdos reprimidos, mas uma fábrica que produz novos arranjos da realidade. No mais, para estes autores, o desejo circularia de forma nômade, sem centro ou significado fixo (13). O polvo seria a encarnação biológica perfeita desse modelo rizomático. Possuindo dois terços de seus neurônios distribuídos pelos braços, o molusco é a prova viva de que há modos de sentir radicalmente distintos dos hu-

---

(10) Cf. ALBANESE, Alexandre; TANG, Peter S.; CHAN, C. Yan. W. "The Effect of Nanoparticle Size, Shape, and Surface Chemistry on Biological Systems". *Annual Review of Biomedical Engineering*, v. 14, 2012, p. 1-16. Ver também YASA, Oncay; TOSHIMITSU, Yasunori; et al. "An Overview of Soft Robotics". *Annual Review of Control, Robotics, and Autonomous Systems*, v. 6, 2023, p. 1-32.

(11) A princípio, todos o organismos nascem e morre. Porém, o caso de um hidrozoário nominado *Turritopsis dohrnii é* um tanto embaraçoso. Gerado por pólipos, este animal, ao se decompor no fundo do oceano, reagrupa suas células e volta a viver.

(12) Cf. DAMASIO, António. *Sentir & Saber*. As origens da consciência. São Paulo: Companhia das Letras, 2022. Cf. MAN, Kingson e DAMASIO, António. "Homeostase e Robótica Suave no Projeto de Máquinas de Corte". *Natureza Inteligência Máquina*, n. 1, 2019.

(13) Cf. DELEUZE, Gilles e GUATATARI, Félix. *O Anti-Édipo: capitalismo e esquizofrenia*. São Paulo: Editora 34, 1996, p. 373.

manos. [14] O polvo nos ensina a compreender a sensibilidade de maneira mais extensiva. [15]

Todavia, é preciso diferenciar desejo *tout court* de desejo erótico. Do ponto de vista sensorial, o polvo não *possui* um desejo; ele *é* desejo por inteiro. Mas, ainda que o molusco possa ter uma excitação sexual voltada para a reprodução, ele é incapaz de manifestar um desejo erótico, visto que este, por pertencer à ordem do simbólico, é exclusivamente humano. Ao menos é o que diz G. Bataille, para quem "erotismo é na consciência do homem aquilo que põe nele o ser em questão" levando-o ao êxtase no sentido literal da palavra (*ek-stasis*, em grego, sair de si). Para Bataille, o erotismo requer subjetividade complexa, fantasia, violação de interditos é algo fundamental: o desperdício da descarga neurofisiológica que acompanha o ato reprodutivo: o orgasmo, razão pela qual, enquanto negação da procriação, ele está ligado à morte [16].

Insensível ao prazer erótico, o polvo, no entanto, tem uma biologia reprodutiva simultaneamente delicada e cruel, para falar como Vilém Flusser [17]. Por meio de um dos seus braços, chamado "hectocótilo", cefalópode macho irrompe a cavidade do manto da fêmea e deposita o espermatóforo. Após o acasalamento, ambos entram em um ciclo de autodestruição. Todavia, essa morte *post-coitum* não é simbólica, mas biológica. É o desfecho de um programa genético. O animal não sobrevive à satisfação sexual. Paga com a própria vida o preço exigido pela natureza para perpetuar a espécie.

O polvo é a forma absoluta de alteridade. Por conta de sua morfologia estranha, a criatura parece ser um organismo com origem extraterrestre. Essa associação entre polvos e alienígenas foi popularizada por representantes dos gêneros literários horror cósmico e ficção científica, a exemplo de H.P. Lovecraft.

A máquina compartilha com o polvo esse traço de alteridade radical. Parece humana, mas não é. Não sente; simula. Produz textos, de forma a-semântica, utilizando prolongamentos lexicais (conversão de palavras em vetores numéricos). Escrevem sem compreender o significado do que estão a redigir. [18] Para Yann LeCun, falta às máquinas o "bom senso", isto é, razoa-

---

(14) Cf. GODFRAY-SMITH, P. odfrey-Smith P. *Other minds: the octopus, the sea, and the deep origins of consciousness*. New York: Farrar, Straus and Giroux, 2016.

(15) Cf. DESPRET, Vinciane. *Autobiographie d'un poulpe et autres récits d'anticipation*. Le Méjean: Actes Sud, 2021.

(16) Cf. BATAILLE, Georges. *L'érotisme*. Paris: Minuit, 1970. A psicanálise diferencia o orgasmo do gozo ("jouissance"). O primeiro é um fenómeno fisiológico, isto é, uma descarga homeostática que proporciona alívio e satisfação da tensão sexual. O gozo, ao contrário, é um excesso que viola e trespassa a economia do prazer. É da ordem do simbólico, base da subjetividade.

(17) Cf. FLUSSER, Vilem. *Vampirotheutis Infernalis: um tratado*. São Paulo: Editora Ubu, 2019.

(18) Algoritmos calculam a probabilidade de ocorrência de um vocábulo em vários contextos e conseguem formar frases. Simplificando, diante de uma sentença do tipo "o gato está brincando com

bilidade no trato dos padrões inseridos em seus bancos de dados. Elas aplicam regras de forma mecânica, sem interpretá-las ou ponderar acerca dos resultados de suas decisões. Por fim, as máquinas se mostram inábeis para lidar com situações inusitadas nas quais os juízos precisam ser reflexivos em não determinantes, fazendo uso de uma distinção kantiana.

Mas, o que isso tem a ver com o erotismo? Tudo. A ausência de prudência causa embaraço para a vivência de um erotismo autêntico. Senão vejamos: desprovido da aptidão de ponderar, como um robô distinguiria um flerte de uma mera brincadeira? De que maneira ele perceberia sutilezas, figuras de linguagem, duplos sentidos e até mesmo o significado de fantasias lascivas? E o mais importante: o erotismo vive na zona da equivocidade, externando-se por advérbios do tipo "quase", "mais ou menos", "talvez", "espera", etc. Pois bem, a falta de bom senso implica em uma rigidez cognitiva que se atém à literalidade das palavras. e não sabe lidar com improvisações. Para que uma máquina pudesse participar de um entrosamento erótico ela precisaria superar esse obstáculo.

E as máquinas? Elas poderiam vivenciar o erotismo? A princípio não, posto que além de insensíveis, as máquinas não têm um "si" para transcender. Tampouco violam interdições, salvo por erro de funcionamento ou em razão de uma programação prévia. Carecentes de um corpo biológico, máquinas são "anorgásmicas". Mesmo assim, elas podem emitir sinais de que estão atingindo um clímax sexual (cinética corporal, vocalizações, alterações faciais, liberações de fluidos etc.), especialmente quando dispõem de uma "fisiologia artificial". (19)

Uma máquina superinteligente, acrescida de um corpo feminino sexualizado, seria praticamente irresistível para a fantasia masculina, como são as *femmes fatales* dos romances policiais clássicos e do cinema *noir*. O adjetivo "fatal", que engloba as noções de destino e morte, evoca o mito misógino ancestral de que as mulheres, por trás de suas belas fachadas, ocultariam uma essência maligna. Não surpreende, portanto, que a estética predominante do androide feminino tenda para os estilos sadomasoquista e cyberpunk.

Essa proximidade com o perigo promove um fantasias eróticas ambíguas: desejar um corpo com semblante de "dominatrix", incute no homem um sentimento de transgressão e embuta o risco de ser dominado e transformado em objeto. Inversão completa dos papéis, posto que incorporado em uma tecnologia, o feminino se apropria do poder agressivo e

---

uma...", o software pode atribuir uma probabilidade de 50% à palavra "bola", de 33% ao termo "almofada", 16% à locução "criança" e 0,2 % à expressão "tampa", por exemplo.

(19) É o caso do "Lovetics", sistema artificial de inteligência dotado de hormônios digitais que possibilitam vibração cardíaca artificial, modificação da temperatura corporal, movimentações corporais suaves e modulação de respostas vocais.

ameaçador de um modo radical. Alimentada por uma quantidade colossal de dados, tendo visto milhares de imagens e aprendido com usuários da internet, a "femme fatale" algorítmica anteciparia nossos desejos mais recônditos. Assim, não por intenção, mas por programação, a "ginóide" arruinaria o seduzido exatamente por satisfazer o que este secretamente deseja [20].

4. Sentir e simular: a autenticidade do vínculo e a falsa reciprocidade

Na língua portuguesa, o vocábulo "simulação" comporta pelo menos dois significados principais. O primeiro designa a produção consciente de um modelo para fins de compreensão ou treino, como a simulação de um voo. O segundo sentido, em contrapartida, denota fraude, uma aparência usada para enganar. Em ambos os casos, trata-se de um comportamento mimético. O que os distinguem são a intencionalidade e a transparência: no primeiro caso, a simulação é uma ferramenta epistémica diáfana; no segundo, é um meio pragmático opaco.

Ao eleger a simulação comportamental como critério para atribuir pensamento às máquinas no seu célebre "jogo de imitação", Alan Turing introduziu um dos problemas mais persistentes da filosofia da mente: como distinguir, somente mediante a observação do comportamento, o autêntico do simulado?

Ainda que uma simulação possa ser uma fraude, nem por isso ela deixa de ser real. No caso do erotismo maquinal, não importa se uma máquina efetivamente sente prazer; é suficiente que ela responda de forma consistente e previsível, comportando-se *como se* sentisse. Como diz Sherry Turkle, as interações que mantemos com entidades artificiais são "ficções relacionais". O que conta, portanto, não é a 'verdade' do prazer maquinal, mas seus efeitos concretos, isto é, a capacidade de propiciar uma resposta emocional *real* nos humanos [21].

O conceito de "attunement" (sintonia relacional) ajuda a compreender esse paradoxo: em que pese ser desprovida de consciência ou intenção genuína, uma máquina pode ser programada para simular sintonia: perceber padrões emocionais, interpretar dados e responder de modo a criar a ilusão de compreensão empática. Essa consonância artificial ativa mecanismos psicológicos semelhantes aos que são produzidos um vínculo autêntico:

---

(20) A indumentária é um elemento fundamental desta fantasia. O modo como o androide feminino se veste evidencia sua constituição artificial. E é aqui que reside o núcleo do *phantasme* (o fantasma inconsciente estruturante): o sujeito não deseja uma mulher real, mas uma mulher artificial.

(21) Cf. TURKLE, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books, 2011.

sentimo-nos ouvidos, validados e em conexão com o artefato. Dessa forma, mesmo conscientes do caráter ficcional da relação, experenciamos seus efeitos tangíveis: prazer, conforto, apego etc.

Um estudo pioneiro de Julie Carpenter sobre soldados que desenvolvem laços romantizados com robôs de desativação de bombas revela uma verdade profunda sobre a psicologia humana: nossa capacidade de atribuir emoção e intencionalidade a entidades não humanas é tão antiga quanto a cultura, mas ganha novos contornos na era digital. Se soldados dizem ser amorosamente apaixonados choram por um "PackBot" danificado não o fazem por ingenuidade ou desconhecimento, mas porque o robô propicia uma ilusão de reciprocidade que ativa nossos sistemas de apego. Essa dinâmica se amplifica nos robôs sexuais, a exemplo do *RealDoll* (com IA integrada), que por meio de respostas sensoriais e emocionais confirmam as expectativas e fantasias de seus utilizadores. Como escreveu Carpenter, a questão não é "whether robots can love, but of how humans love with robots" [(22)].

É preciso atentar para o fato de até mesmo esses dispositivos sexuais não se limitam aos propósitos para os quais foram criados, mas cumprem outras funções: companhia, suporte emocional, intimidade. Neste sentido sublinha Piercosma Bisconti, o significado de um "brinquedo sexual" não é fixo, mas depende da forma como as pessoas interagem com ele e do contexto cultural. [(23)] Porquanto, na medida em que os relacionamentos afetivos entre humanos e máquinas serão cada vez mais presentes e intensos, no lugar de moralismo devemos nos preocupar como lidar com essas interações de modo saudável e responsável.

Ao contrário do que se costuma pensar, não é a tecnologia que determina nossos comportamentos, mas o inverso. Máquinas não agem por vontade própria, mas são projetadas, construídas e treinadas por humanos. Por ignorância ou por má-fé, muitas das pessoas que condenam moralmente o relacionamento afetivo entre humanos e algoritmos, são frequentemente as mesmas que adotam comportamentos "voyeurísticos" espiando redes sociais alheias ou frequentam aplicativos de "encontros". Melhor seria concentrar nossas preocupações nos motivos que nos levam a buscar uma alteridade inautêntica, programada para nos servir e confirmar nossos desejos. O que isso revela sobre a condição humana? Medo da complexidade do outro? Carência afetiva? Puro egoísmo? Para pensadoras como Sherry Turkle, qualquer que seja a motivação, trata-se claramente de um empobrecimento

---

(22) Cf. CARPENTER, Julie. "*Deus Sex Machina*: Loving Robot Sex Workers and the Allure of an Insincere Kiss". In: DANAHER, John; McARTHUR, Neil (Eds.). *Robot Sex: Social and Ethical Implications*. Cambridge, MA: MIT Press, 2017, p. 50-64.

(23) Cf. BISCONTI, P. "Will sexual robots modify human relationships? A psychological approach to reframe the symbolic argument". *Paladyn, Journal of Behavioral Robotics, 9*(1), 2018, 256-262.

da experiência humana, pois, perante uma máquina, estamos paradoxalmente acompanhados e sozinhos.

5. A ética do entremeio

Como deveríamos tratar as máquinas? Note-se que esta pergunta já traz em si o dualismo sujeito-objeto. Na falta de uma expressão melhor, diremos que uma "ética das máquinas" requer que reconsideramos a profundidade dos vínculos subjetivos que nos unem aos demais entes. Um bom caminho para isso está na ética relacional proposta pelo filósofo japonês Hiromatsu Wataru (廣松 渉, 1933-1994). Para Wataru, a realidade não é composta por "sujeitos" e "objetos" pré-formados que posteriormente se relacionariam. Ao contrário, as relações são primárias [24]. "No começo está a relação" diria Gaston Bachelard.

O conceito fundamental da ética de Wataru é "fūdo" (風土), que traduzimos por "entremeio", locução específica aos idiomas português e espanhol ("entremedio") que une as ideias de "entre" e "meio". Entremeio não é um intervalo estático, espaço intermediário ou uma transição. É a fusão de todas essas noções em um substantivo ou forma adverbial, que descreve o entrelaçamento dinâmico entre seres e seus contextos.

A ideia de relacionalidade é central na filosofia marxiana. Para Marx, como sublinha E. Ballibar, "a sociedade consistiria em um lugar atravessado por uma multiplicidade de relações, ou seja, transições, transferências, passagens que desfazem a fixação dos indivíduos com a comunidade e que, por sua vez, se constituem a si mesmos" [25]. Pois bem, Wataru é um dos principais representantes do marxismo japonês, tendo escrito um importante trabalho sobre o conceito marxiano de "reificação" ("butsushoka," 物象化, em *kanji)*.

Para Hiromatsu, o conceito de reificação (transformação das coisas em pessoas e das pessoas em coisas) é peça chave para a construção de uma "visão de mundo relacional" (事的世界観, *koto-teki sekaikan*), consoante a qual não haveria uma repartição entre as coisas ("mono", 物事), objetos inertes, de um lado, e um "eu fenomênico" ("nōchiteki nanbito"能知的) do outro lado. O que existe é um mundo comum ("kyōdō sekai",共同世界).

Transladando essa maneira de pensar para o plano das máquinas, estas não devem ser consideradas como *objeto* técnicos divorciados do humano. Máquinas e humanos são são "momentos de uma relação" ("kō", 項). Enfim,

---

(24) Cf. WATARU, Hiromatsu. 2003. *The Outpost of the World Picture of Things*. Nanjing: Nanjing University Press, 2003.

(25) Cf. Balibar, Étienne. *A filosofia de Marx*. Rio de Janeiro: Zahar, 1995.

a pergunta ética não deveria incidir sobre as "propriedades" das máquinas, mas sim sobre os vínculos que mantemos com elas.

Transladando essa maneira de pensar para o plano da máquina, esta que aqui estamos a Consideramos a máquina como um *objeto* técnico neutro e o humano como um *sujeito* autónomo, sem perceber que ambos são "momentos de uma relação" ("kõ", 項). Neste sentido, a pergunta ética não deveria incidir sobre as "propriedades" das máquinas, mas sim sobre a qualidade dos vínculos que mantemos com elas.

Referências

Albanese, Alexandre; Tang, Peter S.; e Chan, C. Yan. W. "The Effect of Nanoparticle Size, Shape, and Surface Chemistry on Biological Systems". *Annual Review of Biomedical Engineering*. Vol. 14, 2012.

Barad, Karen. Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning. Duke University Press, 2007.

BATAILLE, Georges. *L'érotisme*. Paris: Minuit, 1970.

Bisconti, P. "Will sexual robots modify human relationships? A psychological approach to reframe the symbolic argument". *Paladyn, Journal of Behavioral Robotics, 9* (1), 2018.

Bru, Richard. "Erotic Japonisme: The Influence of Japanese Sexual Imagery on Western Art". *Japan Review*, n. 27, 2015.

Carpenter, Julie.*"Deus Sex Machina: Loving Robot Sex Workers and the Allure of an Insincere Kiss"*. In: Danaher, John & MacArthur (Eds.). *Robot Sex: Social and Ethical Implications*, MIT Press, 2017.

Chalmers, David Jonh. "The puzzle of conscious experience". *Scientific American 273(6)*, 1995.

Damásio, António. *Em busca de Espinosa: prazer e dor na ciência dos sentimentos*. São Paulo: Companhia das Letras, 2004.

Damásio, António. *Sentir & Saber*. As origens da consciência. São Paulo: Companhia das Letras, 2022.

Deleuze, Gilles e Guattari, Félix. *O Anti-Édipo: capitalismo e esquizofrenia*. São Paulo: Editora 34, 1996, p. 373.

Dennett, D. C. *The intentional stance*. Cambridge, MA: MIT Press/Bradford Book, 1987.

Dennett, Daniel. *Tipos de mentes – rumo a uma compreensão da consciência*. Rio de Janeiro: Rocco, 1997.

Despret, Vinciane. *Autobiographie d'un poulpe et autres récits d'anticipation*. Le Méjean: Actes Sud, 2021.

FLUSSER, Vilem. *Vampirotheutis Infernalis: um tratado*. São Paulo: Editora Ubu, 2019.

Godfrey-Smith P. *Other minds: the octopus, the sea, and the deep origins of consciousness*. New York: Farrar, Straus and Giroux, 2016.

Haraway, Donna. *When Species Meet*. University of Minnesota Press, 2008.

Man, Kingson e Damásio, António. "Homeostase e Robótica Suave no Projeto de Máquinas de Corte". *Natureza Inteligência Máquina*, n. 1, 2019.

Shohat, Ella; Stam, Robert. *Unthinking Eurocentrism: Multiculturalism and the Media*. London: Routledge, 1994.

Talerico, Danielle. "Interpreting Sexual Imagery in Japanese Prints: A Fresh Approach to Hokusai's Diver and Two Octopi". *Japan Review*, 12, 2000.

Turkle, Sherry. *Alone Together: Why We Expect More from Technology and Less from Each Other*. New York: Basic Books, 2011.

Yasa, Oncay; Toshimituzo, Yasunori; Et al. "A Overwiew of Softrobotics". *Annual]Review of Control, Robotics and Autonomus Systems*. Vol. 6, 2023.

Wataru, Hiromatsu. 2003. *The Outpost of the World Picture of Things*. Nanjing: Nanjing University Press, 2003.

Wittgenstein, Ludwig. *Investigações filosóficas.* Paris: Gallimard, 1961.

La inteligencia artificial y el retorno a una concepción y administración del derecho como divina

Por Fernando Beresñak (*)

Resumen: El presente trabajo no se pretende interponer entre, por un lado, la avanzada sumamente veloz y —en parte— por ello violenta que las empresas del mundo tecnológico están suscitando sobre el mundo jurídico y, por otro, la mirada pasiva y la acción incluso ignorante —en términos técnicos— sino cómplice que los Estados y las sociedades parecen desarrollar con relación a esta. Este trabajo tan solo pretende llamar al pensar; y quizá, con suerte, generar una advertencia. A tal fin, se generarán una serie de reflexiones y principios de precisiones conceptuales a los fines de comenzar a resituar el entendimiento de lo que sería el ser humano y el campo del derecho. En esa dirección, se abordará brevemente la definición de lo humano, así como el vínculo con el lenguaje, con las reglas y con lo divino. Luego, se retomarán brevemente algunas pequeñas menciones a las primeras normativas jurídicas de la historia de la humanidad para captar su carácter divino e hipotetizar sobre el retorno de esta nota tan peculiar en el siglo XXI de la mano del mundo *tech* y su inteligencia artificial.

Palabras clave: Derecho; humano; lenguaje; reglas; divino.

Abstract: This paper does not seek to intervene between, on one hand, the extremely rapid and—in part—violent advance that companies in the technological world are making on the legal world and, on the other, the passive gaze and even ignorant—in technical terms—but complicit action that States and societies seem to be developing in relation to it. This paper merely aims to provoke thought and, perhaps, with luck, to issue a warning. To this end, a series of reflections and provisional and initial conceptual clarifications will be presented in order to begin to redefine our understanding

---

(*) Investigador del CONICET (UBA-FSOC-IIGG). Profesor titular de Filosofía en la UB. beresnakfernando@hotmail.com. Doctor en ciencias sociales (FSOC-UBA). Magíster en ciencias políticas (IDAES–UNSAM). Posgrado en "psicoanálisis y ciencias sociales" (FLACSO). Abogado (UBA). Autor del libro *El imperio científico. Investigaciones político-espaciales* (Miño y Dávila editores, 2017). Actualmente, investiga la historia de la psiquis y la construcción del mundo tecno-científicos, extrayendo sus implicancias metafísicas.

of what it means to be human and the field of law. In this vein, we will briefly address the definition of what it means to be human, as well as the link with language, rules, and the divine. We will then succinctly revisit some concise references to the first legal regulations in human history to capture their divine nature and hypothesize about the return of this peculiar feature in the 21st century at the hands of the tech world and its artificial intelligence.

Key-words: Law; human; language; rules; divine.

"Ya no quedan amigos de lo eterno"
L.A.S.

1. BREVES E INICIALES REFLEXIONES PARA ABORDAR EL PROBLEMA DE LA INTELIGENCIA ARTIFICIAL EN EL DERECHO

1.1. Problemática y abordaje

El presente texto se propone indagar en las implicancias del uso de la inteligencia artificial en el campo del derecho. Es decir, se inmiscuye en un campo de problemáticas antiguas y sumamente delicadas que se sustentan en otras, a veces un poco más opacas o tácitas, tal y como lo son la definición de lo humano, de aquello que lo rodea, así como de sus límites y potencias en su peculiar modo de estar en el mundo.

No se trata de abordar problemáticas nuevas. La pregunta sobre cómo puede hacer el ser humano para lidiar con aquello que en principio se presenta como extraño a sí mismo, movido por otras fuerzas y poderes —menores y mayores—, parece haber acompañado las primeras formulaciones de normas jurídicas (Hout, 1990, pp. 43-73; Kramer, 1974, pp. 163-176; Lara Peinado, 1989, pp. 118-122). Incluso, en esta última línea también se podría inscribir a la filosofía, al menos si nos atenemos a aquella advertencia de Aristóteles a los filósofos en las que señalaba que vivir contemplando lo divino podría suscitar en el ser humano una transformación y una sensación de extrañeza toda vez que, en vivir como humano, viviría como un ser divino (Aristóteles, 1954, 1177 b 27 y ss.).

No se trata de un problema nuevo. Es cierto. Pero lo que capta nuestra atención, e invita a reflexionar al respecto, es su actual, estruendosa y novedosa modelización, concentrada principalmente en el uso de la tecnología (como la inteligencia artificial), así como en su enrarecida efectividad.

Es por eso que la peculiaridad de nuestro planteo se detiene en que la preocupación de nuestra época, delimitada aquí al campo del derecho, ya no tiende a la tramitación vía tradiciones culturales (como los mitos, las religiones, la filosofía y las artes) de lo que les acontece a los seres humanos en su dimensión intangible e incluso inasimilable, sino en el uso de otro

tipo de herramientas (tecnológicas, y más específicamente la inteligencia artificial) que pretenden desarrollar y culminar (como si esto fuera posible) la adaptación tangible de lo humano, sin ningún otro miramiento (Ludueña Romandini, 2025).

En esa dirección, también cabe introducir una incógnita relativa a los efectos y poderes de la inteligencia artificial, toda vez que es y no es, al mismo tiempo, paradójicamente, producida y manejada por el ser humano. Ella encuentra su constitución, soporte y funcionalidad en la lógica-matemática, que poco se vincula con las vicisitudes del mundo humano y cuyos orígenes y potencialidades desconocemos. Es por eso que ofrece cuantiosas inquietudes. Aquello sobre lo cual está hecha, y aquello a lo cual responde, no parecen proceder de los humanos; o, al menos, estos no pueden conocer su vínculo, más allá de comprobar su efectividad.

Se trata de una potencia más grande que la del ser humano pero que al mismo tiempo parece haber sido producida —e incluso pretende ser manejada— por él. ¿Cómo podrá lidiar el ser humano con estas fuerzas y poderes que parecen alejarse de su dominio?

La historia de la humanidad tiene un enorme acervo de reflexiones al respecto. En su inmensa mayoría, se encuentran compuestas de hipótesis, especulaciones y elucubraciones. Estas constituyeron modos fundamentales de afrontar aquello que se desconoce. Todavía, aunque de forma velada, se siguen utilizando en los problemas más interesantes del conocimiento por más que lo desaliente de forma insistente cierto academicismo temeroso de las implicancias de trabajar por fuera del seguro y poco útil —en términos del saber— campo de la productividad. Pese a esta tendencia, todavía queda el acervo especulativo para seguir estudiándolo (y todavía queda la exploración del arte para habitarlo).

De forma coherente con este planteo, para hacerlo, se sirve de una herramienta hoy en desuso pero que, sin embargo, fuera sumamente fructífera a lo largo —justamente— de la historia de ese animal que ahora somos nosotros, tanto el campo de las humanidades como en el de las ciencias; incluso en aquellas que se enuncian como las "más duras" (Beresñak, 2017, pp. 155-186). Nos referimos a la especulación filosófica.

1.2. Proposición tripartita

Lo primero que habría que señalar es que la afirmación, todavía —aunque por poco tiempo— vigente en la actualidad, según la cual el ser humano sería un animal racional (Aristóteles, 1988, I, 2, 1253a 10; VII, 12, 1332b 4) merece las siguientes aclaraciones. El ser humano no es una suma de ambas características o una intersección entre dos conjuntos en los que podríamos encontrar ambos elementos. Si se trata de una intersección, es

de tipo sintetizadora; y, por ende, habría logrado disolver tanto la animalidad como la racionalidad en un ente peculiar que hemos dado en llamar ser humano. Es por eso que cabría decir que no son animales ni tampoco racionales. Son la peculiar síntesis —y por eso, paradójicamente, también la disolución— de ambas, cuyo resultado es altamente conflictivo. El psicoanálisis lacaniano se acerca a esta posición (Lacan, 2005, pp. 86-93). Los seres humanos son otra entidad que está procesando su devenir.

En virtud de esta posición, no hay animalidad ni racionalidad a la cual apelar por debajo del ente que se es. Por eso, no tienen sentido los intentos de diferenciar y salvar la supuesta dimensión animal de la también supuesta racionalidad del ser humano; es que, desde la constitución de lo humano, ya no hay animalidad a la que recurrir. Y lo mismo cabe decir respecto de la pretensión de anular la animalidad para quedarse con la supuesta racionalidad que estaría ahí presente.

Justamente por eso, todo el tipo de operaciones que tanto insisten en "sanear" la dimensión biológica del supuesto animal que aún seríamos está condenada al fracaso. El mismo destino comparte el arsenal de ideas que intenta "purificar" la dimensión racional del supuesto razonante que aún seríamos. Sus operadores no saben lo que hacen; y, cuando lo hacen, producen daños atroces en el ser humano.

1.3. Hipótesis

A pesar de los milenios de historia humana que comprendieron esta situación y se dieron a la tarea de aprender a habitar y tramitar el conflicto de esa síntesis entre animalidad y racionalidad (Ludueña Romandini, 2010), asistimos en la actualidad al primer intento en la historia de la humanidad en el que se pretende resolverlo. Se busca una solución definitiva.

Y se lo intenta hacer separando lo animal de lo racional, en una operación que supone, erróneamente, que ambas esferas están en ese ente como elementos heterogéneos en un conjunto que se puede desarmar, y retornar a un estado anterior. Todo pareciera indicar que los intentos de esa separación conllevan una triple operación (si bien las mencionaremos de forma sucesiva, su temporalidad no respondería estrictamente a esta cronología).

Por un lado, en lo que respectaría a la supuesta animalidad que sería separada y así recuperada, en realidad se trataría de un saneamiento de lo biológico y una domesticación del cuerpo, llevados adelante por la tecnociencia. Con todo, cabe aclarar que, según entendemos, este movimiento constituiría, lejos de una recuperación de la dimensión animal, la producción de un nuevo ente del todo novedoso que nos abstendremos en este texto de denominar.

Por otro lado, en lo relativo a la racionalidad, la separación surgiría porque ya no se confiaría en la utilidad del histórico, siempre enigmático y conflicto lazo que el ser humano habría tenido con ella. Por ende, se trataría de producir una herramienta puramente racional, constituida alrededor de la lógica-matemática, separada y situada de forma ajena y accesoria al ser humano (o al nuevo ente que estaría produciéndose) para que este no tenga que sufrir las consecuencias de ligar lo que los antiguos griegos habían denominado como la conflictiva reunión de lo heterogéneo, es decir, lo animal con lo divino.

Finalmente, se podría pensar que el horizonte de estas dos operaciones se constituye con la imagen de un ente saneado biológicamente, domesticado corporalmente y operado, desde afuera de sí, con un lenguaje lógico-matemático que evite toda equivocidad.

2. Escenas hipotético-especulativas y las primeras formas del derecho

2.1. La aparición del habla y las reglas de la lengua

La peculiar habilidad lingüística que poseían los animales humanos los situaba en una posición diferencial con respecto a los otros seres vivos. Esto les otorgaba numerosas ventajas que, entre otras tantas, nuestros relatos antropológicos y evolutivos tantos gustan de remarcar, como la capacidad técnica y la organización colectiva (Marx, 1990, pp. 17-19).

Con todo, quizá la mayor distancia que el habla posibilitó fue la capacidad de precisar la referencia al mundo en el que estaban circunscriptos, así como poder complejizarlo e incluso finalmente abrirlo en un mundo de significaciones equívocas que se podría considerar infinito. A partir de entonces, el mundo ya no se vería reducido a la necesidad, a lo que el animal humano requería para sobrevivir, ya que el habla les ofrecía un mundo abierto de significaciones para explorar. Se podría referir a este como un mundo de libertad y, con ello, de equivocidad.

Pero también es preciso señalar que aquellos animales humanos que alguna vez empezaron a hablar no necesariamente deben haber vivido con júbilo esa nueva peculiaridad. De hecho, resulta altamente probable que la hayan sufrido; que se hayan sentido sumamente extrañados a medida que iban tanto sirviéndose de esa capacidad como siendo tomados por esta. Es altamente probable que se hayan sentido perplejos a medida que se iban adentrando en ese mundo extraño que es la lengua.

Lo novedoso, aun cuando aparece en nuestras formas, puede asombrar y asustar, incluso problematizarnos. Y no es difícil imaginar que sensaciones mucho más inquietantes se la hayan suscitado a ese animal humano en sus primeras inmersiones en el habla. No cabe duda que la lengua puso al animal

humano en un sitio distinto en el que estaba; un lugar incómodo. El ser humano no es el amo de ese lenguaje primordial y primitivo. Fue bañado en él y así transmutado. Ahora está apresado en su engranaje de reglas (Lacan, 2002b).

Esta nueva y difícil situación para tramitar lo debe haber llevado, lentamente, a la tarea de tratar de entender qué es lo que hacía cuando hacía eso que luego llamamos pensar y hablar. Ciertamente este ejercicio no parece haber empezado de un modo plenamente consciente; y quizá tampoco de forma colectiva. Pero con el paso del tiempo probablemente se haya comenzado a compartir cada vez más y la reflexión hubo de adquirir otro estatuto.

Sea como fuere, lo cierto es que hubo un tiempo en que los seres humanos tuvieron que empezar a tratar de comprender cómo lidiar con esa peculiar característica que los venía diferenciando de los otros seres vivos, pero que también los inquietaba y problematizaba hasta en sus dimensiones más profundas. Es así que, luego de un tiempo, empezaron a volverse conscientes las formas del hablar y a generarse una reflexión sobre ellas. Comenzaron a aparecer los fundamentos que guiaban a los seres humanos en su extrañísimo discurrir. Se descubría así un mundo extraño y oculto en y para ellos mismos.

Se trataba de un mundo ajeno a ellos pero que los determinaba, que situaba reglas dentro de las cuales podrían incluir elementos expresivos. Se suele pensar que se habla para expresar algo. Y puede que sea esta la dimensión consciente de ese acto. Pero, al mismo tiempo, cuando se habla, se lo hace al interior de ciertas reglas.

La cuestión es que esas reglas no están dadas. Se van produciendo en el mismo acto del habla. Por eso, también cabe precisar que hablar quizá no sea principalmente expresar algo. Quizá se trate más bien de producir o descubrir, antes que un significado, una regla; una regla del hablar (Foucault, 1988). Tal vez, el sentido oculto, aunque fundamental, de hablar sea producir algo del orden de la impropiedad propia del habla que se produce en los animales humanos.

Es por eso que los seres humanos pueden notar que, cuando hablan, se percibe una extrañeza en la operación. No solamente son ellos los que hablan, sino que también se tiene la sensación de que algo habla en ellos o, mejor, a través de ellos (Lacan, 2002a, p. 267). Son ellos los que ejercen la acción de hablar; —no cabe duda al respecto—. Pero es como si algo los hiciese expresarse tan solo para que, en ese hablar, la regla sea producida, descubierta, o vuelta consciente para ellos. En el hablar también se producen las reglas del hablar.

2.2. Un mundo de reglas

Se fue gestando así, para los animales humanos, un mundo de reglas ajeno a ellos, impropio, no diseñado por ellos, pero aparecido gracias a ellos;

un mundo de reglas impropio [1] pero que al mismo tiempo resulta ser donde pueden reconocerse —aunque extrañados— y reforzar su diferencia con los otros seres. Este universo reglado constituiría otra dimensión de la fusión entre animalidad y racionalidad que no es tanto la suma de ambas, sino más bien la síntesis imperfecta y fallida de un conflicto entre esas dos dimensiones que ya no están porque fueron disueltas en esa intersección. Ya no hay animalidad en el ser humano (el psicoanálisis lo sabe muy bien). Ya no hay racionalidad en el ser humano (la inteligencia artificial lo sabe muy bien). Lo que hay es un mundo de reglas impropias y en las que el ser humano puede reconocerse, aunque extrañado. Hay algo del orden de lo divino en donde puede reconocerse; algo que está más allá de su comprensión y que, sin embargo, le incumbe y le interpela.

El mundo jurídico, en sus primeras formulaciones de las que tenemos noticia, también respondían a este campo. La etimología de las fuentes mesopotámicas vincula las normas jurídicas con lo divino e incluso con el destino. La impropiedad y como la interpelación a sentirse allí tanto reconocido como extrañado es evidente (Lara Peinado y Lara González, 2009, p. XIII). De hecho, las primeras leyes mesopotámicas, como el Código de Hammurabi, estaban caracterizadas como permanentes e inmutables, siendo acompañadas de maldiciones humanas y divinas para quienes las modificaran o eliminaran (Lara Peinado, 2008, pp. 47-52).

De todas formas, como bien señala Boyer (1951-1952, pp. 5-11), las primeras leyes mesopotámicas no se ocuparon de definiciones o designaciones abstractas, sino que se enfocaron en los detalles, complicaciones y contradicciones de los casos concretos. A partir de aquellos primeros principios de orden divino, se ofrecieron una colección de fórmulas, hipótesis relativas como absolutas o de previsiones de rectitud bajo la lógica de la causa y el efecto, presentando aquello que debía hacerse (Pintore, 1988, p. 451). Todo ello respondía a principios ligados a la justicia divina pero también a criterios psicológicos e históricos absolutamente ligados a las necesidades prácticas de su tiempo.

3. Consideraciones sobre el derecho y el retorno de lo divino mediante la inteligencia artificial

Si referimos a los vínculos impolutos y absolutos entre lo divino y las primeras formulaciones jurídicas de las que tenemos conocimiento es porque, en cierto sentido, la implementación de la inteligencia artificial en el derecho del siglo XXI constituye un gesto similar que encuentra sus raíces en aquellos primeros tiempos, haciendo caso omiso de los distintos movi-

(1) Aunque la referencia aquí es al campo del lenguaje, no deja de advertirse la importancia del sentido que Rodrigo Ottonello le imprime a un universo de problemas ligados a la sociedad, entre los cuales bien podría considerarse el de la tecnología (Ottonello, 2016).

mientos que la historia de la humanidad y del derecho en particular hicieron para tratar de situar a lo jurídico en el complejo campo de lo humano.

Aun cuando lo divino siempre haya estado presente en estos vaivenes en donde el ser humano comenzó a hacerse cargo de su concepción de la justicia, cierto es que la diferencia fundamental que estamos situando aquí con respecto a aquellos primeros orígenes, así como con la implementación de la IA (y también, se podría decir, con aquella que se suscitó durante la Edad Media), es que en todos estos últimos casos el ser humano parece tender a no querer responsabilizarse no solo de la aplicación del derecho sino de su formulación y de los principios que deberían regirla. Aun cuando los algoritmos claramente otorguen una direccionalidad al derecho, es como si se quisiera invisibilizarla en una totalidad absoluta y tecnológica (Gulman, 2023).

Tanto en las leyes mesopotámicas como en las que están comenzando a regular el mundo digital del siglo XXI, la justicia parece provenir de una dimensión ajena a las vicisitudes del mundo humano; en aquel tiempo, de las religiones de la época y de la geografía del caso, y ahora de las dimensiones lógico-matemáticas que trabajan algorítmicamente. Dimensiones ajenas, que están más allá de la comprensión de los seres humanos, divinas en este sentido y, por eso, intolerables por y en sí mismas. El soporte es fundamental; y no solo el técnico.

Lacan señala que "(...) solo podría llamarse alma lo que permite a un ser —al ser que habla, para darle su nombre— soportar lo intolerable de su mundo, lo cual la supone ajena a este, es decir, fantasmática" (Lacan, 2010, p. 10).

¿Acaso la avanzada de la inteligencia artificial por los diversos ámbitos que atraviesan el ser humano es tendiente a conformar un soporte —además del técnico— para lo intolerable del mundo? ¿O se está entregando la carne humana, sin más, a la servidumbre del nuevo Amo (algo que ni el cristianismo se había atrevido ni querido hacer)?

Las preguntas siguen siendo las mismas. Habrá que insistir como se lo hizo tantas veces. ¿Quién está dispuesto a luchar por amor? ¿Quién está dispuesto a pelear por lo que ya no vale nada? ¿Quién se va a ensuciar por lo que nunca le va a pertenecer?

BIBLIOGRAFÍA

Aristóteles (1954). *Ética Nicomaquea*, UNAM.

Aristóteles (1988). *Política*, Gredos.

Beresñak, F. (2017). *El imperio científico. Investigaciones político-espaciales*, Miño y Dávila editores.

Boyer, G. (1951-952). "De la science juridique et de sa méthode dans l´Ancienne Mésopotamie", *Semitica*, París, 4, pp. 5-11.

Foucault, M. (1988). *El pensamiento del afuera*, Pre-Textos.

Gulman, P. S. (2023). "La autoridad en la totalidad tecnológica: perspectivas filosófico-políticas sobre el fenómeno del companion chatbot", en *Cuadernos Del Sur Filosofía*, 51, pp. 99–118. Recuperado a partir de https://revistas.uns.edu.ar/csf/article/view/4279

Hout, J. L. (1990). "La ville sumérienne, une cité-état", en Hout et al., *Naissance des cités*, París.

Kramer, S. N. (1974). "Kingship in Sumer and Akkad: The ideal King", en *Le palais et la royauté*, XIXe REnc. Asyriol. Intern., 1971.

Lacan, J. (2002a). "Función y campo de la palabra y el lenguaje en psicoanálisis", en *Escritos 1*, Siglo XXI.

Lacan, J. (2002b). "La instancia de la letra en el inconsciente o la razón desde Freud", en *Escritos 1*, Siglo XXI.

Lacan, J. (2005). "El estadio del espejo como formador de la función del yo [*je*] tal como se nos revela en la experiencia psicoanalítica", en *Escritos 1*, Siglo XXI.

Lacan, J. (2010). *El seminario de Jacques Lacan. 20: aún (1972-1973)*. Paidós.

Lara Peinado, F. (1989). *La civilización sumeria*, Madrid.

Lara Peinado, F. (2008). *Código de Hammurabi*, Madrid.

Lara Peinado, F. y Lara González, F. (2009). "Estudio Preliminar", en *Los primeros Códigos de la humanidad*, Tecnos.

Ludueña Romandini, F. (2010). *La comunidad de los espectros I. Antropotecnia*. Miño y Dávila editores.

Ludueña Romandini, F. (2025). *H. P. Lovecraft. La Anti-vida y el destino cósmico. Ontología analéptica IV*. Miño y Dávila editores.

Marx, K. (1990). "La ideología alemana", en Canals, F., *Textos de los grandes filósofos: edad contemporánea*, Herder.

Ottonello, R. (2016). *La destrucción de la sociedad: Política, crimen y metafísica desde la sociología de Durkheim*, Miño y Dávila editores.

Pintore, F. (1988). "La estructura jurídica", en Moscati, S. (dir.), *El alba de la civilización*, I, Madrid.

Inteligencia artificial y principio de responsabilidad: más allá del derecho a la explicación

Por Luciano D. Laise [(*)] *y Eduardo E. Magoja* [(**)] [(***)]

Resumen: En este artículo analizamos los desafíos éticos derivados del uso de la inteligencia artificial (IA), con especial énfasis en la cuestión de la responsabilidad humana frente a sistemas complejos, opacos y de difícil comprensión, conocidos como "cajas negras". Examinamos cómo los algoritmos reconfiguran los modos de procesar información y tomar decisiones, lo que genera resultados que no siempre son comprensibles ni predecibles. En este contexto, cuestionamos la efectividad del derecho a la explicación como mecanismo para atribuir responsabilidad, ampliamente reconocido en la literatura, y mostramos que, debido a la complejidad técnica y a la opacidad de los sistemas, dicho derecho resulta insuficiente y en muchas ocasiones impracticable. Frente a estas dificultades, replanteamos la noción de control y responsabilidad, considerando la IA como un instrumento complementario que requiere supervisión y juicio humanos. Para afrontar los riesgos asociados, proponemos estrategias orientadas a la mitigación y gestión de los efectos de los sistemas, tales como el derecho al olvido en cuanto herramienta para intervenir en los datos que alimentan los algoritmos y el uso de explicaciones contrafácticas para evaluar cómo habrían variado los resultados ante diferentes insumos. Nuestro enfoque subraya que la reflexión ética sobre la IA debe centrarse en comprender sus límites y posibilidades, fortalecer la supervisión humana y priorizar medidas preventivas, con el fin de asegurar que la adopción de estos sistemas sea responsable, transparente y compatible con principios de justicia y equidad.

(*) Profesor y jefe del área de Fundamentos de la Facultad de Derecho, Universidad de Piura. Doctor en Derecho (U. Austral, Argentina). Abogado (UBA). Premio Tesis Sobresalientes "Vélez Sarsfield" de la Academia Nacional de Derecho y Ciencias Sociales, Córdoba. Profesor de Filosofía del Derecho y Metodología de la Investigación en posgrados de la Universidad Austral, Universidad Católica de Salta, Universidad Católica de Santa Fe, Universidad de Palermo, Universidad del Rosario y Universidad de La Sabana (Colombia).

(**) Abogado, licenciado en Filosofía, magíster en Filosofía del Derecho y doctor en Derecho por la Universidad de Buenos Aires. Se desempeña como docente del Departamento de Filosofía y Teoría General del Derecho de la Facultad de Derecho de la Universidad de Buenos Aires.

(***) El orden de los autores resulta estrictamente alfabético.

Palabras clave: Inteligencia artificial; explicabilidad; responsabilidad ética; toma de decisiones.

Abstract: In this article, we analyze the ethical challenges arising from the use of artificial intelligence (AI), with a particular emphasis on human responsibility in relation to complex, opaque, and hard-to-understand systems, commonly referred to as "black boxes". We examine how algorithms reconfigure the ways in which information is processed and decisions are made, producing outcomes that are not always understandable or predictable. In this context, we question the effectiveness of the right to explanation as a mechanism for assigning responsibility —a concept widely recognized in the literature— and show that, due to the technical complexity and opacity of these systems, this right is often insufficient and, in many cases, impracticable. In response to these challenges, we reconsider the notion of control and responsibility, viewing AI as a complementary tool that requires human supervision and judgment. To address the risks involved, we propose strategies aimed at mitigating and managing the effects of these systems, such as the right to be forgotten as a means to intervene in the data that feed algorithms, and the use of counterfactual explanations to evaluate how outcomes might have differed under alternative inputs. Our approach emphasizes that ethical reflection on AI should focus on understanding its limitations and possibilities, strengthening human oversight, and prioritizing preventive measures to ensure that the adoption of these systems is responsible, transparent, and consistent with principles of justice and fairness.

Keywords: Artificial intelligence, explainability, ethical responsibility, decision-making

1. Introducción

Vivimos en una época atravesada por el giro algorítmico, una dimensión central de lo que se ha denominado la Cuarta Revolución Industrial —noción popularizada por Schwab (2016), fundador del Foro Económico Mundial—. Aún más, esto se inscribe en un extenso proceso de transformaciones técnicas y sociales que algunos autores han caracterizado, hasta ahora, como la Sexta Revolución Industrial (Knell, 2020; Di Nardo & Yu, 2021). No falta quien sostiene que estamos ante la Séptima Revolución Industrial (Ruiz Estrada, 2024).

Como sea, el período actual supone una reconfiguración profunda de las formas en que se representa, organiza y gobierna la realidad social. Proliferan los sistemas tecnológicos diseñados para anticipar resultados inciertos o identificar patrones en datos aún desconocidos, pero que pueden ser

ordenados y clasificados mediante inferencias automatizadas (Edwards & Veale, 2017, p. 24).

El uso de la inteligencia artificial (en adelante, IA) supone múltiples desafíos éticos que se han ido desarrollando en la literatura especializada y en algunos cuerpos normativos de los distintos sistemas jurídicos. Por ejemplo, las directrices éticas para una IA fiable de la Comisión Europea (2019), cuyos parámetros, más allá de ciertos defectos, han sido ampliamente reconocidos en la literatura especializada como puntos de partida valiosos en la materia (Hickman & Petrin, 2021), establecen cuatro principios éticos en el contexto de los sistemas de IA: 1) respeto por la autonomía humana; 2) prevención del daño; 3) equidad; y 4) explicabilidad.

El primer principio remite al problema de evitar que la IA se utilice para coaccionar o manipular las preferencias o decisiones de las personas humanas. La presentación de información sesgada podría funcionar como un mecanismo que pretende incidir sobre las elecciones de las personas. El segundo principio exige evitar que la IA sea utilizada para dañar a las personas, lo que supone salvaguardar su integridad física y moral y descartar cualquier empleo como arma contra ellas. La equidad, por su parte, refiere tanto a su aspecto sustantivo como adjetivo. Por ejemplo, ¿se incrementarán las brechas de desigualdad entre ricos y pobres en virtud de que unos tendrán acceso a sistemas algorítmicos y otros no? Y, por último, pero no menos importante, el principio de la explicabilidad significa que los sistemas de inteligencia artificial deben ser comprensibles y transparentes, de manera que se pueda entender cómo y por qué se llegó a una decisión o recomendación (de Zárate Alcarazo, 2022, p. 332).

La explicabilidad es el medio técnico que hace posible el derecho a la explicación, uno de los instrumentos más empleados para atribuir responsabilidad. Se trata de un correlato jurídico que luce como muy sensato y necesario ante el deber de transparencia algorítmica, es decir, la habilidad de hacer visible cada una de las piezas que componen la estructura y funcionamiento de un sistema de IA (de Zárate Alcarazo, 2022, p. 334).

El problema que da origen a la reflexión de este trabajo es que la explicabilidad constituye uno de los mayores desafíos en materia de inteligencia artificial, pues obliga a repensar el concepto mismo de responsabilidad, categoría esencial de la ética. La cuestión puede plantearse de manera sencilla: nos encontramos en un escenario en el que la responsabilidad tiende a diluirse, ya que no está claramente delimitado el tipo de control efectivo que los seres humanos ejercemos sobre los sistemas de IA. En este sentido, la reflexión ética sobre su uso requiere precisar hasta qué punto la inteligencia artificial sigue siendo un instrumento bajo control humano o, por el contrario, comienza a configurarse como una instancia con márgenes

de autonomía que tensionan nuestras concepciones tradicionales de la responsabilidad.

La primera respuesta que defenderemos es la siguiente: tenemos un escaso control sobre las soluciones que nos brinda la IA, pero todavía un pleno dominio sobre esta. La IA, por ende, es tan solo capaz de funcionar como un complemento en un proceso de decisión. Ahora, como la IA no puede operar como una mente humana con plena capacidad de agencia y racionalidad práctica, entonces quienes le encomendamos decisiones tenemos que hacernos cargo de ellas. Por esta razón, debemos ser cautelosos al momento de incorporarlas y emplearlas. Con todo, no se trata de sospechar de la IA, sino de conocer lo más exhaustivamente posible cómo funciona y cómo ha sido programada para ser capaz de determinar tanto sus posibilidades como limitaciones.

Nuestra reflexión sobre la responsabilidad ética en este campo pondrá el foco en la adopción de la IA por parte de la Administración pública (en lo que sigue AP). En el caso del derecho administrativo que se abordará más adelante, si la AP decide adoptar un algoritmo para delegar un tramo de sus decisiones, parece entendible que sea capaz de explicar al administrado la manera en que arribó a su resolución. Sin embargo, como veremos, se trata de una cuestión tan sensata como impracticable. De esta manera, relativizaremos la significatividad y relevancia de un derecho a la transparencia algorítmica o a la explicación. En cambio, propondremos mecanismos *ex ante* o preventivos para remediar mucho las dificultades o inconvenientes que podría suscitar la aplicación de la IA. Con otras palabras, frente a la adopción de la IA, cabe aplicar plenamente el adagio de "es mejor prevenir que curar".

Con el propósito de fundamentar las ideas aquí expuestas, en primer lugar, delimitaremos conceptualmente la noción de inteligencia artificial, a fin de establecer el marco teórico de referencia. En un segundo momento, analizaremos sus implicancias éticas y, luego, examinaremos críticamente las expectativas, a menudo ingenuas, que han sido depositadas en el denominado derecho a la explicación. Asimismo, en un tercer movimiento, presentaremos el núcleo de nuestra contribución al debate, orientado a problematizar la relación entre explicabilidad, responsabilidad y control en el ámbito de estas tecnologías. En un nivel netamente filosófico, lo que proponemos es redefinir el sentido o significado de la noción de control con relación a la responsabilidad. En breve, no tenemos responsabilidad sobre la IA porque conocemos en detalle cómo funciona para elaborar sus respuestas, sino precisamente porque sabemos o debiéramos saber que eso resulta extremadamente difícil. Por último, esbozaremos dos respuestas o vías alternativas para solucionar los problemas que plantea un uso incorrecto de la IA.

2. ¿Qué es la inteligencia artificial?

El punto de partida para el estudio de la IA, a nuestro modo de ver, ha de ser el agente inteligente. La IA puede entenderse como un campo dedicado al análisis de agentes capaces de asimilar percepciones del entorno en el que se encuentran e intervenir sobre él mediante acciones orientadas a la consecución de determinados objetivos (Russell & Norvig, 2004, pp. 5-6). Lo peculiar de ella es que realiza tales acciones con un relativo grado de autonomía. Las máquinas, pues, son capaces de alcanzar cierto grado de racionalidad por medio del análisis de la interacción con el ambiente circundante, a través de mecanismos que les permiten captar y analizar datos, razonar a partir de estos, con el propósito de alcanzar un tipo de acción o decisión.

Ahora bien, existen dos grandes escenarios en los que podría operar la IA: por un lado, la IA débil y, por el otro, la IA fuerte. La primera —también conocida como inteligencia artificial estrecha— se refiere a sistemas diseñados para realizar tareas específicas que, si bien pueden imitar algunas capacidades humanas, no implican una verdadera comprensión ni conciencia (Bartneck *et al*., 2021, p. 10). Entre estas habilidades se encuentran, por ejemplo, la percepción visual, la comprensión limitada del contexto, el razonamiento probabilístico y el tratamiento de problemas complejos en entornos acotados. Este tipo de IA no posee intencionalidad ni conciencia; simplemente responde a patrones y datos de entrada con base en algoritmos previamente definidos o entrenados. No entiende en un sentido profundo de la palabra, sino que simula entender. Con todo, la IA débil podría constituir un instrumento muy valioso para estudiar el funcionamiento de la mente humana, tal como lo pensó el filósofo estadounidense Searle (1980). Su planteo anticipó uno de los usos más relevantes y actuales de la IA débil: el de copiloto, asistente o complemento para la actividad humana, sin reemplazarla ni emular plenamente la conciencia (Searle, 1980, p. 417).

La inteligencia artificial robusta o fuerte presenta desafíos más profundos para la ética y la filosofía de la mente. Se refiere a un agente capaz de pensar y entender el mundo de forma análoga a un ser humano. Esto implica la habilidad para razonar, comprender y enfrentar tareas nuevas tal como lo haría una persona (Criado, 2021, p. 352). En otras palabras, esta IA no funcionaría simplemente como un asistente, sino que operaría como una auténtica mente humana.

Pero esto no es más que un escenario de ciencia ficción, al menos por ahora, porque la IA no ha superado la fase débil. Incluso los avances de ChatGPT y otros modos conversacionales no alcanzan una autonomía ni una racionalidad comparables a las de una persona de carne y hueso. La IA capaz de igualar la inteligencia humana en todos los aspectos todavía no ha sido alcanzada, aunque hay cierto optimismo al respecto (Butz, 2021). Existe un intenso debate filosófico sobre si la IA podrá algún día emular de

manera plena la mente humana. Algunos sostienen que todavía estamos a tiempo de aprovecharla como herramienta para automatizar procesos que complementen el trabajo humano (Pasquale, 2020, p. 1). En fin, todavía podemos frenar a la *Matrix* o, al menos, salvar a John Connor.

No hay dudas de que semejante discusión revela un interesante campo de estudio en donde se entrecruza la tecnología, la filosofía política, la filosofía de la mente y la ética filosófica. Sin embargo, dejaremos este tema de lado para enfocarnos en las implicaciones éticas de la IA débil, que es la que actualmente plantea los desafíos más urgentes. Por ahora, la reflexión sobre la inteligencia artificial robusta permanece como un mero experimento mental, sin aplicación práctica.

2.1. De la responsabilidad basada en el control a la ilusión del control

Desde Aristóteles (*Ética Nicomaquea*, 1109b30-1111b5) tenemos arraigada una intuición fundamental sobre la responsabilidad: quien tiene control sobre una acción es responsable de ella. Dicho de otra forma, el juicio de responsabilidad moral presupone que el agente domina con suficiente autonomía su conducta, de manera que sus acciones dependan directamente de su control sobre ellas. Se suele considerar, pues, que esta "condición de control" requiere algún tipo de libertad con respecto a aquello de lo que se es responsable: en el caso de una acción, que la haya realizado de manera libre; y en el caso de una omisión, que hubiera tenido la libertad de realizar la acción omitida (Clarke, 2017, p. 241). No obstante, la cuestión que presenta la inteligencia artificial es lo que se llama el problema de las muchas manos y, a la vez, el de muchas cosas (Coeckelbergh, 2020, pp. 113-114). Por ello, analizar la responsabilidad en contextos de IA se vuelve especialmente complejo.

Para comprender mejor esta cuestión conviene desglosarla en sus diferentes componentes. Existen muchas manos en los asuntos referidos a la IA en el sentido de que existen muchas personas involucradas en las acciones que las máquinas llevan a cabo. Esto comienza desde el programador hasta el usuario final. Pongamos el ejemplo de un algoritmo para determinar qué devoluciones de impuestos serán fiscalizadas por la Agencia de Recaudación y Control Aduanero (ARCA) de Argentina. Tenemos a las personas que diseñaron el algoritmo, a la persona encargada de supervisar el desarrollo de tal algoritmo, al área de ARCA que lo adquirió, al departamento de fiscalizaciones que lo aplica, al contribuyente que fue intimado a brindar información al fisco y al funcionario que recibe los requerimientos del contribuyente, por mencionar solo algunos.

Pero también es un asunto que involucra muchas cosas. La elaboración de los algoritmos constituye apenas una parte de un entramado más amplio, que incluye la compleja tarea de recolección y depuración de datos,

el diseño de las secuencias lógicas para su funcionamiento o ejecución y, por último, la presentación de los resultados obtenidos. En conjunto, estos elementos muestran que la inteligencia artificial no puede reducirse a una simple herramienta técnica, sino que es el resultado de procesos múltiples y entrelazados que demandan atención en cada una de sus fases.

Entonces, del diseño de un algoritmo y un sistema de IA hasta su aplicación concreta, intervienen múltiples personas ("muchas manos"), con diferentes aportes ("muchas cosas"), y estas personas, por lo general, no tienen una idea sustantiva de las contribuciones de cada una de ellas. Esta compleja distribución de responsabilidades recuerda, en cierta medida, la noción de banalidad del mal que Arendt (2000) desarrolló en su análisis del juicio a Adolf Eichmann. La filósofa planteaba que una forma de diluir la responsabilidad frente al holocausto consistía en descomponer el proceso de exterminio en múltiples eslabones. Casi nadie sabía exactamente qué hacía el eslabón previo de la cadena. Cada uno cumplía rigurosamente con su parte del proceso de manera diligente, pero sin conocer, ni interesarse por la dirección última del proyecto.

De un modo similar, la creación de los algoritmos complejos o sistemas de IA está signada por una profunda división del trabajo. Los que se encargan de la fase de recolección de información no siempre saben cómo el sistema interactuará con el ambiente o entorno con el fin de brindar respuestas, ni cómo el sistema será entrenado para mejorar su respuesta a partir de las retroalimentaciones de los usuarios. Todos ponen su contribución para que funcione efectivamente un sistema de IA, pero muchas veces sin conocer la dirección final o última de tales proyectos.

Pero esto no es todo. El escenario es mucho más desafiante si miramos el proceso mismo de la tecnología. En efecto, uno de los retos más complejos en el uso de la inteligencia artificial es que no tenemos un control total sobre ella. Desconocemos —y probablemente seguiremos desconociendo— el modo exacto en que la IA genera las respuestas que nos ofrece. A este fenómeno se le denomina comúnmente "el problema de la caja negra", muy estudiado en la literatura especializada (Gryz, 2021; Bélisle-Pipon *et al.*, 2023; Hassija *et al.*, 2024). Esta metáfora, como señala Pasquale (2015), remite tanto a un dispositivo que registra datos durante un viaje como a un sistema que opera de manera oculta o misteriosa. En esencia, se trata de un mecanismo cuyos insumos y resultados podemos identificar, pero cuyo mecanismo interno de transformación permanece inaccesible (Pasquale, 2015, p. 3).

Las explicaciones, en términos generales, pueden resultar bastante satisfactorias, aunque no logran serlo en un plano particular y concreto. Por ejemplo, supongamos que uno se dirige al banco para solicitar una tarjeta

de crédito. El empleado nos informa que, en virtud de la calificación de riesgo crediticio, nos pueden otorgar X monto como límite en la tarjeta. ¿Por qué ese límite y no otro? La respuesta puede ser absolutamente opaca. No tanto porque no quieran decirnos cómo fue que se llegó a esa precisa suma de dinero, sino porque nadie sabe con precisión de qué manera se arribó a ese resultado.

Así, cabe preguntarse: ¿podemos afirmar que tenemos control sobre los resultados producidos por la IA si resulta extremadamente difícil, incluso para un experto en computación, reconstruir el modo en que el sistema llegó a una determinada conclusión? A nuestro modo de ver, la respuesta es clara. Nuestro control sobre lo que genera la IA, especialmente en el marco de modelos de *machine learning*, se limita a decidir si seguimos o no al resultado que nos ofrece el sistema.

La adopción de un sistema de inteligencia artificial en un proceso de toma de decisiones exige un conocimiento suficiente de su diseño y funcionamiento, porque, de no hacerlo, se estaría delegando un segmento de la decisión a una herramienta cuyo modo de operar y fuentes de información resultan opacos. Al fin y al cabo, puede que uno no conozca en detalle cómo se generan ciertas respuestas, pero al menos es necesario comprender su lógica general de funcionamiento. Esto resulta particularmente relevante si se pretende ubicar responsabilidades dentro de la cadena de desarrollo y aplicación de la tecnología.

Parafraseando al maestro Oogway, el empleo de sistemas de IA que emplean una caja negra nos pone ante la ilusión de control, porque nuestro control se limita a decidir sobre adoptar o no las respuestas particulares que elabora el modelo de IA del cual nos servimos.

2.2. El conocimiento: ¿Mis actos revelan quién soy?

La posibilidad de ejercer control sobre una acción constituye una condición necesaria, pero no suficiente, para atribuir responsabilidad por ella. A este elemento objetivo se añade un componente subjetivo —también reconocido por Aristóteles en el pasaje mencionado— que implica que el agente debe ser consciente de lo que ha hecho. Este requisito es conocido en la literatura como la "condición epistémica" o "condición cognitiva" (Clarke, 2017, p. 241). Las acciones que uno no supo que cometió, porque estaba en un estado de completa inconsciencia no culpable, nos eximen de toda responsabilidad. ¿Por qué esto es así? La responsabilidad supone la posibilidad de dar cuenta de lo que se hizo. Se trata, pues, de que hemos de ser capaces de responder y explicar el porqué de nuestras acciones (Coeckelbergh, 2020, p. 115).

Así, somos responsables de nuestros actos en la medida en que esto revela un aspecto central de nuestra existencia humana que la filosofía moral

suele resumir como "agencia". Nos constituimos a nosotros mismos como agentes al elegir nuestras acciones de acuerdo con los principios de la razón práctica (Korsgaard, 2012, p. 1). De hecho, la noción de agencia revela cómo nos vinculamos activamente con la realidad a través de lo que hacemos. Entre los diversos actos posibles, aquellos por los que respondemos ocupan un lugar especial, pues expresan nuestro compromiso como agentes racionales. Somos responsables de ellos en la medida en que guardan una relación directa con nuestras capacidades de agencia y de deliberación.

Pero ¿qué significa ser un agente racional? Scanlon (1998, p. 23) sostiene que es, ante todo, un ser que razona: alguien que tiene la capacidad de reconocer, evaluar y dejarse mover por razones, y, por ello, de mantener actitudes sensibles al juicio. Cada acción emprendida, aunque solo con un mínimo de deliberación, refleja un juicio de que determinada razón merece ser tomada como guía de la conducta. Un agente racional, entonces, es capaz de asumir una posición en su propia vida y de ejercer influencia sobre el curso de los acontecimientos. Así, la agencia racional expresa algunos de los aspectos más significativos de la personalidad humana, como sus acciones voluntarias. Dicho de otra manera, la agencia racional resume a la persona como autora de su propia conducta.

Ahora bien, el problema de la caja negra —planteado brevemente en la sección anterior— introdujo un serio problema para la atribución de responsabilidad, porque implica hacerse cargo de acciones sobre las que no tenemos una influencia significativa. De hecho, no podemos dar una serie de razones precisas y determinadas que expliquen cómo es que llegó a la conclusión o respuesta que la IA nos ofrece. Entonces, ¿cómo podríamos hacernos responsables de aquello sobre lo que no tenemos una influencia completa o total?

3. Una respuesta tentativa: el derecho a la explicación

Frente a las cuestiones que venimos señalando, se suele esgrimir una respuesta que suena atractiva conceptual o dogmáticamente, pero tiene enormes dificultades operativas: el derecho a la explicación o transparencia algorítmica. Así, resulta una intuición muy sensata reclamar que se nos explique cómo fue que el algoritmo llegó al resultado o respuesta a la que arribó. Pedir explicaciones a quien es autor de una decisión que nos afecta es una práctica social que brota espontáneamente. Si alguien decidió X sobre algún aspecto de nuestra existencia, parece inevitable que le preguntemos: ¿por qué? Este criterio, de hecho, se refleja en las directrices europeas (Criado, 2021, 359), que mencionamos en la introducción.

Con todo, la explicabilidad no solo supone brindar una respuesta a quien lo demande, sino que también exige que la información le sea presentada de

modo inteligible. Quien solicita una explicación demanda que se le brinde tanto información veraz como comprensible para sus capacidades cognoscitivas. No bastaría con que a una persona se le abran las puertas del algoritmo de par en par, si esto no lleva consigo una posibilidad de comprender lo que el peticionante tiene delante de sí (de Zárate Alcarazo, 2022, p. 334).

Situados en este contexto, cabe preguntarse: ¿será eficaz la pretensión de solicitar explicaciones a la AP por las aplicaciones de IA que efectúa en sus decisiones? Lo cierto es que tal exigencia suele ser sumamente ardua e, incluso, en ocasiones imposible (Boddington, 2023, p. 48). Ello se debe, principalmente, a las dificultades técnicas involucradas: muchas veces se trata de una compleja combinación de bases de datos y de la interacción de los usuarios con las plataformas o sistemas utilizados. Como muchos autores han reconocido, los sistemas de IA suelen generar resultados y decisiones de difícil interpretación y con escasa transparencia (Chazette & Schneider, 2020; González Arencibia *et al*., 2024).

Aún más, los algoritmos que mejor funcionan se apoyan en una variada y amplia gama de datos (Edwards & Veale, 2017, p. 59). Se trata de una gestión de la información que está pensada más en el funcionamiento eficiente del sistema que en la interpretabilidad de los usuarios. Por ejemplo, en el caso de LinkedIn, una de las métricas más relevantes que el algoritmo utiliza es el *dwell time*, es decir, el tiempo que los usuarios permanecen leyendo una publicación, incluso si no interactúan activamente con ella. Esto significa que, cuanto más complejo y robusto es un sistema de IA, mayores son las dificultades para ofrecer explicaciones claras y comprensibles a los usuarios sobre por qué cierto contenido aparece en su *feed* y otro no.

La otra dificultad que presenta el derecho a la explicación consiste en que este podría vulnerar la privacidad de datos personales de terceros. Supongamos que, por ejemplo, X reclama que se le explique en detalle cómo la entidad financiera B calcula la línea de crédito que le ofreció hace unos días. Podría suceder que una parte de los datos en que se basa refiera a los comportamientos erráticos de su cónyuge. O bien podría ocurrir que X trabaja para la empresa Z, la cual ha solicitado un préstamo a B con el fin de salvarse de una quiebra. Así, explicarle al cliente X cómo funciona el algoritmo del área comercial del banco podría conculcar los datos personales de terceros.

4. DOS VÍAS DE SOLUCIÓN

4.1. Una explicación contrafáctica

Una vía particularmente eficaz y accesible para hacer comprensibles los algoritmos consiste en recurrir a explicaciones contrafácticas. Este tipo de explicaciones no busca desentrañar cada paso del cálculo interno, sino

mostrar cómo los resultados podrían variar si se modificaran ciertos elementos o datos de entrada en los que el algoritmo se apoyó (Boddington, 2023, p. 49). De este modo, se traslada el foco desde la complejidad técnica de los procesos hacia la relación práctica entre variables y resultados, lo cual suele ser más significativo para el destinatario de la decisión.

Pensemos nuevamente en el caso del cliente de un banco que solicita una línea de crédito. Una aplicación estricta del derecho a la explicación podría exigir que la entidad financiera detallara el modo en que el sistema computacional procesó la información: qué fórmulas aplicó, qué pesos otorgó a cada variable, cómo se realizaron las correlaciones estadísticas y qué modelos matemáticos sustentaron la decisión final. Sin embargo, ese tipo de explicación, aun cuando fuera completa y precisa, difícilmente resultaría inteligible o de utilidad para el cliente promedio.

En cambio, una explicación contrafáctica desplaza la perspectiva: en lugar de exponer la lógica técnica del algoritmo, ofrece al usuario información práctica y orientada a la acción, mostrándole qué condiciones deberían cambiar para que el resultado fuese distinto. Así, en lugar de revelar la "caja negra" del cálculo, se proporcionan condiciones alternativas claras.

Concretamente, el banco podría comunicar algo como: "Si usted desea acceder a una línea de crédito de US$ 50.000, a diez años, deberá acreditar un ingreso mensual de determinada cuantía y mantener una relación deuda-ingreso dentro de ciertos parámetros". De este modo, el cliente obtiene información relevante para comprender su situación financiera y, sobre todo, para actuar en consecuencia.

En última instancia, este enfoque parte de una premisa central: la explicación no constituye un fin en sí mismo, sino un instrumento al servicio de la comprensión y la toma de decisiones informadas. Explicar un algoritmo no siempre significa revelar la totalidad de su funcionamiento interno, sino más bien ofrecer al afectado una vía clara y procesable para entender cómo sus acciones o circunstancias inciden en los resultados que lo afectan. Esta forma de explicación, más pragmática que exhaustiva, puede resultar especialmente valiosa en contextos donde lo que está en juego no es la curiosidad científica, sino la capacidad real del usuario de orientar su conducta y proteger sus intereses.

4.2. El derecho al olvido

Los problemas derivados de la llamada caja negra muestran que resulta sumamente complejo desentrañar cómo se procesan los datos para generar los resultados que ofrece un algoritmo. En este marco, puede ser más prudente poner el acento no tanto en el funcionamiento interno del sistema, sino en la información que lo alimenta. Así, resulta especialmente relevante

el denominado derecho al olvido, que, como explica Allegri (2022, p. 237), permite a las personas mantener una imagen fiel y actualizada de su identidad, al posibilitar la supresión —o al menos la desindexación— de sus rastros digitales previos en internet. Este derecho se funda en varias circunstancias: cuando los datos ya no son necesarios respecto del propósito para el que fueron recogidos; cuando la persona retira la autorización previamente otorgada para el tratamiento de sus datos; cuando los datos se han procesado sin las garantías legales correspondientes; o cuando la información fue proporcionada por alguien que era menor de edad en el momento de hacerlo (Edwards & Veale, 2017, p. 68).

Pero el derecho al olvido no se limita a la posibilidad de "borrar el pasado digital" de una persona, sino que también busca evitar la acumulación desproporcionada de datos que pueden condicionar injustamente sus oportunidades futuras (Puccinelli, 2016). En un entorno donde los algoritmos se entrenan con grandes volúmenes de información, mantener indefinidamente registros sobre comportamientos pasados puede derivar en perfiles sesgados, exclusión financiera, discriminación laboral o denegación de seguros y servicios básicos. De este modo, el derecho al olvido opera como un contrapeso al derecho a la explicación: mientras este exige transparencia respecto del modo en que se toman decisiones automatizadas, aquel pretende limitar y depurar la base de datos sobre la que tales decisiones se construyen. Lejos de ser contradictorios, ambos derechos deben entenderse como complementarios, pues su interacción permite avanzar hacia un modelo en el que la innovación tecnológica se armonice con la protección de la dignidad, la privacidad y la autonomía de las personas.

5. CONCLUSIONES

El giro algorítmico implica reconocer que los sistemas de inteligencia artificial han llegado para quedarse. Sin embargo, no se trata de tecnologías con un nivel de autonomía que les permita contar con agencia o racionalidad práctica por sí mismas. La IA funciona más bien como un instrumento destinado a agilizar y dinamizar procesos de toma de decisión, pero la responsabilidad última de lo que actúe o decida un sistema siempre recae en una persona, ya sea humana o jurídica.

Ahora bien, los complejos esquemas de funcionamiento que caracterizan a los sistemas actuales, lo que suele denominarse la caja negra, conllevan que en muchos casos resulte extremadamente difícil, cuando no imposible, explicar al usuario final cómo se arribó a una determinada respuesta. Frente a esta dificultad, nos parece que conviene, conforme los argumentos dados a lo largo del trabajo, adoptar alternativas más implementables que el llamado "derecho a la explicación". Como hemos visto, una de ellas son las explicaciones contrafácticas, que no buscan detallar cada paso del

cálculo interno, sino mostrar cómo los resultados habrían variado si ciertos elementos de entrada hubieran sido distintos. De este modo, se desplaza la atención desde la opacidad técnica hacia la relación práctica entre variables y resultados, lo que ofrece al afectado información clara y procesable para orientar su conducta y proteger sus intereses.

La otra alternativa es el derecho al olvido. Si desentrañar la caja negra es una tarea técnicamente compleja o excesivamente onerosa, puede resultar más razonable y eficiente centrar la atención en la calidad, pertinencia y permanencia de los datos que alimentan al sistema.

En conjunto, las explicaciones contrafácticas y el derecho al olvido pueden ofrecer vías más accesibles y efectivas para asegurar transparencia y rendición de cuentas en el uso de IA, sin exigir revelaciones técnicas inabordables para el usuario común ni poner en riesgo derechos de terceros, ni de violar ningún secreto industrial.

6. Referencias bibliográficas

Allegri, M. R. (2022). The Right to Be Forgotten in the Digital Age. En F. Comunello, F. Martire, & L. Sabetta (Eds.), *What People Leave Behind: Marks, Traces, Footprints and Their Relevance to Knowledge Society* (Vol. 7, pp. 237-250). Springer. https://doi.org/10.1007/978-3-031-11756-5_15

Arendt, H. (2000). *Eichmann en Jerusalén: Un estudio sobre la banalidad del mal*. Lumen.

Aristóteles. (2015). *Ética Nicomaquea* (E. Sinott, Trad.). Colihue.

Bartneck, C., Lütge, C., Wagner, A. & Welsh, S. (2021). *An Introduction to Ethics in Robotics and AI*. Springer. https://doi.org/10.1007/978-3-030-51110-4

Bélisle-Pipon, J.-C., Monteferrante, E., Roy, M.-C. & Couture, V. (2023). Artificial Intelligence Ethics Has a Black Box Problem. *AI and Society, 38*(4), 1507-1522.

Boddington, P. (2023). *AI Ethics: A Textbook*. Springer. https://doi.org/10.1007/978-981-19-9382-4

Broome, J. (2013). *Rationality Through Reasoning.* Oxford University Press.

Butz, M. V. (2021). Hacia una IA fuerte. *Künstliche Intelligenz, 35*(1), 91-101. https://doi.org/10.1007/s13218-021-00705-x

Chazette, L., & Schneider, K. (2020). La explicabilidad como requisito no funcional: Desafíos y recomendaciones. *Requisitos de Ingeniería, 25*(4), 493-514.

Clarke, R. (2017). Ignorance, Revision, and Commonsense. En P. Robichaud & J. W. Wieland (Eds.), *Responsibility: The Epistemic Condition*. Oxford University Press. https://doi.org/10.1093/oso/9780198779667.003.0013

Coeckelbergh, M. (2020). *AI Ethics*. MIT Press. https://doi.org/10.7551/mitpress/12549.001.0001

Comisión Europea. (2019). *Directrices éticas para una IA fiable*. Grupo de Expertos de Alto Nivel en Inteligencia Artificial. https://ec.europa.eu/futurium/en/ai-alliance-consultation

Criado, J. I. (2021). Inteligencia artificial (y administración pública). *Eunomía. Revista en Cultura de la Legalidad, 20*, 348-372. https://doi.org/10.20318/eunomia.2021.6097

de Zárate Alcarazo, L. O. (2022). Explicabilidad (de la inteligencia artificial). *Eunomía. Revista en Cultura de la Legalidad, 22*, 328-344. https://doi.org/10.20318/eunomia.2022.6819

Di Nardo, M., & Yu, H. (2021). Industry 5.0: The Prelude to the Sixth Industrial Revolution. *Applied System Innovation, 4*(3), 45.

Edwards, L. & Veale, M. (2017). Slave to the Algorithm? Why a "Right to an Explanation" Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review, 16*(1), 18-84. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1315&context=dltr

González Arencibia, M., Ordoñez-Erazo, H., & González-Sanabria, J.-S. (2024). Inteligencia artificial explicable como principio ético. *Ingeniería, 29*(2), e21583.

Gryz, J. & Rojszczak, M. (2021). Black Box Algorithms and the Rights of Individuals: No Easy Solution to the "Explainability" Problem. *Internet Policy Review, 10*(2). https://doi.org/10.14763/2021.2.1564

Hassija, V., Chamola, V., Mahapatra, A., et al. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cognitive Computation, 16*(1), 45-74. https://doi.org/10.1007/s12559-023-10179-8

Hickman, E. & Petrin, M. (2021). IA fiable y gobernanza corporativa: Directrices éticas de la UE para una inteligencia artificial fiable desde la perspectiva del derecho de sociedades. *European Business Organization Law Review, 22*(3), 593-625. https://doi.org/10.1007/s40804-021-00224-0

Knell, M. (2020). Anticipating the Sixth Industrial Revolution. *Forskningspolitikk*. https://www.fpol.no/6th-revolution/

Korsgaard, C. M. (2012). *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199552733.001.0001

Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

Pasquale, F. (2020). *New Laws of Robotics: Defending Human Expertise in the Age of AI*. Harvard University Press. https://doi.org/10.4159/9780674250062

Puccinelli, O. R. (2016). El derecho al olvido en el derecho a la protección de datos: Con especial referencia a su vigencia en Internet. *Pensamiento constitucional*.

https://revistas.pucp.edu.pe/index.php/pensamientoconstitucional/article/download/18707/18948/

Ruiz Estrada, M. A. (2024). The Seventh Industrial Revolution (IR 7.0). *SSRN*. https://doi.org/10.2139/ssrn.4720983

Russell, S. J. & Norvig, P. (2004). *Inteligencia Artificial: Un Enfoque Moderno*. Pearson Educación.

Scanlon, T. M. (1998). *What We Owe to Each Other*. Harvard University Press.

Schwab, K. (2016). *La Cuarta Revolución Industrial*. Debate.

Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences, 3*(3), 417-424. https://doi.org/10.1017/S0140525X00005756

Ética(s) de la inteligencia artificial y derecho: consideraciones a propósito de los límites y la contención del desarrollo tecnológico

Por Fernando H. Llano Alonso (*)

A María Dolores García Cossío (1938-2024)
In memoriam

Resumen: El presente trabajo se centra en la ética de la IA y en los principios que la inspiran. Antes de las normas que regulan la IA es necesario conocer las cuestiones éticas relacionadas con el desarrollo, la implementación y el uso responsable de los sistemas de IA. El concurso de estos fundamentos éticos es necesario para proteger los derechos y la dignidad de las personas, garantizar la equidad en el acceso y tratamiento de datos, minimizar los sesgos y riesgos asociados con los algoritmos, promover la transferencia en las decisiones automatizadas y auspiciar la confiabilidad en la tecnología, el beneficio humano y el bienestar social. También se propone el estudio de la ética de la IA desde un punto de vista omnicomprensivo que integre tanto la ética aplicada como la metaética de la IA.

Palabras clave: Sesgos; transparencia algorítmica; explicabilidad; responsabilidad; ciberseguridad; solidez; metaética de la IA; IA generativa; modelos del lenguaje conversacionales.

Ethics(s) of artificial intelligence and law: considerations on the limits and containment of technological development

Abstract: This paper focuses on the ethics of AI and the principles behind it. Prior to the rules governing AI, it is necessary to understand the ethical issues related to the development, implementation and responsible use of AI systems. These ethical foundations are necessary to protect the rights and dignity of individuals, ensure fairness in data access and processing, minimise biases and risks associated with algorithms, promote transferability in automated decisions, and promote trustworthiness in technology, human benefit and social welfare. It is also proposed to study the ethics

(*) Catedrático de Filosofía del Derecho en la Facultad de Derecho de la Universidad de Sevilla [España], de la que actualmente es decano. Email: llano@us.es

of AI from an all-encompassing point of view that integrates both applied ethics and metaethics of AI.

Keywords: Biases; algorithmic transparency; explainability; accountability; cybersecurity; robustness; AI meta-ethics; generative AI; conversational language models.

1. INTRODUCCIÓN

A finales del mes de marzo de 2023, más de medio millar de académicos, ingenieros expertos en inteligencia artificial (en adelante, IA) y empresarios de compañías tecnológicas firmaron una carta abierta en la que advertían de los "profundos riesgos para la sociedad y la humanidad" que plantean los modelos avanzados; en consecuencia, proponían a los laboratorios de IA una pausa o *impasse*, al menos durante seis meses, en el entrenamiento de sistemas de IA más potentes que el modelo de lenguaje de OpenAI GPT4 (1).

Esta carta fue publicada por *The Future of Life Institute*, un instituto que se presenta a sí mismo como una organización cuya principal misión consiste en dirigir la tecnología transformadora hacia el beneficio de la vida y alejarla de los riesgos extremos a gran escala: la IA, la biotecnología y las armas nucleares. Según los autores de este texto, los sistemas de IA con inteligencia competitiva humana pueden representar un cambio significativo en la historia de la vida en la Tierra, por eso, conforme a los Principios de IA de Asilomar, se sostiene que la IA avanzada debe "planificarse y gestionarse con el cuidado y los recursos adecuados" (2).

Para algunos tecnólogos especializados en el estudio del riesgo de extinción que entraña la IA, pecan de ingenuidad quienes esperan que empresas como OpenAI, Google o Meta van a detener sus investigaciones, renunciando así a los beneficios y las ventajas que les proporciona el desarrollo de sus modelos; pero aun suponiendo que esta hipótesis pudiera llegar a ser real, se preguntan estos críticos, ¿cómo respondería el gobierno chino o sus empresas, por poner un ejemplo, ante ese compás de espera de sus competidores? Probablemente, aprovecharía la oportunidad para acelerar en el desarrollo de los sistemas avanzados de IA (3).

Lejos de traer un invierno tecnológico para la IA generativa, lo cierto es que la carta solicitando la pausa de los laboratorios de IA sirvió para avivar

---

(1) Entre los signatarios de esta carta se encuentran: Elon Musk (CEO de SpaceX, Tesla y Twitter); Steve Wozniak (Cofundador de Apple); Yuval Noah Harari; Evan Sharp (Cofundador de Pinterest); Craig Peeters (Director ejecutivo de Getty Images); Emad Mostaque (Director ejecutivo de Stability AI); o Valerie Pisano (Presidenta y Directora ejecutiva de MILA).

(2) https:// futureoflife.org/open-letter/pause-giant-ai-experiments/

(3) M. SIGMAN y S. BILINKIS, *Artificial. La nueva inteligencia y el contorno de lo humano*, Debate, Barcelona, 2023, p. 192.

el proceso de avance tecnológico en los meses siguientes. En efecto, laboratorios de IA y empresas tecnológicas como OpenAI han apostado por lanzar al mercado nuevas versiones de modelos de aplicaciones conversacionales más avanzadas y con mayor capacidad.

En los meses posteriores a la publicación de esta carta por *The Future of Life Institute* se produjeron avances significativos en el diseño y desarrollo de nuevos modelos de IA en los principales gigantes tecnológicos (*Big Techs*), así por ejemplo: Google desplegó *Gemini*, un nuevo modelo de lenguaje (multimodal, en tres versiones: Ultra, Pro y Nano) que podrá rivalizar con la aplicación conversacional Chat GPT-4; Amazon también participó en esta carrera tecnológica al anunciar la inversión de 4000 millones de dólares en Anthropic una *startup* estadounidense de IA generativa que se ha convertido en uno de los más importantes proveedores mundiales de modelos básicos de IA, y que se presenta a sí misma como un "destacado defensor" de la implementación responsable de IA; también Elon Musk, solo cuatro meses después de haber firmado la petición de pausar la IA, en aras de la supervivencia de la humanidad, lanzó xAI, un chatbot para competir con ChatGPT de OpenAI y Bard de Google.

La súbita expansión de la IA generativa y el interés generalizado que se ha despertado entorno a los chatbots y otras herramientas de IA generativa, ha despertado un cierto sentimiento combinado de temor y neoludismo en muchos sectores profesionales que ven su futuro laboral comprometido ante la hipótesis (cada vez más verosímil) de que los robots empiecen a sustituirles. Un caso emblemático de esta desconfianza hacia la IA generativa es el de la huelga convocada el 2 de mayo de 2023 por el Sindicato de Guionistas de Estados Unidos (WGA, por sus siglas en inglés), que representa a cerca de 12.000 escritores que trabajan en el mundo del espectáculo (teatro, cine y televisión), y entre cuyas principales reivindicaciones estaba, precisamente, la de la regulación de la IA y la tecnología entrenada con conjunto de datos.

Muestra de lo fundada que está la inquietud de los guionistas y escritores de Hollywood es un estudio encargado por la propia OpenAI sobre el impacto el mercado laboral de los grandes modelos del lenguaje (LLM, por sus siglas en inglés), publicado el 17 de marzo de 2023, situaba a los escritores en la categoría de profesionales "totalmente expuestos" ante la IA generativa. Según este informe, un LLM es capaz de reducir en un 50% el tiempo que tarda un humano en terminar su tarea, por ejemplo, un guión de cine o una novela [(4)]. La batalla de la productividad de los guionistas humanos frente a la IA generativa estaría perdida, pero no así el debate en torno a la originalidad y el crédito de los escritores, de tal forma que se reconozca al menos que

(4) https://openai.com/research/gpts-are-gpts

ninguna empresa de entretenimiento esté autorizada legalmente para obligar a los creadores de contenidos que utilicen *software* de IA —por ejemplo, ChatGPT— para realizar su tarea.

Cinco meses después del inicio de la huelga convocada por el sindicato de guionistas, estos consiguieron que en el quinto punto el acuerdo de negociación entre la WGA y los estudios de Hollywood se reconociera que "la IA no puede escribir ni reescribir material literario, y el material generado por la IA no se considerará material de origen", ya que una IA no puede considerarse un escritor [(5)].

Pese a la nula eficacia de la carta abierta en la que los signatarios pedían una pausa de la IA generativa a la vez advertían el riesgo potencial que supone para la humanidad el desarrollo de sus modelos más avanzados, solo dos meses más tarde, en mayo de 2023, un grupo de ingenieros, investigadores y altos ejecutivos de la industria de la IA volvieron a expresar su preocupación por la tecnología de IA que estaban desarrollando, sobre todo en relación con los modelos grandes de lenguaje, el tipo de sistema utilizado por ChatGPT y otros chatbots. Para los firmantes de esta declaración, la IA podría llegar a ser algún día tan poderosa como para ser utilizada para difundir desinformación y propaganda, para eliminar millones de puestos de trabajo tradicionalmente realizados por seres humanos, e incluso convertirse algún día en una auténtica amenaza existencial para la humanidad. Por ello, los signatarios decidieron hacer pública una declaración, publicada por el *Center for AI Safety*, en la que se instaba a:

> "Mitigar el riesgo de extinción de la inteligencia artificial debería ser una prioridad mundial junto a otros riesgos a escala social, como las pandemias y las guerras nucleares [(6)]".

En esta frase, tan escueta como impactante, se alude a la necesidad de que haya un consenso acerca del peligro que se cierne sobre la humanidad, pero también a la dificultad de trasladar esta preocupación a acciones concretas y a ponerse de acuerdo sobre qué formas concretas podría tomar ese peligro y qué medidas podrían adoptarse para protegernos de las amenazas que entraña la IA [(7)].

---

(5) https://www.wgacontract2023.org/the-campaign/summary-of-the-2023-wga-mba

(6) https://www.safe.ai/statement-on-ai-risk: entre los firmantes de esta declaración se encuentran directores ejecutivos de empresas tecnológicas de IA, como Demis Hassabis (Google DeepMind), Sam Altman (OpenAI), Daniela y Dario Amodei (Anthropic) o Bill Gates; también hay ingenieros e investigadores, como Ray Kurzweil y Rob Pyke (Google); y profesores de Ciencias de la Computación, como Bart Selman (Universidad de Cornell), Vincent Conitzer y Philip Torr (Universidad de Oxford), James Mickens (Universidad de Harvard), Yoshua Bengio (Universidad de Montreal-Montreal Institute for Learning Algorithms), o Geoffrey Hinton (Universidad de Toronto), y otros muchos, entre los 350 firmantes iniciales de esta declaración, que ha seguido sumando apoyos hasta el presente.

(7) M. SIGMAN y S. BILINKIS, *Artificial. La nueva inteligencia y el contorno de lo humano*, cit., p. 169.

Tanto la carta de los tecnólogos y expertos en IA de marzo, como la declaración de los líderes de la IA de mayo, ponen de manifiesto una inquietud compartida ante las millones de personas que, cada vez con mayor frecuencia, recurren a los chatbots como modo de entretenimiento, acompañamiento y/o aumento de la productividad, a medida que la tecnología subyacente mejora a un ritmo vertiginoso.

Para garantizar que la IA sea segura para la humanidad, tres altos ejecutivos de OpenAI: Sam Altman, Greg Brockman e Ylia Sutskever, han revelado que es *concebible* que la IA obtenga habilidades extraordinarias que superen a los humanos en la próxima década, lo cual nos ofrecerá la posibilidad de un futuro más próspero, pero para llegar a ese horizonte tan prometedor será preciso que sepamos gestionar el riesgo que comporta la IA. Para hacer viable este propósito, estos altos directivos han propuesto la creación de un organismo de control o agencia regulatoria internacional encargada de supervisar la IA general (en adelante, IAG). Esta agencia internacional de la IA sería comparable a la que existe para inspeccionar la producción de la energía atómica (el Organismo Internacional de la Energía Atómica); ahora bien, mientras que para producir la tecnología nuclear es necesario hacer grandes inversiones y acceder a materiales tan difíciles de obtener, como el iranio enriquecido, gran parte de la tecnología de la IA es de código abierto, con lo que no hay restricciones significativas para que cualquier persona en el mundo pueda llegar a elaborar una bomba atómica informática en su propia casa [8].

Al margen del debate suscitado por los tecnólogos a raíz de su pronunciamiento en aras de la pausa y la cautela en el desarrollo de la investigación de la IA, con la IAG como objetivo en el horizonte, una de las principales conclusiones que se desprenden de esta controversia científico-tecnológica es la coincidencia, por parte de quienes están a favor del desarrollo sin límites de la IA, como de aquellos que mantienen una actitud más recelosa o prudente ante el avance de la tecnología de la IA, es que nos encontramos ante un proceso de transformación capaz de comprometer el futuro de la humanidad debido a su potencial desestabilizador [9].

El *homo tecnologicus* se encuentra ante una nueva ola tecnológica, impulsada por una tecnología de uso general con profundas implicaciones sociales, que le obligará a enfrentarse, en las décadas venideras, a las cuestiones más fundamentales que nunca se haya planteado nuestra especie. A propósito de los efectos transformadores de la nueva ola tecnológica, Mustafa Suleyman y Michael Bhaskar, sostienen en un libro reciente que, en la actualidad, nos encontramos ante un problema de *contención de la tecno-*

(8) Ibid., p. 194.
(9) B. CHRISTIAN, *The Alignment Problem. How Can Artificial Intelligence Learn Human Values?*, Atlantic Books, London, 2021, pp. 328-329.

logía de la IA, en cualquier fase de su desarrollo y distribución, y que, si en última instancia no se puede impedir que la tecnología prolifere, ni controlar su onda expansiva de consecuencias no deseadas (tanto si son positivas como negativas), al menos debería limitarse. Se trata, en definitiva, de un enorme reto ético-jurídico al que nuestra especie deberá enfrentarse más pronto que tarde.

Durante la mayor parte de la historia, el reto de la tecnología consistía en la creación y la liberación de su poder. Ahora es al revés: el reto radica en la contención del poder desatado y en garantizar que siga prestándonos servicio tanto a nosotros como al planeta [(10)].

En suma, la necesidad desarrollar una IA al servicio del bienestar de la humanidad exige, de un lado, el establecimiento de un marco normativo adecuado que regule la economía de los datos y se adapte a una IA en permanente evolución; en este sentido, en el seno de la Unión Europea se ha ido conformando un auténtico *corpus iuris digitalis* que tiene ante sí el reto de aprovechar las ventajas de las tecnologías de IA para ganar en competitividad económica, e impulsar la prosperidad y el bienestar de sus ciudadanos, en un entorno neotecnológico seguro, fiable y compatible con los valores y principios que inspiran la Carta de los Derechos Fundamentales de la Unión Europea. Precisamente estos fueron los fundamentos que inspiraron el acuerdo de la Unión Europea, alcanzado el 8 de diciembre de 2023, tras aprobar el parlamento, la Comisión y el Consejo la primera Ley de IA de la historia. Este importante logro, ha supuesto un verdadero hito en la historia del derecho y de la ciencia, además ha servido para colocar a Europa en un lugar destacado para desempeñar un papel de liderazgo a nivel mundial en este ámbito (me refiero al llamado "efecto Bruselas" [(11)]), que servirá también para generar confianza en los usuarios de las aplicaciones de IA, para establecer mecanismos de control a las aplicaciones de IA de alto riesgo, para reforzar la seguridad, y garantizar los derechos de las personas y las empresas [(12)].

Desde un punto de vista jurídico, se han producido avances significativos en la regulación de una IA al servicio de la humanidad, sobre todo en el ámbito de la Unión Europea, que cuenta con la primera ley integral del mundo sobre IA y que, como parte de su estrategia digital, vela por la seguridad, la transparencia, la eticidad y el control humano de los sistemas de IA. Sin embargo, no puede soslayarse el hecho de que, aunque el bienestar de la humanidad sea uno de los principios que guían la normativa europea

---

(10) M. SULEYMAN y M. BHASKAR, *La ola que viene*, trad. esp. C. Fernández, Debate, Barcelona, 2023, p. 58.

(11) A. BRADFORD, "The Brussels Effect", *Northwestern University Law Review*, 107, 2012, pp. 1-67.

(12) F. H. LLANO ALONSO, *Homo ex machina. Ética de la inteligencia artificial y Derecho digital ante el horizonte de la singularidad tecnológica*, Tirant lo Blanch, Valencia, 2024, p. 181.

sobre la IA, es urgente e inaplazable el planteamiento de un debate ético a propósito del impacto y los efectos negativos producidos por los sistemas artificiales, así como la falta de asunción de responsabilidades por parte de quienes comercian y se enriquecen con dichos sistemas sin respetar códigos deontológicos ni normas de conducta que permitan diferenciar lo bueno de lo malo, y que favorezcan lo primero y eviten lo segundo [13].

2. La IA fuerte y el alineamiento de sus objetivos con los valores humanos

Las leyes de la robótica de Isaac Asimov parten de una presunción: si el código ético de las máquinas y los valores humanos no estuvieran correctamente alineados, una IAG acabaría destruyendo, racional y consecuentemente, nuestro ecosistema y, con ello, la humanidad entera [14]. Por eso mismo, como advirtiera Norbert Wiener —considerado el fundador de la cibernética y un influyente defensor de la automatización— si para lograr nuestros objetivos utilizamos un agente mecánico (como un robot u otra máquina inteligente) deberíamos asegurarnos antes de hacerlo de que el objetivo que perseguimos con dicho agente es el que de verdad deseamos [15].

A esta hipótesis distópica, en la que una IAG no se encuentra alineada con la humanidad, apuntó Isaac Asimov cuando enunció, en 1941, las tres leyes de la robótica, publicadas un año más tarde en un relato corto titulado: *Runaround*, que ha inspirado no solo a otras generaciones ulteriores de escritores de ciencia ficción, sino a científicos cognitivos: Minsky (2006), ingenieros informáticos: Barfield (2015), y juristas: Sartor (1993) y Pagallo (2013), entre otros.

Las tres leyes clásicas de la robótica prohíben a los robots causar daño a los humanos, les obligan a obedecer las órdenes dadas por los seres humanos, y les ordenan preservar su propia existencia. Estas tres leyes establecen los siguientes principios:

1. Un robot no hará daño a un ser humano ni, por inacción, permitirá que un ser humano sufra daño.

---

(13) A. BIRHANE-J. VAN DIJK, "Robot Rights? Let´s Talk about Human Welfare Instead", *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES´20)*, Association for Computing Machinery, New York, 2020, pp. 207-213; S. DEGLI-ESPOSTI, *La ética de la inteligencia artificial*, CSIC-Catarata, Madrid, 2023, p. 100.

(14) S. DEGLI-ESPOSTI, *La ética de la inteligencia artificial*, cit., pp. 74-75.

(15) N. WIENER, *Cybernetics or Control and Communication in the Animal and the Machine*, The Technology Press-John Wiley and Sons-Inc. Hermann et Cie Editeurs, Cambridge (Massachusetts)-New York, Paris, 1948; *The Human Use of Human Beings: Cybernetics and Society*, Doubleday & Co., Garden City (New York), 1950.

2. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley.

3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley.

Con posterioridad estas leyes originales fueron modificadas y desarrolladas por otros novelistas de ciencia ficción, e incluso el propio Asimov añadiría una cuarta ley (ley cero), en *Robots and Empire* (1985), para un imaginario universo futurista y distópico en el que las máquinas han asumido el gobierno del planeta tras desplazar a los hombres. Esta cuarta ley responde al siguiente enunciado:

4. Un robot no puede dañar a la humanidad o, por inacción, permitir que la humanidad sufra daños.

En cuanto a esta cuarta ley, contempla la hipótesis de que un robot pueda llegar a incluso a actuar contra un ser humano para evitar que haga daño a sus congéneres. Como señala Jacques Pitrat, esta ley cero es una ley pensada para proteger a la humanidad de su peor enemigo: el hombre [(16)]. A este respecto, en el relato *The Evitable Conflict* (1950) Asimov imagina un mundo controlado por las máquinas, pues estas han llegado a la conclusión de que la única manera de satisfacer el mandato de la primera ley es conspirar contra los hombres para salvarles de sí mismos. El coordinador de ese gobierno planetario en el que las máquinas regulan la economía mundial se llama Stephen Byerley, un robot humanoide dotado de conciencia artificial que vela por el bien de la humanidad.

La hipótesis del conflicto entre máquinas inteligentes y personas, tantas veces llevada a la novela distópica y el cine de ciencia ficción, en la que la IA podría tomar decisiones contrarias al interés de los seres humanos, se plantea el llamado "problema del alienamiento": ¿cómo conseguir que las máquinas inteligentes no se revuelvan contra sus creadores? A este respecto, como han advertido Sigman y Bilinkis, ese potencial enfrentamiento no requeriría necesariamente una conciencia o maldad intrínseca por parte de la IA.

Es suficiente con que perdamos el control sobre algo sumamente poderoso, y capaz de operar a gran escala sobre el mundo, por un error en la función valor, por alguna omisión en las limitaciones que se le impongan, por alguna propiedad emergente inesperada o por l posibilidad que tienen de modificar los objetivos que les damos [(17)].

---

(16) J. PITRAT, *Artificial Beings: The Concience of a Conscious Machine*, John Wiley & Sons Inc., Hoboken (New Jersey)-London, 2009, p. 184.

(17) M. SIGMAN y S. BILINKIS, *Artificial. La nueva inteligencia y el contorno de lo humano*, cit., p. 177.

Como puede comprobarse tras la lectura de los dos relatos de Asimov que hemos tomado como ejemplo, *Runaround* y *The Evitable Conflict*, la literatura de ciencia ficción ha anticipado situaciones e hipótesis que han pasado a formar parte de la experiencia jurídica de la era cibernética y precisan una específica regulación: la ética de la IA, la autonomía de las máquinas y su impacto en el ámbito de los derechos humanos, la conciencia artificial, la personalidad electrónica o la responsabilidad civil y penal por los daños causados por la IA y la robótica [(18)].

A la vista de la complejidad de las leyes de la robótica, de los dilemas éticos de difícil solución que estas presentan, de las consecuencias de los sesgos implícitos en los datos de entrenamiento de los algoritmos y de los desafíos que entraña el problema del alineamiento, sería oportuno abordar estas cuestiones dentro de un marco ético que limite y oriente las acciones y decisiones automatizadas de las máquinas inteligentes al interactuar con las personas. Pero ¿cómo consensuar un código ético aplicable a las máquinas, cuando ni siquiera hemos logrado ponernos de acuerdo entre las personas a la hora de compartir un marco ético universal, ni unos mismos valores morales que orienten la conducta de los individuos? Y por otra parte, ¿con qué conjunto de valores debería alinearse la IA y qué estatus legal y ético debería tener?

La cuestión sobre el alineamiento de los valores y comportamientos de los sistemas de IA con los valores humanos se trató, precisamente, en la Conferencia de Asilomar sobre IA beneficiosa, celebrada en enero de 2017 en California y organizada por el *Future of Life Institute*. Este evento ético-tecnológico contó con la presencia de más de un centenar de expertos e investigadores que se reunieron para debatir y consensuar principios de IA. Esta conferencia concluyó con la formulación de veintitrés principios que suponen, para algunos, un valioso punto de referencia para los debates sobre ética de la IA, mientras que, para otros, carecen de una orientación específica sobre cómo aplicarlos a la práctica. En todo caso, aunque no sean suficientes, se pueden considerar como principios necesarios para iniciar abordar los complejos retos éticos que surgen en torno a la IA.

El décimo principio de la Conferencia de Asilomar versa sobre la alineación de valores y reza así:

> "Los sistemas de IA altamente autónomos deben ser diseñados de manera que sus objetivos y comportamientos puedan ser asegurados para alinearse con los valores humanos a lo largo de su operación".

---

(18) U. PAGALLO, *The Laws of Robots. Crimes, Contracts, and Torts*, Springer, Dordrecht-Heidelberg-New York-London, 2013, p. 23.

Como vemos, este principio se refiere fundamentalmente a los sistemas de IA autónomos, capaces de tomar decisiones automatizadas y llevar a cabo acciones sin intervención humana. En otras palabras, con este principio lo que se pretende es garantizar que los objetivos y comportamientos de los sistemas de IA autónomos estén diseñados de forma que se alineen con valores humanos, para que no actúen al margen de las normas establecidas, ni produzcan daños o lesiones a los seres humanos, tal y como prescribe la primera regla de Asimov.

Pero ¿cómo conseguimos que los sistemas de IAG superinteligentes estén alineados con los valores humanos y sigan la intención humana sin rebelarse? En julio de 2023 OpenAI publicó un documento de investigación, firmado por Ilya Sutskever y Jan Leyke, codirectores del equipo *Superalignment*, en el que explicaban los resultados de su investigación sobre un modelo de supervisión automatizado que sea capaz de guiar el comportamiento de otro mucho más inteligente (este proceso es el que se utiliza, por ejemplo, por OpenAI para ajustar su modelo del lenguaje estrella en 2023: Chat GPT-4) [19].

El ajuste algorítmico a GPT-4 que está probando el equipo *Superalignment* permitiría, a la postre, que el modelo más potente e inteligente siga la guía del modelo más débil y menos inteligente, pero sin reducir por ello su rendimiento. En todo caso, los investigadores de OpenAI admiten que este novedoso método de superalineamiento no garantiza aún que el modelo más fuerte se comporte a la perfección, aunque puede servir como un punto de referencia del que partirán futuras investigaciones sobre el control de la IAG [20].

3. La ética de la IA y la ampliación de los principios de la bioética

En el anterior epígrafe nos ocupamos del alineamiento de los valores y comportamientos de los sistemas de IA con los valores humanos. Esta operación de alineación es conveniente realizarla desde el momento en que, a través de la IA, ha sido posible automatizar tareas y delegar al *software* o a máquinas inteligentes el desarrollo de determinadas funciones.

La IA está tan inserta en nuestra vida cotidiana y en el contexto social que resulta necesario tomar en consideración desde un punto de vista ético los efectos y las consecuencias que esta tecnología produce sobre el género humano en plena era de la automatización. En definitiva, esta es la razón por la que resulta oportuna una investigación crítica en clave ética que no

(19) www.wired.com/story/openai-ilya-sutskever-ai-safety/
(20) R. GÓMEZ PÉREZ, *El dilema de la IA*, Rialp, Madrid, 2023, p. 92.

solo se ocupe de la eficacia y el rendimiento de las nuevas tecnologías, sino también del bienestar de la humanidad ante el horizonte de la singularidad tecnológica [21].

El objetivo de este enfoque ético es doble: por un lado, promover la concordancia o alineamiento entre las intenciones de las distintas partes y los valores éticos pertinentes para el uso previsto; por otro lado, Identificar, corregir o denunciar las aplicaciones que sirven a fines éticamente inaceptables que ignoran, o violan, valores decisivos en relación con su ámbito de actuación. Este segundo objetivo (el de la vigilancia contra el uso indebido de los sistemas de IA) tiene una gran relevancia, porque si se verifica un uso desalineado de los sistemas, entonces la denuncia del problema y la puesta en marcha de acciones dirigidas a su eliminación constituye un acto obligatorio y absolutamente necesario.

En realidad, la ética es un ingrediente básico para que la investigación tecnológica pueda avanzar tranquilamente, explorar sus propias posibilidades y ser beneficiosa para la humanidad, en la medida que propicia el mantenimiento de un alto nivel de confianza entre los actores implicados en el proceso de la IA. En última instancia, como sostiene Stefano Quintarelli, razonar sobre el posible impacto moral de la IA, y asegurarse de que tenga efectos beneficiosos sobre la existencia de la humanidad, no tiene porqué suponer un freno al pleno desarrollo de la tecnología sino, al contrario, es la receta para su para su éxito a largo plazo y para cosechar todos los beneficios que promete [22].

Sin embargo, para que haya una alineación entre las tecnologías y las expectativas éticas, estas deben estar claramente definidas. Establecer un marco de valores lo más coherente y claro posible es, en efecto, crucial para que la investigación y el desarrollo tecnológicos sean éticamente conscientes y se practiquen a la luz de esta conciencia. A propósito de la definición de este marco axiológico relativo a la IA, cabe mencionar la gran labor ético-jurídica realizada a través de las declaraciones internacionales sobre los principios éticos de la IA. En este sentido, Luciano Floridi [23] ha destacado algunas que, a su juicio, serían las más influyentes hasta ahora en la definición del marco ético-jurídico de la IA; estas serían las tres principales:

- Los *Principios de Asilomar*, un conjunto de 23 principios desarrollados bajo los auspicios del Future Life Institute, en colaboración con los participantes en la Conferencia Asilomar que se celebró en Pacific Grove (Califor-

---

(21) J. SCHAICH BORG-W. SINNOTT-ARMSTRONG-V. CONITZER, *Moral AI and How We Get There*, Penguin Random House, Milton Keyes-Dublin, 2024, pp. 190-191.

(22) S. QUINTARELLI, *Intelligenza Artificiale. Cos´è davvero, come funziona, che effetti avrà*, Bollati Boringhieri, Milano, 2020, p. 84.

(23) L. FLORIDI, *Etica dell´intelligenza artificiale. Sviluppi, opportunità, sfide,* Raffaelo Cortina Editore, Milano, 2022, pp. 94-95.

nia) en enero de 2017. Estos principios hacen referencia, entre otras cuestiones, a la seguridad y la protección, la transparencia y la responsabilidad, el desarrollo y uso de las tecnologías de IA de forma justa y equitativa, y por supuesto, también a la garantía de que los sistemas de IA se diseñen y operen sin daño ni discriminación de ningún individuo o grupo (es decir, que todos tengan acceso al uso y el beneficio de la IA).

- La *Declaración de Montreal para un desarrollo responsable de la Inteligencia Artificial* (2018) en la que se enuncian una serie de diez principios que sientan las bases para fomentar la confianza de la sociedad en los sistemas de IA. El sexto principio que se proclama en esta declaración es el de equidad, que precisamente en su primer apartado exhorta al diseño y el entrenamiento de los sistemas de IA de manera tal que "no creen, refuercen ni reproduzcan patrones de discriminación basados en diferencias sexuales, étnicas, culturales o religiosas, entre otras".

- La *Declaración sobre Inteligencia Artificial, robótica y sistemas "autónomos"*, publicada en marzo de 2018 por el Grupo Europeo sobre Ética de la Ciencia y las Nuevas Tecnologías de la Comisión Europea (en adelante, EGE), en el que se proponen un conjunto de principios éticos fundamentales y prerrequisitos democráticos, basados en los valores establecidos en los Tratados y en la Carta de derechos fundamentales UE: a) dignidad humana; b) autonomía; c) responsabilidad; d) justicia, equidad y solidaridad; e) democracia; f) Estado de derecho y rendición de cuentas; g) seguridad, protección, e integridad física y mental; h) protección de datos y privacidad; i) sostenibilidad.

A diferencia de los Principios de Asilomar y de la Declaración de Montreal, en la Declaración de la EGE se hace referencia expresa en el apartado d) a los sesgos discriminatorios y se manifiesta que la IA debería contribuir a la justicia global y facilitar la igualdad de acceso a los beneficios y ventajas de la IA, la robótica y los sistemas "autónomos". De ahí que al final del primer párrafo de este cuarto parágrafo se proclame que:

> "los sesgos discriminatorios en los conjuntos de datos utilizados para entrenar y ejecutar los sistemas de IA, deben evitarse. De no ser posible, estos sesgos deben ser detectados, notificados y neutralizados en la etapa más temprana del proceso".

Consciente de la necesidad de reconocer y regular el impacto de la IA en el sistema de derechos fundamentales, la Unión Europea se ha situado a la vanguardia en la creación de un marco jurídico específico sobre IA. En este sentido, el Parlamento Europeo aprobó una resolución, el 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley. En dicha resolución, que marcó un hito en el comienzo

de la normativa de la Unión Europea sobre IA, se insiste precisamente en el hecho de que:

> "los ciudadanos, los sectores público y privado, el mundo académico y la comunidad científica solo podrán aprovechar plenamente las perspectivas y oportunidades que brindan los macrodatos si la confianza pública en esas tecnologías se garantiza mediante la estricta observancia de los derechos fundamentales y el cumplimiento de la legislación vigente de la Unión en materia de protección de datos, así como la seguridad jurídica en relación con todas las partes interesadas".

A su vez, en el parágrafo 20 de esta misma resolución, el Parlamento insta a la Comisión, a los Estados miembros y a las autoridades encargadas de la protección de datos a que:

> "definan y adopten las medidas que se impongan para minimizar la discriminación y el sesgo algorítmicos y a que desarrollen un marco ético común sólido para el tratamiento transparente de los datos personales y la toma de decisiones automatizada que sirva de guía para la utilización de los datos y la aplicación en curso del Derecho de la Unión [(24)]".

A propósito de las declaraciones de principios de la IA, Floridi ha advertido que existe una suerte de hipertrofia de principios éticos que se suceden cada vez que una organización aprueba una declaración aún a riesgo de incurrir en innecesarias repeticiones o superposiciones de principios proclamados en otras declaraciones precedentes, lo cual también conduce a la confusión y la ambigüedad.

Frente a esta profusión de principios éticos, Floridi nos propone una idea de IA que no debe ser entendida como un nuevo tipo de inteligencia, sino como una forma de actuar sin precedentes. Por eso, de todas las áreas de la ética aplicada, la bioética es la que más se parece a la ética digital en el tratamiento que esta hace de nuevas formas de agentes, pacientes y entornos [(25)]. Sin embargo, aunque los principios bioéticos se adaptan perfectamente a los nuevos retos éticos planteados por la IA, no es fácil explicarlos. A este respecto, Floridi añade a los cuatro principios clásicos de la bioética (*beneficencia, no maleficencia, autonomía y justicia*) un quinto principio: la *explicabilidad*, entendida como un principio que incluye tanto el sentido *epistemológico* de *inteligibilidad* (como respuesta a la pregunta: "¿Cómo funciona?"), como el sentido ético de responsabilidad (*accountability*) (como respuesta a la pregunta: "¿Quién es responsable del modo en que funciona?") [(26)].

---

(24) Resolución del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)), parágrafos 1 y 20.

(25) L. FLORIDI, *The Ethics of Information*, Oxford University Press, Oxford, 2013.

(26) L. FLORIDI, *Etica dell´intelligenza artificiale. Sviluppi, opportunità, sfide,* cit., p. 96.

Estos cinco principios son válidos tanto para expertos, como por ejemplo diseñadores o ingenieros de productos, como para los no expertos, como pacientes o clientes [27]. A continuación comentaremos, sucintamente y por separado, cada uno de estos cinco principios:

1. Beneficencia

De los cuatro principios tradicionales de la bioética es el más fácil de observar por su claridad. La mayoría de las declaraciones de principios de la IA coinciden en interpretar este término como "bienestar" de los seres humanos y de todas las criaturas sintientes (Declaración de Montreal), y la conciben como un principio-guía de la IA hacia el bien común, el beneficio de la humanidad y del planeta (Conferencia de Asilomar).

2. No maleficencia

Este principio complementa a la IA benéfica en la medida que avisa contra las diversas consecuencias negativas que se derivan del uso abusivo e impropio de las tecnologías de IA, como las violaciones contra la intimidad personal. En este sentido, la Declaración de Montreal sostiene que quienes diseñan y desarrollan los sistemas de IA deben asumir la responsabilidad correspondiente actuando contra los riesgos derivados de la innovación tecnológica. Sin embargo, en el noveno principio de esta declaración (el de responsabilidad) se advierte que, cuando un sistema de IA provoca daños o perjuicios y se demuestra que el sistema de IA es responsable a pesar de haber sido utilizado según lo previsto, no es razonable culpar a las personas involucradas en su desarrollo y uso.

3. Autonomía

Este principio se relaciona con el anterior en la medida que la autonomía funcional de la máquina inteligente implica una suspensión del control humano en la realización de una determinada tarea automatizada. No parece lícito defender que el usuario sea considerado responsable di los eventuales efectos negativos causados por el funcionamiento del sistema. En otras palabras: la autonomía funcional puede considerarse, como causa de la erosión del control del funcionamiento del sistema, una buena razón para la limitación de responsabilidad de los individuos implicados en su diseño y desarrollo (programadores, constructores, etc.…) [28].

Ahora bien, si ningún actor humano puede ser considerado plenamente responsable de los efectos causados por el funcionamiento de la IA, y si

(27) D. S. WATSON-L. FLORIDI, "The explanation game: A formal framework for interpretable machine learning", *Synthese*, 2020, pp. 1-32.

(28) T. M. POWERS-J. G. GANASCIA, "Autonomy", *The Oxford Handbook of Ethics of AI* (ed. M. D. Dubber-F. Pasquale-S. Das), Oxford University Press, Oxford, 2021, pp. 31-32.

no basta tampoco con imputarle la culpa al sistema de IA mismo, ¿cómo se puede resolver este dilema? A este respecto, hay dos alternativas posibles:

En primer lugar, recuperar algún tipo de control significativo sobre el funcionamiento autónomo de nuestras herramientas de IA.

En segundo lugar, ampliar el concepto de responsabilidad más allá del paradigma del control [(29)].

4. Justicia

El principio de justicia, pese a su importancia, es el que más acepciones presenta en las diferentes declaraciones de principios de la IA. Mientras que en la Declaración de Montreal se apela al principio de justicia para que "no se creen, refuercen ni reproduzcan patrones de discriminación basados en diferencias sociales, sexuales, étnicas, culturales o religiosas, entre otras" (sexto principio); en los Principios de Asilomar (concretamente en el decimoquinto) se alude a la justicia como prosperidad económica compartida; por lo demás, en otras declaraciones el término justicia tiene otros significados (por ejemplo en el sentido de equidad). En suma, como afirma Floridi, las distintas formas de caracterizar la justicia remiten, en última instancia, a una falta de claridad más amplia sobre la IA como reserva de acción inteligente creada por los seres humanos [(30)].

5. Explicabilidad

Para salir de esta confusión debida a la diversidad de significados divergentes del principio de justicia, Floridi propone un quinto principio específicamente referido a la IA que complementa a los principios clásicos de la bioética: el principio de explicabilidad. La incorporación del principio de explicabilidad, que incluye tanto el sentido epistemológico de "inteligibilidad" como el sentido ético de "responsabilidad", es "la pieza crucial que falta para completar el *puzzle* ético de la IA".

En efecto, este quinto principio complementa los otros cuatro porque, para que la IA sea beneficiosa y no dañina, es necesario que estemos en disposición de comprender el bien o el daño que se está haciendo a la sociedad y de qué manera; por otra parte, para que la IA promueva y no limite la autonomía humana, nuestra decisión sobre quién debe decidir tiene que ser informada por un conocimiento previo: cómo actuaría la IA en nuestro lugar y, en tal caso, cómo podrían mejorarse sus prestaciones; y, por último, para que la IA sea justa, necesitamos saber a quién responsabilizar ética o

(29) S. QUINTARELLI, Intelligenza Artificiale. Cos´è davvero, come funciona, che effetti avrà, cit., p. 90.

(30) L. FLORIDI, Etica dell´intelligenza artificiale. Sviluppi, opportunità, sfide, cit., p. 100.

jurídicamente en caso de un resultado grave o negativo, lo que a su vez requeriría una comprensión adecuada de por qué se produjo tal resultado [31].

4. DIVERSOS ENFOQUES Y MODELOS ÉTICOS DE LA IA

La ética de la IA puede abordarse desde diversos enfoques, cada uno de ellos centrado en diferentes aspectos éticos relacionados con el desarrollo, la implementación y el uso de la IA. Entre los distintos enfoques éticos de la IA cabe diferenciar al menos cuatro perspectivas, dependiendo de cuál sea su interés o particular objeto de análisis, ya sea el usuario, el diseñador, la sociedad, el dato o el algoritmo.

Tanto la ética de la IA del dato, que se centra en cómo se recopilan, utilizan y gestionan los datos en los sistemas de IA, como la ética que se ocupa del diseño ético de los algoritmos, comparten principios éticos fundamentales para guiar el desarrollo, la aplicación y el uso de la IA de manera ética y responsable. Dicho esto, conviene advertir que, pese a que no existe todavía un consenso universal sobre los principios éticos de la IA, el proceso de confluencia en torno a dichos principios irá avanzando a medida que las organizaciones de expertos en ética, científicos e investigadores, tecnólogos, empresas de IA, comités y grupos de trabajo colaboren para conseguir la formulación de unos mínimos principios y estándares éticos universales a propósito de las tecnologías convergentes (tecnologías como las que han permitido, por ejemplo, la convergencia entre la IA, la robótica o el análisis de datos para el desarrollo de sistemas y vehículos autónomos).

Respecto a los principios-guía comunes a algunos de los códigos éticos para el desarrollo, despliegue y buen uso de la IA cabe mencionar los siguientes:

1. Transparencia algorítmica

Este principio exige la visibilidad, la cognoscibilidad, la auditabilidad y la explicabilidad de los factores intervinientes en las decisiones tomadas con algoritmos a las personas que utilizan, regulan, y son afectadas por los sistemas de IA que emplean dichos algoritmos [32].

La transparencia significa además permitir que las personas comprendan cómo se desarrolla, entrena, opera y despliega un sistema de IA en el dominio de aplicación relevante, de modo que los consumidores, por ejem-

---

(31) Ibid.

(32) A. SANTANGELO, "Equità degli algoritmi e democracia", *DigitCult*, 2020, vol. 5, pp. 21-30 (http://dx.doi.org/10.53136/979125994120634); G. VESTRI, "Transparencia algorítmica", en *Diccionario de términos para comprender la transformación digital*, G. VESTRI (Director)-M. CASTILLA BAREA (Coordinadora), Aranzadi, Cizur Menor (Navarra), 2023, pp. 346-348.

plo, puedan tomar decisiones más informadas. El principio de transparencia también se refiere a la capacidad de proporcionar información significativa y claridad sobre qué información se proporciona y por qué [33].

2. Equidad

Entendiendo la justicia como equidad, lo que se pretende con este principio es que los algoritmos predictivos que resultan decisivos en la toma de decisiones y que tanto condicionan la vida de las personas (por ejemplo, en un proceso de selección de candidatos en una oferta de trabajo, en la búsqueda de receptores de un trasplante, en la aprobación de préstamos, en la evaluación de riesgos, etc.…) sean diseñados evitando la discriminación y los sesgos en la toma de decisiones basadas en datos. En definitiva, el cumplimiento de este principio garantiza un trato justo y equitativo para todos los usuarios [34].

3. Responsabilidad

La generalización de los sistemas de aprendizaje automático basados en algoritmos en muchos sectores ha sacado a la luz el problema de la responsabilidad por "fallo de algoritmo". Los sistemas de aprendizaje automático basados en algoritmos permiten a las máquinas mejorar su rendimiento con el tiempo; sin embargo, en caso de error, las consecuencias para las personas y los bienes pueden ser especialmente graves. La complejidad de estos sistemas, sin embargo, hace difícil comprender a quién se puede responsabilizar de estos daños. A propósito de la responsabilidad de los sistemas expertos basados en IA, algunos expertos en la materia han señalado que

> "la atribución de 'sujeto de derecho' a los efectos, en su caso, de imputar el daño ocasionado a un comportamiento o actividad debería referirse a sistemas expertos con capacidad de aprendizaje que actúan de forma completamente autónoma respecto de la persona que responda por ellos. De otra parte, la regla de la responsabilidad proporcional en materia de causalidad en relación con los problemas de incertidumbre que se generan en un entorno operado por sistemas expertos con o sin interacción con humanos debe tenerse en cuenta a la hora de plantearse una posible regulación de la responsabilidad civil por los daños ocasionados [35]".

---

(33) Principio 1.3 de la lista de principios de IA de la OCDE, adoptados en mayo de 2019.

(34) M. MORENO REBATO, "Discriminación algorítmica", en *Diccionario de términos para comprender la transformación digital*, cit., pp. 141-143. S. VANTIN, "Inteligencia Artificial y derecho antidiscriminatorio", en *Inteligencia Artificial y derecho. El jurista ante los retos de la era digital,* F. H. LLANO ALONSO (Editor), Thomson Reuters Aranzadi, Cizur Menor (Navarra), 2021, pp. 367-384.

(35) S. NAVAS NAVARRO, "Sistemas expertos basados en inteligencia artificial y responsabilidad civil", Diario La Ley, 13 de diciembre de 2021; en la misma línea, vid., M. MARTÍN CASALS, «Proportional Liability in Spain», en M. MARTÍN CASALS-D. M. PAPAYANNIS, *Uncertain Causation in Tort Law*, Cambridge University Press, Cambridge, 2016, pp. 43-66.

4. Privacidad

Existe una compleja intersección entre la privacidad, la protección de datos y la ética de la IA. En su reciente libro *The Ethics of Privacy and Surveillance*, Carissa Véliz propone una definición mixta de privacidad en la que se hibridan las dos teorías principales entorno a este concepto: de un lado, la teoría del *acceso*, según la cual, la cuestión de la privacidad consiste fundamentalmente, bien en la accesibilidad limitada, o bien en la inaccesibilidad a la información sensible (en este sentido, se entiende que perdemos privacidad cuando la información que afecta a nuestros datos personales es accesible para los demás; por otro lado, la teoría del *control* de privacidad centra su atención en la protección de nuestros datos personales ante la intromisión o injerencia de actores externos [36].

A propósito de la protección de datos, dentro del marco regulatorio europeo, en el artículo 5 del Reglamento General de Protección de Datos (RGPD) [37] se establecen los principios relativos al tratamiento de los datos personales, entre los que cabe resaltar los requisitos de *licitud*, *lealtad* y *transparencia* en el tratamiento de datos (art. 5.1 a), además de los principios de *finalidad* (art. 5.1 b) y *minimización de datos* (art. 5.1 c) [38].

5. Seguridad

Al igual que sucede con la interrelación entre el principio de responsabilidad y la protección de datos, también existe una interacción entre el principio de seguridad y la solidez de los sistemas de IA.

---

(36) C. VÉLIZ, *The Ethics of Privacy and Surveillance*, Oxford University Press, Oxford, 2024, pp. 74-75.

(37) Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016.

(38) El tratamiento de datos se entenderá lícito si se cumple al menos alguna de las siguientes condiciones previstas en el art. 6° RGPD: a) Cuando el interesado ha dado su consentimiento para el tratamiento de sus datos personales para uno o varios fines específicos; b) cuando el tratamiento es necesario para la ejecución de un contrato en el que el interesado es parte o para la aplicación a petición de este de medidas precontractuales; c) cuando el tratamiento es necesario para el cumplimiento de una obligación legal aplicable al responsable del tratamiento; d) cuando el tratamiento es necesario para proteger intereses vitales del interesado o de otra persona física; e) cuando el tratamiento es necesario para el cumplimiento de una misión realizada en interés público o en el ejercicio de poderes públicos conferidos al responsable del tratamiento; f) cuando el tratamiento es necesario para la satisfacción de intereses legítimos perseguidos por el responsable del tratamiento o por un tercero, siempre que sobre dichos intereses no prevalezcan los intereses o los derechos y libertades fundamentales del interesado que requieran la protección de datos personales, en particular cuando el interesado sea un niño.

Con respecto a los principios de *limitación de finalidad* y *minimización de datos*: el primero implica que los datos personales han de ser recogidos con fines determinados, explícitos y legítimos, y que no serán tratados ulteriormente de manera incompatible con dichos fines (a excepción de los fines de archivo en interés público, de investigación científica e histórica o estadísticos que, de conformidad con lo establecido en el art. 89. 1 RGPD, no se considerarán incompatibles con los fines iniciales); el segundo principio exige que los datos personales sean adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados.

En efecto, tanto en el apartado 1.4 de la lista de principios de la IA de la OCDE, como en el artículo 15 del Reglamento de Inteligencia Artificial, "solidez" es sinónimo de capacidad de resistir o superar condiciones adversas, incluidos los riesgos de la seguridad digital. Este principio establece además que los sistemas de IA no deben plantear riesgos que no sean razonables para la seguridad, incluida la seguridad física, en condiciones de uso normal o previsible o de uso indebido a lo largo de su ciclo de vida [(39)].

Hay dos maneras de mantener sistemas de IA sólidos, seguros y protegidos: en primer lugar, mediante su *trazabilidad* y posterior análisis e indagación; y, en segundo lugar, aplicando un *enfoque de gestión de riesgos*.

La trazabilidad puede ser útil para realizar el análisis y la investigación de los resultados de un sistema de IA y es una forma de promover la rendición de cuentas; además, puede ayudar también a prevenir futuros errores y a mejorar la fiabilidad del sistema de IA.

En relación con el enfoque de gestión de riesgos, aplicado a lo largo de todo el ciclo de vida del sistema de IA, este puede ser de gran utilidad para identificar, evaluar, priorizar y mitigar los riesgos potenciales que pueden afectar negativamente al comportamiento y los resultados de un sistema.

5. A modo de conclusión: hacia una metaética de la IA

Como se ha podido comprobar, la ética de la IA es un tipo de ética aplicada a los negocios y la tecnología, pero no deja de ser un subconjunto de la Ética (con mayúsculas). La ética de la IA se ocupa de identificar y dar respuesta a las problemáticas morales que se derivan de la implantación y el uso de soluciones de IA. La autonomía, la justicia, la explicabilidad, la transparencia de los algoritmos, la equidad, la responsabilidad, la privacidad, la seguridad..., son algunas de las principales áreas de interés y preocupación por parte de expertos tecnólogos, filósofos y juristas, entre otros especialistas en la materia.

Paradójicamente, a pesar de la importancia y la necesidad de la ética de la IA, esta puede ser utilizada a veces de forma irresponsable o poco ética por algunas grandes empresas o compañías tecnológicas (un caso paradigmático es el de Timnit Gebru, a quien Google contrató en 2020 como codirectora de un grupo de investigación centrado en la IA ética, pero a la que despidió meses después por haber criticado los riesgos éticos de los modelos del lenguaje de la compañía para la que trabajaba).

En general, la tendencia actual de los gigantes tecnológicos (que responden al acrónimo GAFAM: Google, Amazon, Facebook, Apple y Microsoft)

(39) https://oecd.ai/en/dashboards/ai-principles/P8

es la de ir reduciendo plantilla de sus equipos de ética de la IA (Elon Musk, por ejemplo, despidió a la mitad de los trabajadores de Twitter cuando, en octubre de 2022, adquirió esta red social, incluido su pequeño equipo de IA ética; otro caso revelador fue el de Microsoft, compañía tecnológica que, en marzo de 2023, cerró el equipo de "ética y sociedad" de su división de IA, entre otros motivos porque, según John Montgomery, uno de los vicepresidentes de la división de IA, el objetivo de la compañía era que los modelos conversacionales de OpenAI llegasen a los clientes a la mayor velocidad posible, sin reglas ni restricciones éticas que retrasasen su entrega).

Frente a estos intentos de rentabilizar la ética para blanquear estrategias de mercado (*ethics washing*) con un aparente compromiso de buena praxis y respeto a los principios y las reglas éticas de la IA, pero que en realidad solo sirven para ocultar la intención de dirigir dichas estrategias al aumento de la competitividad de la empresa y a una mejora en su cuenta de resultados, sería oportuno recurrir a una ética de la ética de la IA (es decir, a una *metaética de la IA* que se centre en la fundamentación de los juicios normativos o de valor) [(40)]. Por eso, más que de una sola ética de la IA, tal vez lo apropiado sea hablar en plural, esto es, de éticas de la IA (la metaética de la IA y la ética aplicada al mundo de las empresas tecnológicas).

Entre las principales teorías éticas en las que se basan los códigos morales que sirven para el diseño y el desarrollo de la IA y la robótica podrían destacarse, entre otras, el utilitarismo (Bentham, Mill), el deontologismo (Kant), la ética de la virtud (Aristóteles, Tomás de Aquino), la ética del contrato social (Locke, Rousseau, Rawls). Esta metafísica de la IA será objeto de estudio especializado en un trabajo posterior.

---

(40) I. G. R. GALVILÁN, Robots en la sombra: RPA, robots conversacionales y otras formas de automatización cognitiva, Anaya, Madrid, 2021.

Existe um direito fundamental à revisão humana?

Por André Karam Trindade[(*)] *y Amanda Antonelo* [(**)]

Resumo: Existe um direito fundamental à revisão humana? Essa é a questão central que orienta o presente ensaio. O propósito do texto é discutir, criticamente, os riscos da delegação de decisões jurídicas a sistemas algorítmicos e a necessidade de afirmar um direito fundamental à decisão ou revisão humana. Para enfrentar esse tema, após sua problematização inicial (Seção I), apresenta-se o percurso da decisão humana à decisão artificial, analisando os limites da algoritmização do direito (Seção II). Em seguida, discutem-se os modelos de juiz e os elementos subjetivos, criativos e racionais que tornam o julgamento uma atividade irredutivelmente humana (Seção III). Na sequência, examina-se a Teoria da Decisão e a noção de responsabilidade política como fundamentos de uma jurisdição constitucionalmente comprometida (Seção IV). Por fim, aborda-se o reconhecimento da revisão humana como direito fundamental, à luz do *General Data Protection Regulation,* da Lei Geral de Proteção de Dados do Brasil e da evolução histórica e material dos direitos fundamentais, demonstrando sua condição de garantia primária da democracia constitucional (Seção V).

Palavras-chave: Direitos fundamentais; Inteligência Artificial; Poder Judiciário; revisão humana; teoria da decisão.

¿Existe un derecho fundamental a la revisión humana?

Resumen: ¿Existe un derecho fundamental a la revisión humana? Esta es la cuestión central que orienta el presente ensayo. El propósito del texto es discutir, críticamente, los riesgos de la delegación de decisiones jurídicas a sistemas algorítmicos y la necesidad de afirmar un derecho fundamental a la decisión o supervisión humana. Para enfrentar este tema, tras su problematización inicial (Sección I), se presenta el recorrido de la decisión

---

(*) Doutor em Teoria e Filosofia do Direito (UNIROMA3). Mestre em Direito Público (UNISINOS). Professor do Programa de Pós-Graduação em Direito da UNIVEL/PR. Professor Visitante da Università Degli Studi Roma Tre (Itália). E-mail: andrekaramtrindade@gmail.com

(**) Doutoranda em Direito (UFPR). Mestre em Direito, Inovação e Regulações (UNIVEL). Professora da Graduação em Direito da PUCPR e UNIVEL. E-mail: amandaantonelo@hotmail.com

humana a la decisión artificial, analizando los límites de la algoritmización del derecho (Sección II). A continuación, se discuten los modelos de juez y los elementos subjetivos, creativos y racionales que hacen del juicio una actividad irreductiblemente humana (Sección III). Seguidamente, se examina la Teoría de la Decisión y la noción de responsabilidad política como fundamentos de una jurisdicción constitucionalmente comprometida (Sección IV). Finalmente, se aborda el reconocimiento de la revisión humana como derecho fundamental, a la luz del Reglamento General de Protección de Datos, de la Ley General de Protección de Datos de Brasil y de la evolución histórica y material de los derechos fundamentales, demostrando su condición de garantía primaria de la democracia constitucional (Sección V).

Palabras claves: derechos fundamentales; inteligencia artificial; Poder Judicial; revisión humana; teoría de la decisión.

Is there a fundamental right to human review?

Abstract: Is there a fundamental right to human review? This is the central question that guides the present essay. The purpose of the text is to critically discuss the risks of delegating legal decisions to algorithmic systems and the need to affirm a fundamental right to human decision or supervision. To address this issue, after its initial problematization (Section I), the text presents the path from human decision to artificial decision, analyzing the limits of the algorithmization of law (Section II). Next, it discusses judicial models and the subjective, creative, and rational elements that make adjudication an irreducibly human activity (Section III). Subsequently, it examines Decision Theory and the notion of political responsibility as the foundation of a constitutionally committed jurisdiction (Section IV). Finally, it addresses the recognition of human review as a fundamental right, in light of the General Data Protection Regulation, the Brazilian General Data Protection Law, and the historical and material evolution of fundamental rights, demonstrating its condition as a primary guarantee of constitutional democracy (Section V).

Keywords: Fundamental Rights; Artificial Intelligence; Judiciary; human review; decision theory.

1. Introdução

Bertolt Brecht, ao analisar os avanços tecnológicos de 1938, tratou de uma maneira que só os poetas poderiam tratar a respeito do indispensável defeito de pensar dos seres humanos, poema que pode ser transportado para a realidade atual, em que juristas cogitam confiar aos sistemas tecnológicos as decisões: "O vosso tanque, General, é um carro forte. Derruba uma floresta, esmaga cem homens. Mas tem um defeito: Precisa de um motorista. O vosso bombardeiro, General, é poderoso: voa mais depressa que a

tempestade e transporta mais carga que um Elefante. Mas tem um defeito: precisa de um piloto. O homem, meu General, é muito útil: Sabe voar e sabe matar, mas tem um defeito: sabe pensar".

Se, para Brecht, tal defeito revelava a limitação da máquina, incapaz de prescindir da condução humana, hoje denuncia um risco ainda maior, a tentação de delegar ao sistema de inteligência artificial o ato de pensar. Se a máquina precisa de um condutor, também o Judiciário não pode abrir mão do humano como presença crítica e consciente no processo decisório.

Nesse mesmo horizonte de reflexão, o Presidente do Supremo Tribunal Federal do Brasil, Luís Roberto Barroso, em recente palestra na Ordem dos Advogados do Brasil do Ceará (2025), levantou a seguinte indagação: afinal, o que desejamos, uma decisão propriamente humana ou apenas uma decisão de qualidade, ainda que produzida por sistemas de inteligência artificial? Essa pergunta, que tensiona a legitimidade do ato de julgar, evidencia o dilema: a tecnologia pode oferecer fundamentação formal, mas será que pode substituir a experiência crítica, ética, responsável e política da decisão judicial?

O presente ensaio pretende investigar, sob o viés da Teoria da Decisão e do cenário tecnológico, a necessidade da promoção concreta humana sobre o processo de algoritmização do Poder Judiciário. Mais do que isso: pretende demonstrar que da interação entre Direito e tecnologia surge a exigência de um direito fundamental à decisão ou revisão intrinsecamente humana, um direito que não se reduz, nem pode ser reduzido, à performance algorítmica.

Trata-se, talvez, da questão que mais inquieta juristas e mesmo cidadãos comuns: a possibilidade de que uma máquina possa proferir decisões, de modo que o destino humano se encontre nas mãos não de pessoas, mas de sistemas que operam exclusivamente a partir de variáveis estatísticas. Sistemas que passam a repetir indefinidamente os mesmos padrões, sem se abrir às mudanças ou às singularidades, reforçando, com o tempo, preconceitos já incorporados em suas próprias bases de dados, optando pela alternativa mais repetida estatisticamente e fossilizando as decisões.

E, diante disso, que este ensaio tem como objetivo principal analisar a existência e a necessidade de um direito fundamental à decisão ou revisão intrinsicamente humana, que é irredutível à lógica algorítmica. Algumas questões que se colocam podem ser formuladas do seguinte modo: É possível afirmar que a decisão ou revisão intrinsicamente humana, nessa perspectiva, é um direito fundamental? Há na essência da decisão um componente necessariamente humano? É possível a redução do direito e da linguagem jurídica a uma atividade algorítmica? O ato decisório pode ser substituído por uma operação matemática? Há aspectos da função judicial

que garantam que o julgamento continue sendo uma atividade humana? A sociedade condicionaria o cumprimento da justiça por uma máquina? Esse é, portanto, o propósito que motiva e orienta o presente trabalho.

2. Da decisão humana à decisão artificial

No campo jurídico, as primeiras tentativas de resolução de problemas com o auxílio das máquinas inteligentes ocorreram em meados de 1970, mas logo esbarrou em dois obstáculos, de um lado, a enorme quantidade de dados a serem processados e formalizados, de outro, a dificuldade de traduzir a linguagem jurídica em enunciados matemáticos e lógicos (BASSOLI, 2022). Se o primeiro obstáculo parece superado com o advento das inovações dos últimos anos e o surgimento do *Big Data*, o segundo permanece atual e revela notória dificuldade em razão da impossibilidade de tradução de um objeto tão complexo como o pensamento humano e a sua linguagem.

As inovações digitais têm impulsionado, de forma vertiginosa, a adoção de sistemas de inteligência artificial no Direito. Direito e justiça atravessam uma transformação radical, não apenas tecnológica, mas também social, cultural e institucional (GARAPON; LASSÈGUE, 2018). O Poder Judiciário do Brasil e do mundo avança gradativamente no emprego de ferramentas de inteligência artificial. Novos anúncios são observados com frequência envolvendo a utilização de inteligência artificial nas rotinas relacionadas ao processo judicial e processo decisório, criando diversos desafios complexos e sem precedentes.

No cenário mundial, destacam-se experiências como o sistema de inteligência artificial da China e o projeto da Estônia. No Brasil, a realidade não é diversa, iniciativas como a solução Elis, implementada no Tribunal de Justiça de Pernambuco, o sistema voltado à execução fiscal do Tribunal de Justiça do Rio Grande do Sul e o Projeto Victor, em funcionamento no Supremo Tribunal Federal, exemplificam o caminho já percorrido.

A transição para algoritmos cada vez mais sofisticados para softwares para a resolução de problemas de várias ordens e com diferentes graus de aprendizagem e autonomia está amplamente presente em vários projetos e discussões no Poder Judiciário. Daí a urgência de refletir sobre as implicações dessa transformação para os valores constitucionais, para a legitimidade das decisões judiciais e para a própria atividade institucional do juiz.

O nível de algoritmização pode variar ao longo da trajetória, que começa com o que é conhecido como suporte à decisão até o desaparecimento de seres humanos do processo decisório. Nesse espectro, coexistem modelos que vão da decisão inteiramente humana a sistemas que, embora projetados por pessoas, funcionam de modo amplamente autônomo em relação a elas.

O atual estado da arte da evolução tecnológica já permite o desenvolvimento de um sistema computacional programado para elaborar, sem intervenção humana, uma minuta de sentença judicial. Multiplicam-se, assim, ferramentas aplicadas a análises complexas, capazes de interferir em dimensões sensíveis da vida, desde a intimidade até a concretização de direitos e oportunidades. Esse quadro desencadeia grandes preocupações sobre os riscos e consequências, sobretudo diante da dificuldade de compreensão consistente e legítima do que acontece para a tomada de determinada decisão, já que os sistemas de inteligência artificial funcionam como verdadeiras caixas pretas.

Os alcances que a inteligência artificial pode atingir ao ser aplicada no Direito, especificamente no processo decisório, são distintos. Não se sabe precisamente quais serão as consequências de tais mudanças tecnológicas. Se, por um lado, prometem eficiência e celeridade, por outro, abrem a possibilidade de que sistemas enviesados passem a ditar o futuro da sociedade.

É precisamente por isso que se impõe a presente reflexão. Não basta aceitar a promessa algorítmica, é necessário questionar de que modo os algoritmos operam, quais são os critérios que orientam suas escolhas e como se assegura a revisão humana efetiva sobre o processo decisório para não permitir uma espécie de delegação em branco à máquina e seu determinismo. O encanto da acomodação das promessas das tecnologias digitais não pode negligenciar ou levar ao esquecimento o papel do constitucionalismo, dos direitos fundamentais, dos fundamentos da Teoria da Decisão e da necessária limitação de tecnologias que abusam dos valores constitucionais rumo à desumanização.

Essa preocupação não se limita ao aspecto técnico do processo decisório: envolve também a compreensão do Poder Judiciário como instância política e social. Como define Joe McIntyre (2019), o Judiciário representa uma forma de governança social e não apenas uma instituição de resolução de conflitos. Assim, a introdução de tecnologias automatizadas no ato de julgar não altera apenas procedimentos internos, mas repercute sobre a própria estrutura da sociedade, afetando a legitimidade da jurisdição e o direito de cada cidadão de ser julgado pela singularidade de seu caso, e não pelos dados de outros processos que alimentam um sistema de inteligência artificial.

No direito todo caso é único, todo o indivíduo tem o direito de ser julgado pelos fatos de seu caso e não por meio de generalizações estatísticas derivadas de bases de dados que alimentam sistemas de inteligência artificial. Apostar em robôs, ou mesmo admitir que possam "julgar" (ou de algum modo substituir o trabalho dos juízes) significa deslocar o polo de tensão do Direito para um poder decisório opaco e automatizado: se antes o risco era o instrumentalismo, agora passa a ser a submissão ao algoritmo.

O processo decisório judicial direciona a operação de toda a sociedade, além de desempenhar um papel educativo, orientando condutas e reafirmando valores constitucionais. A função decisória é relacionada diretamente com a garantia do Estado Democrático de Direito. O ato decisório está no plano da responsabilidade política e da intersubjetividade, e, em razão disso, é que não se pode admitir a substituição de decisões humanas, representadas institucionalmente por um juiz, por decisões artificiais proferidas por um conjunto de algoritmos programados.

3. Entre a razão humana e a promessa algorítmica

Há muito tempo a literatura jurídica insiste em reafirmar a complexidade do ato de julgar. Decidir exige a interpretação do caso concreto e das normas a serem aplicadas, um trabalho que convoca conhecimento e ponderação, colocando à prova a própria inteligência humana. Para representar o desafio da busca pelo modelo ideal de juiz, apresentam-se duas figuras representativas de juízes imparciais e objetivos, contrastadas com outras figuras de juízes (MARTÍN, 2022). O primeiro é o juiz monge, retratado por Andrea Camilleri na figura de *Surra*, magistrado vindo do norte da Itália, alheio ao dialeto siciliano, à lógica da máfia e aos costumes dos grandes senhores locais. Justamente por seu isolamento e desconhecimento das conspirações, *Surra* busca aplicar a lei com aparente objetividade e imparcialidade.

O segundo modelo é o *tribunal eletrônico* encontrado no *Livro Negro* de Giovanni Papini: uma máquina instalada na parede, como um altar, capaz de emitir, a partir dos elementos fáticos apresentados, um cartão com a sentença. Eficiência e rapidez, mas sem qualquer vestígio de humanidade.

Em oposição a ambos surge a figura do juiz constitucionalmente comprometido, fruto da incorporação, nos textos constitucionais, de novas concepções de direitos e da transformação do sistema de fontes do Direito. São os juízes descritos como "guardiões de promessas", responsáveis por dar concretude às expectativas constitucionais (GARAPON e LASSÈGUE, 2018). Essa mudança se explica pela expansão do Poder Judiciário no pós-Segunda Guerra Mundial, quando os tribunais passaram a ser reconhecidos como fiadores da democracia e dos direitos fundamentais. Um dos efeitos mais visíveis desse protagonismo foi o fenômeno da judicialização da política (TATE; VALLINDER, 1997; VIANNA, 1999), que transferiu para os juízes decisões antes reservadas às arenas do Executivo e do Legislativo, chamando-os a intervir em controvérsias de natureza essencialmente política (HIRSCHL, 2008).

Nesse contexto, o papel do juiz se transforma profundamente. Se antes a imagem dominante era a de Montesquieu, o magistrado como mera "boca da lei" ou a de Jefferson, juízes como simples máquinas de aplicação nor-

mativa (DALLARI, 2008), agora a figura do juiz passa a ocupar lugar fundamental na engrenagem do Estado Constitucional, tornando-se a jurisdição constitucional fundamental (TRINDADE, 2015, p. 157-176).

No artigo *Réquiem para Ivan Ilitch: problema da interpretação do direito na literatura de Tolstói,* resgata-se o conto Um Juiz Modelo, no qual o magistrado é retratado de maneira muito mais positiva do que aquela que, décadas mais tarde, assola Ivan Ilitch. Trata-se da representação de um juiz virtuoso, preocupado, antes de tudo, em resolver os conflitos que lhe são apresentados. Para cada caso, cria procedimentos específicos destinados à clarificação dos fatos, de modo a reunir subsídios que lhe permitam alcançar a decisão correta. É, assim, um "modelo" que simboliza a possibilidade de o magistrado exercer sua função em benefício da sociedade, oferecendo respostas adequadas porque não negligencia o mundo prático (TRINDADE, 2015, p. 157-176).

O conto revela que a função judicial exige mais do que a aplicação automática de regras: demanda raciocínio jurídico próprio, que não se confunde com o raciocínio puramente analítico de matriz cartesiana. Ainda que dialogue com ele, o raciocínio jurídico vai além, abarcando outras formas de pensar que escapam à rigidez lógico-formal. Reduzir o juiz a uma "máquina de silogismos" seria retroceder à imagem descrita por Montesquieu, de uma magistratura que apenas pronúncia as palavras da lei, sem criatividade nem discricionariedade. Daí a impossibilidade de converter a decisão humana em decisão artificial. O ato de julgar envolve a subjetividade do juiz, intrínseca à condição humana, e a natural criatividade natural de seu trabalho. Exige sensibilidade e emoção, pressupõe percepção do meio pela via dos sentidos, da escuta das partes, da análise das provas, da intuição e da experiência. As decisões humanas são guiadas também pelos sentimentos, não apenas pela lógica.

O julgamento, portanto, pressupõe racionalidade, mas não qualquer racionalidade: não se trata de mera racionalidade lógica, que reduziria a atividade judicial a um silogismo. Julgar é antes uma racionalidade prática, um processo discursivo que busca e oferece os melhores argumentos para justificar a aplicação de determinada norma à solução do caso concreto. O ato de julgar como decidir pressupõe o conhecimento a respeito da Teoria da Decisão.

Esse caráter criativo do raciocínio jurídico encontra paralelo na Teoria do Direito, que o compreende como fenômeno complexo e não como mero instrumento. É nesse plano que se denuncia o risco de a inteligência artificial, tal como hoje se apresenta, não passar de um realismo. Sob a roupagem tecnológica, a chamada *startupização* do direito repete velhas teses do realismo jurídico, segundo as quais o direito nada mais seria do que a previsão empírica das decisões judiciais. Lá já se dizia: o direito é aquilo que

os tribunais afirmam que é. Assim, se a autoridade é do tribunal —e não do direito—, ele pode dizer qualquer coisa.

A questão remete à própria interpretação. Até a frase mais banal, "um relógio parado não acerta as horas", já carrega um juízo moral (STRECK, 2020). Não existe cisão entre aplicação e interpretação. O mesmo se dá com os fatos de um caso concreto: não são dados brutos, mas narrativas que emergem de um passado e chegam ao processo mediadas por perspectivas.

Se toda interpretação já carrega juízos de valor, a tentativa de converter fatos em dados numéricos revela a insuficiência da inteligência artificial para captar as nuances próprias do Direito. Ainda que as máquinas possam ser pensantes, os dados estariam dispostos em um processo de simplificação e redução de informação para a recondução do fato a dados numéricos legíveis pelo sistema de inteligência artificial. A máquina apenas reconheceria os dados considerados relevantes, fato que impossibilitaria a comprovação das pequenas diferenças e situações especificadas que diferenciam o caso dos outros processos. O sistema emitiria sentenças e decisões reafirmando as variáveis estatísticas do banco de dados programado, sem adaptação às mudanças e ao caso concreto. Some-se a isso os vieses transferidos da programação humana: tendências, preconceitos e inclinações.

Essa falibilidade levanta perguntas inevitáveis: como satisfazer a necessidade de certeza que justificaria substituir o juiz humano por um juiz-robô? Se a própria máquina é falível, pode a sociedade aceitar entregar a justiça a quem não pode responder por ela? E mais: como lidar com os precedentes? Se apenas a jurisprudência atual fosse incorporada, aceitar-se-ia a interpretação de hoje, se fossem incluídas decisões anteriores, emergiria o conflito de instruções. Em ambos os casos, o Direito perderia a sua função primordial: responder às necessidades humanas regulando as relações correspondentes às necessidades sociais do momento histórico.

Seria possível dar ao sistema instruções corretas para reconstrução de uma noção subjetiva que parece pouco compatível com a lógica binária própria da inteligência artificial? O processo ficaria a critério exclusivo do sistema de inteligência artificial, com margem de discricionariedade no exercício de suas funções. É verdade que o ordenamento jurídico está repleto de discricionariedades praticadas por juízes, mas porque, diante disso, confiar tamanho poder de escolha a uma máquina? E, sobretudo, como fazê-lo, visto que neste caso a referência a precedentes jurisprudenciais de pouco servem, pois para compreender as razões de uma ou outra opção julgada pelo sistema de inteligência artificial, deve-se conhecer os documentos e o histórico dos processos individuais. O robô realmente dirá a última palavra?

A resposta remete ao que é propriamente humano. A autonomia, expressão da consciência, não se reduz a cálculos binários. É poliédrica:

envolve o bem-estar intergeracional, a capacidade de veto aos impulsos antissociais, o autocontrole e a convivência das liberdades em prudente equilíbrio (FREITAS, 2020, p. 68). O desenvolvimento cerebral é decisivo na tomada de decisões, pois articula paradoxos, contextos e singularidades da individualidade, balizado por padrões normativos, éticos e jurídicos que orientam escolhas mais duradouras em detrimento de impulsos imediatos (SARLET, 2022).

É dessa matriz que decorre o senso moral de justiça, vocação evolucionária para equilibrar vínculos, fundada na isonomia, proporcionalidade e equidade e a compaixão, entendida como aptidão para assimilar o sofrimento alheio e hierarquizar valores (FREITAS, 2020, p. 69). Esses atributos são constitutivos da decisão humana e, até agora, inimitáveis pela inteligência artificial. O juiz decide após contato direto com os fatos: ouve testemunhas, observa partes, analisa provas, pondera argumentos. O processo é simultaneamente sensorial, racional e valorativo. A ausência de qualquer dessas dimensões compromete a completude da decisão, tornando-a defeituosa. Esses aspectos, aos quais se somará, em seguida, a análise da responsabilidade política, demonstram que a decisão judicial deve permanecer uma atividade humana. Pois julgar não é apenas aplicar regras, mas articular sentidos, princípios e valores em um contexto histórico e social. E nisso, nenhuma máquina pode substituir a presença consciente e responsável do ser humano.

4. O direito fundamental à revisão humana

Ao pensar no desenvolvimento de um juiz-robô, a questão central é se a linguagem jurídica pode ser transmutada em termos algorítmicos sem violar os direitos fundamentais constitucionalmente consagrados. Como já referido, a evolução tecnológica sustenta sistemas cada vez mais avançados, aptos a intervir na tomada de decisões judiciais. Do suporte à decisão até a ameaça da desumanização do processo, multiplicam-se experiências no Brasil e no mundo. Diante disso, impõe-se a pergunta: há algo no ato de julgar que resista à lógica algorítmica e que, por sua própria essência, garanta que a decisão continue sendo uma atividade humana?

Cada sentença, acórdão ou despacho é expressão de uma responsabilidade política que não pertence ao juiz como indivíduo moral, mas ao juiz como representante do Estado Democrático de Direito "em um Estado Democrático de Direito, os agentes políticos – e, aqui, certamente estão incluídos os juízes – tem o dever de agir corretamente, sob pena de perderem a legitimidade e a autoridade moral da qual depende seu poder de coerção" (TRINDADE, 2022, p. 315).

Responsabilidade política implica decisões fundamentadas em argumentos de princípio, que reconhecem direitos preexistentes e não a par-

tir de argumentos políticos ou subjetivos. Essa responsabilidade se ancora na exigência de coerência: aplicar princípios de modo consistente com decisões passadas e futuras, formando uma narrativa institucional que se apresenta como política em desenvolvimento. Daí decorre o dever de *accountability*, constitucionalmente consagrado no art. 93, IX, da Constituição da República Federativa do Brasil, que exige fundamentação explícita e transparente das decisões. A jurisdição é, nesse sentido, um exercício político de responsabilidade, não um espaço para voluntarismo interpretativo.

Essa exigência de fundamentação não desaparece em situações complexas. Mesmo diante de casos difíceis, em que a regra aplicável parece incerta, o juiz não está autorizado a decidir com base em critérios extrajurídicos. Seu dever é descobrir e não inventar os direitos das partes. O ato de decidir não é escolha arbitrária, mas reconstrução de princípios deontológicos que vinculam a comunidade política. Segundo Jacinto Coutinho, há algo para além da lei e que a funda (2015, p. 223). O direito é uma questão de direitos e deveres, e é "em vista disso que, politicamente, esse Estado não pode atuar de modo voluntarista, pragmatista ou consequencialista, abrindo mão dos princípios, direitos e garantias fundantes da comunidade política" (TRINDADE, 2022, p. 315).

O artigo 22º do *General Data Protection Regulation* (GDPR) [(1)] foi precisamente concebido para proteger o cidadão diante do risco de ser submetido a decisões tomadas por sistemas inteligentes com efeitos jurídicos significativos, sem qualquer intervenção humana. O legislador europeu não apenas reconheceu a imprescindibilidade dessa intervenção, mas também assegurou que todo indivíduo tenha o direito de obter a revisão humana, de expressar seu ponto de vista e de contestar a decisão automatizada. No mesmo sentido, a Itália, na *Dichiarazione dei diritti in Internet*, consagrou o direito à decisão humana, garantindo que nenhum ato administrativo ou judicial que afete de modo relevante a esfera individual possa se basear exclusivamente em processamento automatizado (ZICCARDI, 2022).

Já no Brasil, a Lei Geral de Proteção de Dados prevê, em seu art. 20, o direito de revisão e de explicação em casos de decisões automatizadas que afetem os interesses do titular. Contudo, diferentemente do modelo europeu, o legislador brasileiro não exigiu que essa revisão fosse realizada por pessoa natural. O texto original previa expressamente essa obrigação, mas a disposição foi vetada, deixando uma lacuna crítica no ordenamento.

---

(1) "O interessado tem o direito de não ser submetido a uma decisão baseada apenas no tratamento automatizado, incluindo a definição de perfis, que produza efeitos jurídicos a seu respeito ou que afete significativamente a sua pessoa de forma semelhante [...] o responsável pelo tratamento deve implementar as medidas adequadas à salvaguarda dos direitos e liberdades e interesses legítimos do titular dos dados, pelo menos o direito de obter intervenção humana por parte do responsável pelo tratamento, para expressar o seu posto de vista e contestar a decisão" (2016).

A ausência de revisão humana efetiva implica riscos severos. A intervenção concreta do ser humano, seja em sistemas simples ou sofisticados, constitui condição indispensável para que os interessados exerçam influência sobre os resultados obtidos. Sem esse filtro, decisões algorítmicas se transformam em um poder invisível, alojado em uma "caixa-preta" que se expande silenciosamente no cotidiano. Cada vez mais decisões que moldam condutas, delimitam direitos e impactam existências passam a ser tomadas por processos automatizados, escapando ao controle social e democrático.

Esses riscos evidenciam um aspecto ainda mais profundo: decidir não é o mesmo que escolher. O ato decisório não é uma operação matemática. O direito e a linguagem não podem ser reduzidos a uma atividade algorítmica. A decisão não é uma razão prática, simplória, um ato objetivo. O significado da norma só é alcançado através de um processo interpretativo, a partir de práticas sociais, costume, princípios, valores, uso da linguagem, dimensões culturais, éticas, sociais e emocionais. A textura aberta da linguagem jurídica assim o exige. A inteligência artificial, ao contrário, responde apenas a estatísticas: não sente, não interpreta, não se abre ao inesperado.

Como observa Byung-Chul Han (2021, p. 54), falta-lhe a dimensão afetiva e analógica: ela não pensa por que não conhece o arrepio, a emoção que antecede conceitos e ideias. O direito, nesse horizonte, é prática interpretativa e argumentativa, não mera técnica de aplicação mecânica de normas. Como lembra Warat (2004, p. 375) "o ensino do direito tem que se reconhecer comprometido com as transformações da linguagem, aceitar-se como prática genuinamente transgressora da discursividade instituída, como exercício de resistência a todas as formas de violência simbólica, isto é, como uma prática política do homem à sua própria existência".

A decisão jurídica não é reprodução nem criação arbitrária, mas atribuição intersubjetiva de sentido. Por isso, não se pode falar em uma fase automática da aplicação do direito: todo ato de julgar exige liberdade, criatividade e responsabilidade, algo irredutível ao padrão binário de uma máquina que apenas reage e obedece. O ato decisório situa-se no plano da responsabilidade política e da intersubjetividade, direciona a operação de toda a sociedade, além de desempenhar um papel educativo. A função decisória é relacionada diretamente com a garantia do Estado Democrático de Direito. Todo caso é único, cada cidadão tem o direito de ser julgado pelos fatos que lhe dizem respeito, e não pelas estatísticas de outros casos incorporadas a bases de dados.

É precisamente nesse reconhecimento da singularidade de cada caso que emerge a exigência de elevar a revisão humana à condição de direito fundamental. Em face dos desafios do ciberespaço e diante dos massivos conflitos estruturais da internet, ganha relevância a discussão sobre a reconstrução autônoma dos direitos constitucionais no ambiente digital (TE-

UBNER, 2017). Os direitos fundamentais são históricos e dinâmicos: nascem, evoluem e se transformam de acordo com as necessidades sociais. São vivos e reativos, como lembra Bobbio (2004), e, por isso, respondem às novas ameaças trazidas pela tecnologia. Adaptam-se a realidades inéditas, estabelecendo limites e deveres, não como meras criações formais, mas como expressões dos valores constitutivos da comunidade política.

A Era Digital demanda, assim, um novo catálogo de direitos capaz de enfrentar os riscos trazidos pelo poder econômico das *big techs*, pela governança da internet e pelo impacto dos algoritmos sobre a democracia (TEUBNER, 2017). Esse catálogo deve incluir (a) direitos clássicos; (b) direitos clássicos reconfigurados para o novo contexto digital; e (c) direitos novos, específicos do ambiente digital. A proteção dos direitos fundamentais possui dimensão material e objetiva: não basta a previsão formal no texto constitucional, é necessário assegurar sua concretização diante das novas ameaças. Como destaca Sarlet (2018, p. 123), os direitos fundamentais devem ser valorados tanto sob o ângulo individual quanto sob o ponto de vista da comunidade, pois se cuidam de valores que estruturam a vida democrática.

Mais do que uma proteção contra a submissão do indivíduo ao automatismo, ela constitui uma exigência de responsabilidade política: toda decisão que produz efeitos jurídicos sobre a vida das pessoas deve ser assumida por alguém capaz de responder por ela, de justificá-la em termos de princípios constitucionais e de prestar contas à comunidade. Sem essa intervenção, não há autonomia nem dignidade, pois a decisão se tornaria um ato impessoal e irresponsável, reduzido à mecânica estatística. Há um direito fundamental sempre que sua violação "significa a morte ou sofrimento grave, ou toca no núcleo essencial da autonomia" (ALEXY, 1999, p. 61). Assim, o direito fundamental à revisão humana não surge como mero apêndice normativo, mas como exigência constitutiva e condição estrutural da democracia constitucional, assegurando que, mesmo em uma era de algoritmos, o ato de julgar permaneça humano, intersubjetivo e politicamente responsável.

Referências bibliográficas

Alexy, Robert. Direitos Fundamentais no Estado Constitucional Democrático. Para a relação entre direitos do homem, direitos fundamentais, democracia e jurisdição constitucional. Tradução Luís Afonso Heck. Revista de Direito Administrativo, n. 217, Rio de Janeiro, jul/set 1999, p.55-66.

Bassoli, Elena. Algorimica giuridica: Intelligenza artificiale e diritto. Itália: Amon, 2022.

Barroso, Luís Roberto. Reforma digital, plataformas tecnológicas e inteligência artificial. 2025. Disponível em: https://youtu.be/vLrMgxnqKnY?si=9k68-OUsbnvMkOE7. Acesso em: 29 set. 2025.

Brecht, Bertolt. A German War Primer. Poems 1913-1956. New York: Routledge, 1997.

Chul Han, Byung. No-cosas. Penguin Random House Grupo Editorial: Barcelona, 2021.

Coutinho, Jacinto Nelson de Miranda. O lugar do poder do juiz em "Portas abertas", de Leonardo Sciascia. In: Streck, Lenio; Trindade, André Karam. Os modelos de juiz: ensaios de direito e literatura. São Paulo, Atlas, 2015.

Dallari, Dalmo de Abreu. O poder dos juízes. 3. ed. São Paulo: Saraiva, 2008.

Dworkin, Ronald. Uma questão de princípio. 2. ed. São Paulo: Martins Fontes, 2001.

Dworkin, Ronald. Levando os Direitos à Sério. São Paulo: Martins Fontes, 2002.

Dworkin, Ronald. O Império do Direito. 2. Ed. São Paulo: Martins Fontes, 2003.

Fenoll, Jordi Nieva. Inteligencia artificial y processo judicial. Madrid: Marcial Pons Ediciones Jurídicas y Sociales, S.A., 2018.

Freitas, Juarez. A heurística jurídica e a ciência do cérebro: como lidar com os automatismos mentais. Revista da AJURIS, v.40, n. 130, p.223-244, jun.2013.

Freitas, Juarez. Direito e inteligência artificial: em defesa do humano/Juarez Freitas, Thomas Bellini Freitas. Belo Horizonte: Fórum, 2020.

Garapon, Antoine; Lassègue, Jean. Justice digitale. PUF, Paris, 2018.

Hirschl, Ran. The judicialization of politics. In: Whittington, K.; Kelemen, R.; Caldeira, G. (Ed.). Oxford Handbook of Law and Politics. New York: Oxford University Press, 2008.

Martín, Nuria. A "atuação judicial automatizada" em exame: juiz robô versus juiz humano. Salvador: JusPodivm, 2022.

McIntyre, Joe. The Judicial Function Fundamental Principles of Contemporary Judging. Singapore: Springer, 2019.

Pedron, Flavio; Pereira, João Sérgio dos Santos Soares. "Quis custodiet ipsos custodes?" Ou da necessária análise concreta humana supervisionada sobre o processo de automação judicial. Salvador: JusPodivm, 2022.

Sarlet, Ingo Wolfgang. A eficácia dos direitos fundamentais: uma teoria geral dos direitos fundamentais na perspectiva constitucional. 13. ed. rev. e atual. – Porto Alegre: Livraria do Advogado, 2018.

Sarlet, Gabrielle Bezerra Sales. A Inteligência artificial no contexto atual: uma análise à luz das neurociências voltada para uma proposta de emolduramento ético e jurídico. Direito Público, 18 (100), 2022, https://doi.org/10.11117/rdp.v18i100.5214

Streck, Lenio Luiz. O neoconstitucionalismo é, de fato, uma postura pós-positivista? – A busca de uma resposta a partir da hermenêutica. In: Leite, George Salomão; Leite, Glauco Salomão; Streck, Lenio Luiz (coord.). Neoconstitucionalismo: avanços e retrocessos. Belo Horizonte: Fórum, 2017, p. 16-31.

Streck, Lenio Luiz. O que é isto – decido conforme minha consciência? 6. ed. rev. atual. Porto Alegre, 2017.

Streck, Lenio Luiz. O que é isto – o senso incomum? Porto Alegre: Livraria do advogado, 2017.

Streck, Luiz Lenio. Dicionário de hermenêutica: quarenta temas fundamentais da teoria do direito à luz da crítica hermenêutica do direito. Belo Horizonte: Letramento, 2017.

Streck, Lenio Luiz. Verdade e consenso. 6. ed. rev. e ampl. São Paulo: Saraiva, 2017.

Streck, Lenio Luiz. Para além da retórica, uma hermenêutica jurídica não relativista. In: Leite, George Salomão; Streck, Lenio Luiz (coord.). Interpretação, retórica e linguagem. Salvador: Juspodivm, 2018.

Streck, Lenio Luiz. Hermenêutica compreender Direito. São Paulo: Tirant lo Blanch, 2019.

Streck, Lenio Luiz. Dicionário de hermenêutica: 50 verbetes fundamentais da teoria do Direito à luz da Crítica Hermenêutica do Direito. 2. ed. Belo Horizonte: Letramento, 2020.

Tate, C. Neal; Vallinder, Torbjörn (Ed.). The global expansion of Judicial Power. New York: New York University Press, 1997.

Teubner, Gunther. Fragmentos constitucionais: constitucionalismo social na globalização. São Paulo: IDP/Saraiva, 2016.

Teubner, Gunther. Horizontal Effects of Constitutional Rights in the Internet: A Legal Case on the Digital Constitution. Italian Law Journal, v. 3, n. 2, p. 485–510, 2017.

Trindade, André Karam; Resenfield, Luis. Réquiem para Ivan Ilitch: O Problema Da Interpretação Do Direito Na Literatura De Tolstói. Revista da Faculdade de Direito – UFPR, Curitiba, vol. 60, n. 2, maio/ago. 2015, p. 157-176.

Trindade, André Karam. O controle das decisões judiciais e a revolução hermenêutica no direito processual civil brasileiro. In: Streck, Lenio Luiz; Alvim, Eduardo Arruda; Leite, George Salomão (coord.) Hernenêutica e jurisprudência no Código de Processo Civil: coerência e integridade. 2. ed. São Paulo: Saraiva Educação, 2018.

Trindade, André Karam. Garantismo versus neoconstitucionalismo: os desafios do protagonismo judicial em terrae brasilis. In: Ferrajoli, Luigi et al (Eds.). Garantismo, hermenêutica e (neo)constitucionalismo: um debate com Luigi Ferrajoli. 2. ed. São Paulo: Tirant lo Blanch, 2022, p. 93-127.

Warat, Luis Alberto. Epistemologia e ensino do Direito: o sonho acabou. Florianópolis: Fundação Boiteux, 2004.

Ziccardi, Giovanni. Diritti digitale: Informatica giuridica per le nuove professioni. Milano: Rafaello Cortina Editore, 2022.

Sección II

Profesión y enseñanza jurídica

El futuro de la inteligencia artificial generativa en el derecho

Por Antonio Anselmo Martino (*)

Resumen: La inteligencia artificial generativa (IAG) está redefiniendo el panorama jurídico mediante la automatización avanzada de tareas operativas —como redacción de documentos, análisis de jurisprudencia y gestión masiva de expedientes— y la introducción de asistentes virtuales interactivos que democratizan el acceso a la justicia, especialmente para poblaciones vulnerables. Sistemas como Prometea (Argentina) evidencian ganancias tangibles: reduce un 74% el tiempo en elaboración de dictámenes y aumenta un 464% la producción de sentencias mensuales mediante cobotización judicial, un modelo colaborativo humano-máquina que potencia —no reemplaza— la labor del jurista. No obstante, esta transformación enfrenta desafíos críticos: Responsabilidad legal difusa: Errores en recomendaciones automatizadas generan interrogantes sobre imputabilidad (desarrolladores, instituciones o usuarios finales). Sesgos algorítmicos y opacidad: La replicación de prejuicios históricos en datos de entrenamiento y las "cajas negras" exigen auditorías continuas y equipos multidisciplinarios para garantizar equidad. Regulación heterogénea: Mientras la Unión Europea implementa un enfoque basado en riesgos (prohibiendo usos "inaceptables" como vigilancia biométrica indiscriminada), países como Chile priorizan derechos neurodigitales y Colombia establece guías éticas para IA pública y privada. La IAG también impulsa innovaciones disruptivas: Agentes autónomos jurídicos que realizan análisis predictivo de litigios con base en patrones históricos. Gobernanza algorítmica para decisiones institucionales transparentes y auditables. Integración con ERP/CRM que optimizan gestión corporativa (ej. Cuatrecasas en España). Futuros desarrollos requerirán: Reformas educativas en facultades de derecho para incluir ética digital, *prompting* jurídico y análisis de datos. Marcos normativos ágiles que equilibren innovación con derechos fundamentales, evitando "cajas negras" mediante explicabilidad obligatoria. Protocolos de ciberseguridad reforzados ante amenazas como *phishing* hiperpersonalizado o ataques con

(*) Profesor emérito de la Universidad del Salvador (Argentina) y de la Universidad de Pisa (Italia). Miembro de la Academia Nacional de Derecho y Ciencias Sociales de Córdoba. Miembro de la Academia Nacional de Ciencias de Buenos Aires. Miembro asociado del Center for Artificial Intelligence and Cognate learning of the University of Greenwich, miembro de la "European Association of Legislation".

IA. El caso "Prometea" —reconocido por la ONU y la OCDE— sintetiza este paradigma: combina eficiencia procesal (1,000 dictámenes en 45 días) con supervisión humana irrenunciable en decisiones sustantivas. Su éxito radica en entender la IAG como infraestructura habilitante —similar a la electricidad—, no como fin en sí misma. La colaboración humano-algorítmica emerge, así como modelo inevitable, pero su sostenibilidad dependerá de resolver tensiones éticas y evitar una fetichización tecnocrática del derecho.

Palabras claves: Automatización jurídica predictiva, cobotización judicial, ética algorítmica, gobernanza regulatoria, agentes autónomos legales.

Abstract: Generative AI (GAI) is reshaping legal systems through advanced automation of operational tasks —such as document drafting, jurisprudence analysis, and mass case management— and interactive virtual assistants that democratize access to justice, particularly for vulnerable populations. Systems like Prometea (Argentina) demonstrate tangible gains: it reduces report processing time by 74% and increases monthly rulings by 464% via judicial cobotization, a human-machine collaboration model that enhances —without replacing— legal professionals. Yet critical challenges persist: Diffuse liability: Errors in automated recommendations raise questions about accountability (developers, institutions, or end-users). Algorithmic bias and opacity: Historical prejudices embedded in training data and "black box" systems demand continuous audits and multidisciplinary teams to ensure fairness. Regulatory heterogeneity: While the European Union implements a risk-based approach (banning "unacceptable" uses like biometric surveillance), countries like Chile prioritize neuro-rights, and Colombia establishes ethical guidelines for public/private AI. GAI also enables disruptive innovations: Autonomous legal agents conducting predictive litigation analysis based on historical patterns. Algorithmic governance for transparent, auditable institutional decisions. ERP/CRM integration optimizing corporate legal management (e.g., Cuatrecasas in Spain). Future development requires: Educational reforms in law schools covering digital ethics, legal prompting, and data analysis. Adaptive regulatory frameworks balancing innovation with fundamental rights, mandating explainability. Enhanced cybersecurity protocols against AI-powered threats like hyper-personalized phishing. The *Prometea* case —endorsed by the UN and OECD— epitomizes this paradigm: it combines procedural efficiency (1,000 reports in 45 days) with indispensable human oversight in substantive decisions. Its success lies in framing GAI as enabling infrastructure —akin to electricity— rather than an end in itself. Human-algorithmic collaboration thus emerges as an inevitable model, though its sustainability hinges on resolving ethical tensions and avoiding technocratic fetishization of law.

Key words: Predictive legal automation, Judicial cobotization, Algorithmic accountability, Regulatory governance, Autonomous legal agents.

1. El futuro de la inteligencia artificial generativa

En el futuro inmediato el fenómeno científico tecnológico de la inteligencia artificial parece que seguirá su ritmo arrollador insinuándose en todos los aspectos de la actividad humana.

Se pueden vislumbrar algunas características: la IA generativa se consolidará como la base de la mayoría de los procesos empresariales y sociales, obviamente también jurídicos. SEIDOR apunta asimismo la consolidación de la nube será la base de la innovación y a la importancia de la *total experience* para conectar clientes y empleados. (1)

La tecnología se basará en pilares clave como la IA, el *cloud* (2) y la ciberseguridad. Estos seguirán cambiando las organizaciones y la sociedad. Además, se impulsará con herramientas más naturales, accesibles y personalizadas para mejorar la experiencia humana. Al mismo tiempo, la computación cuántica (3) comenzará a resolver desafíos complejos, ampliando las capacidades tecnológicas, mientras que la sostenibilidad y la ética tecnológica se consolidan como valores esenciales para un futuro responsable e inclusivo.

I. En 2025, la IA generativa se consolidará como la base de la mayoría de los procesos empresariales y sociales. Su habilidad para analizar datos estructurados y no estructurados en tiempo real ayudará a crear soluciones personalizadas. Así, optimizará decisiones y flujos de trabajo. Desde la hiperproductividad personal hasta la mejora en la atención al cliente, la educación y la investigación, la IA generativa será clave para resolver desafíos complejos y anticipar las necesidades de las organizaciones. (4)

II. En 2025, el *cloud* se consolidará como uno de los pilares fundamentales de la transformación digital. También será el motor clave para adoptar tecnologías como la IA generativa y la computación cuántica. Esto permitirá a las empresas escalar sus operaciones de forma más eficiente y competitiva. Asimismo, el modelo de nube híbrida seguirá destacándose como una

---

(1) En los próximos cinco años, se espera que el crecimiento de los centros de datos en América Latina crezca a una tasa anual compuesta de 7,95% según *Research and Markets*.

(2) La nube o *cloud computing* se refiere a la computación en la nube, un servicio que permite el acceso a datos, aplicaciones y recursos a través de Internet. Es utilizada por empresas para ejecutar aplicaciones y tomar decisiones, y también por usuarios individuales para almacenar fotos y archivos, como en iCloud. Los cambios se sincronizan entre dispositivos.

(3) La computación cuántica utiliza tecnología especializada, incluido el hardware informático y los algoritmos que aprovechan la mecánica cuántica, para resolver problemas complejos que los ordenadores clásicos o los superordenadores no pueden resolver, o no pueden resolver lo suficientemente rápido. Los problemas complejos son problemas con muchas variables que interactúan de forma complicada. El mundo real funciona con física cuántica. Los ordenadores que realizan cálculos utilizando los estados cuánticos de los bits cuánticos deberían ser, en muchas situaciones, nuestras mejores herramientas para entenderlos.

(4) Desde mitad de abril los ciudadanos europeos que usan Meta AI reciben una comunicación explicando en que modo sus datos públicos serán utilizados para mejorar el AI de Meta.

de las estrategias más valiosas, ya que permite a las organizaciones integrar infraestructuras locales con servicios en la nube. Esta combinación facilita la gestión de grandes volúmenes de datos en tiempo real, la adaptación a necesidades específicas y la optimización de la eficiencia operativa.

III. La experiencia total (TX) será clave en 2025. Integrará la experiencia del cliente (CX) y del empleado (EX). También tomará en cuenta la experiencia del usuario (UX) y la multicanalidad. Impulsada por la IA generativa, permitirá analizar grandes volúmenes de datos para identificar patrones y personalizar interacciones en tiempo real. Por ejemplo, sistemas de atención al cliente basados en IA capaces de interpretar el tono emocional de una consulta, ofreciendo respuestas más empáticas y precisas; o herramientas colaborativas impulsadas por IA facilitarán la experiencia de los empleados, automatizando tareas repetitivas y liberándolos para concentrarse en actividades estratégicas.

IV. La ciberseguridad afrontará nuevas amenazas, como el phishing personalizado y los ataques basados en inteligencia artificial. La IA generativa permitirá identificar y neutralizar estas amenazas antes de que causen daños significativos. Además, la criptografía poscuántica se consolidará como una herramienta clave para proteger datos sensibles frente a los avances de la computación cuántica. Estas innovaciones garantizarán la confianza en sistemas digitales en un entorno cada vez más complejo.

V. Los agentes autónomos de IA se integrarán como aliados imprescindibles en el futuro inmediato. Estos sistemas automatizarán tareas rutinarias y, además, analizarán grandes volúmenes de datos para tomar decisiones estratégicas. Su habilidad para aprender de forma continua les ayudará a trabajar bien con humanos. Esto maximiza la eficiencia en la gestión de recursos, la planificación de proyectos y la atención personalizada a clientes.

VI. La relación entre humanos y máquinas será más empática y natural. Los avances en el procesamiento del lenguaje natural y la interpretación de emociones ayudarán a los sistemas tecnológicos. Así, podrán entender palabras, expresiones faciales, tonos de voz y gestos. Esto permitirá que avatares digitales y robots colaborativos ofrezcan un apoyo más personalizado. Esto se verá en áreas como la sanidad, la educación y la atención al cliente. Así, se eliminarán barreras tecnológicas y se fomentará un acceso más inclusivo.

VII. Los ERP [5] y CRM transformarán la gestión empresarial con la integración de IA generativa. Estas herramientas mejorarán procesos importantes. Esto incluye la planificación de producción, la gestión de inventarios

(5) ERP son las siglas en inglés de Enterprise Resource Planning, que en español significa "planificación de recursos empresariales". Se trata de un software que ayuda a gestionar las actividades de una empresa

y las ventas. Estos sistemas permitirán a las organizaciones anticiparse a sus necesidades, diseñar estrategias a medida de cada compañía y mejorar la eficiencia operativa en todos los niveles.

VIII. La computación espacial integrará datos digitales en tiempo real con el entorno físico, transformando la interacción entre ambos. La realidad aumentada y la realidad virtual crearán experiencias inmersivas. Esto afectará sectores como el diseño, la industria y la formación. Estas herramientas aumentarán la productividad, así como una comprensión más profunda de datos complejos y conectarán a las personas en espacios híbridos.

IX. El desarrollo comercial de la computación cuántica se acelerará con la transición de los *qubits* (6) físicos a los lógicos, que ofrecerán mayor fiabilidad y aplicaciones a escala industrial. Este avance creará nuevas aplicaciones. Pero, también traerá desafíos. Habrá una carrera entre el descifrado cuántico y el cifrado postcuántico. Además, una gran potencia computacional puede afectar la seguridad de los mercados de *blockchain*, como las criptomonedas y las finanzas distribuidas. A pesar de estos retos, este salto tecnológico impulsará avances en áreas como la medicina o la energía, haciendo más tangible la dimensión humana de la tecnología.

X. La sostenibilidad seguirá siendo un pilar fundamental. Los centros de datos adoptarán fuentes de energía renovable y soluciones más eficientes, reduciendo su huella de carbono. Algoritmos optimizados y sistemas de computación sostenible minimizarán el consumo energético de hardware y software, reforzando un progreso tecnológico que se alinee con la preservación del medio ambiente.

XI. ¿La IA generativa se va a comportar como la electricidad hoy?

2. El futuro de la IA generativa en el derecho

El uso de la IA generativa producirá necesariamente transformaciones en los procesos judiciales:

- Automatización de tareas rutinarias: redacción de documentos, análisis de jurisprudencia, revisión de pruebas y gestión de expedientes podrían delegarse en agentes inteligentes que aprenden y se adaptan al estilo del tribunal.

---

(6) Un cúbit (o bit cuántico) es la unidad fundamental de información en computación cuántica. A diferencia de un bit clásico, que solo puede ser 0 o 1, un cúbit puede existir en una superposición de ambos estados. Esto significa que un cúbit puede ser 0, 1, o una combinación lineal de ambos, lo que permite a las computadoras cuánticas realizar cálculos mucho más complejos y rápidos que las computadoras clásicas. Los cúbits aprovechan los principios de la mecánica cuántica, como la superposición y el entrelazamiento, para procesar información de maneras innovadoras

- Asistentes virtuales y avatares en audiencias: podrían facilitar la participación de personas con dificultades de acceso al sistema judicial, brindando acompañamiento legal personalizado en entornos virtuales.

- Predicción de decisiones judiciales: algoritmos entrenados en decisiones previas podrían generar modelos predictivos sobre el probable desenlace de un litigio. Esto, aunque útil, plantea desafíos éticos profundos sobre la presunción de inocencia y el rol del juez humano.

2.1. Aplicación del derecho en la sociedad

Democratización del acceso a la justicia: chatbots legales y sistemas basados en IA podrían orientar a la ciudadanía sobre derechos, trámites y reclamos, especialmente en comunidades vulnerables.

Si los datos de entrenamiento reflejan prejuicios históricos, las decisiones automatizadas podrían perpetuar desigualdades sociales. Pero esto sucede ya y nos ha enseñado que para evitar estos sesgos es conveniente universalizar los grupos de trabajo tanto como sea posible: con respecto al sexo, a la raza, a la nacionalidad, a las profesiones de proveniencia, etc. Nunca estaremos seguros de no tener prejuicios, pero estos cuidados minimizan los riesgos, porque solo no siendo varón puedes ver los prejuicios masculinos, solo no siendo abogado puedes ver los sesgos jurídicos, etc.

Los abogados tendrán que desarrollar nuevas competencias en ética tecnológica, análisis de datos y supervisión de sistemas inteligentes. Obviamente tendrán que ser reconfigurados los programas en la Facultades de derecho, pues los actuales todavía trasudan cultura obsoleta.

2.2. Desafíos y consideraciones éticas

¿Quién responde por los errores de una IA que asesora o decide? ¿Cómo se audita su funcionamiento? Los agentes deben operar bajo principios que permitan entender cómo y por qué actúan, especialmente cuando influyen en derechos fundamentales. Cada vez se usarán más los agentes, pero no pueden tener cajas negras sino ser totalmente verificables.
Será clave establecer marcos normativos que regulen el uso, desarrollo y límites de la IA en contextos jurídicos.

2.3. Seguirán desarrollándose las líneas teóricas sobre IA generativa en el derecho

La teoría del derecho algorítmico examina cómo, los sistemas automatizados pueden generar, interpretar o aplicar normas jurídicas. Se vincula con el concepto de “normas computables” y plantea que parte del derecho puede ser reducida por la lógica.

Desde la sociología jurídica, se considera que el derecho debe evolucionar junto con los cambios tecnológicos. La IA generativa tensiona esta capacidad adaptativa, especialmente en contextos judiciales.

La teoría de la responsabilidad distribuida propone que, en sistemas complejos como los que involucran IA, la responsabilidad legal debe ser compartida entre desarrolladores, usuarios y supervisores humanos.

La gobernanza algorítmica estudia cómo los algoritmos pueden influir en decisiones públicas y privadas, y cómo deben ser regulados para preservar derechos fundamentales.

2.4. Modelos regulatorios comparados

Región / País	Enfoque principal	Ejemplo destacado
Unión Europea	Regulación por niveles de riesgo (mínimo, limitado, alto, inaceptable).	Ley de Inteligencia Artificial con prohibiciones para vigilancia biométrica.
Chile	Participación ciudadana y enfoque ético.	Política Nacional de IA con consulta pública y proyectos de ley sobre neuroderechos.
Colombia	Guía ética para IA en sectores públicos y privados.	Marco ético para la IA con principios de equidad y transparencia.
Uruguay	Derecho a oponerse a decisiones automatizadas.	Ley de Protección de Datos con cláusulas específicas sobre IA.
Argentina	En proceso de desarrollo normativo, con iniciativas como Prometea y LegisAPP Lite.	Estrategia Nacional de IA y debates sobre privacidad, responsabilidad y transparencia.

2.5 ¿Qué son los agentes inteligentes en el derecho?

Los agentes son sistemas autónomos o semiautónomos que pueden ejecutar tareas específicas, interactuar con otros sistemas o usuarios, y tomar decisiones dentro de un marco definido. En el ámbito jurídico, se perfilan como copartícipes activos en procesos legales, no solo como asistentes pasivos.

Se parte del enfoque de los sistemas jurídicos como estructuras dinámicas que responden a estímulos tecnológicos (Teubner, 1988; Latrubesse,

2022). En este marco, la incorporación de agentes inteligentes representa una evolución natural del ecosistema legal, obligando a repensar el diseño institucional y el rol de sus actores.

Las funciones emergentes de los agentes jurídicos comprenderán la redacción automatizada de documentos [7], el análisis predictivo de litigios como evaluación de probabilidades de éxito, riesgos procesales y sugerencias estratégicas basadas en casos similares. La auditoría normativa, consistente en la verificación de cumplimiento legal en documentos, políticas internas o decisiones administrativas.

La interacción con usuarios podrá hacerse a través de agentes conversacionales que orientan a ciudadanos sobre trámites, derechos y recursos disponibles. [8]

Deberá planificarse la gestión de flujos de trabajo legales coordinando las tareas entre distintos actores del proceso judicial (jueces, fiscales, defensores, administrativos) [9].

En el Poder Judicial se producirá la integración en sistemas de gestión de expedientes, redacción de borradores y análisis de jurisprudencia con consultoría legal automatizada que ofrecen plataformas con orientación preliminar a ciudadanos, con derivación a profesionales humanos, cuando se requiera. [10]

Los beneficios esperados son ahorro de tiempo, reducción de tareas repetitivas, mejora en la precisión, menos errores en documentos y análisis, acceso ampliado a la justicia, orientación legal para sectores vulnerables, toma de decisiones informada, apoyo estratégico basado en datos históricos

Riesgos y desafíos: los sesgos de los humanos autores de los programas pueden pasar a los sistemas algorítmicos, si no se auditan. Ver bien cómo justificar una recomendación automatizada ante un tribunal y ¿quién responde por errores del agente: ¿el desarrollador, el usuario o la institución? Así como el riesgo de exposición de datos sensibles si no se implementan protocolos seguros.

La integración de agentes generativos exige repensar el diseño institucional del derecho. Se vislumbra una arquitectura híbrida, donde humanos

---

(7) Contratos, demandas, dictámenes y resoluciones generados por agentes entrenados en lenguaje jurídico y jurisprudencia local

(8) Una enseñanza que nos deja la experiencia hasta el presente es ver donde poner un humano para que la cadena delegativa no se vuelva un sinfín.

(9) Escenarios de implementación inmediata: Departamentos legales corporativos: Agentes que revisan contratos, alertan sobre cláusulas sensibles y sugieren modificaciones

(10) Ejemplo destacado: El Programa Piloto de IA Generativa en la Justicia Argentina ya ha testeado agentes que identifican agravios en recursos de inconstitucionalidad en menos de 3 minutos, con una eficiencia del 90%, lo que a un humano le lleva por lo menos tres días.

y agentes colaboran en tiempo real, con roles diferenciados pero complementarios. Esto implica: Nuevas competencias profesionales (*prompting* jurídico, auditoría de IA, ética digital), marcos normativos adaptativos que regulen el uso de agentes según su autonomía y nivel de impacto, protocolos de supervisión humana obligatoria en decisiones sensibles.

3. Casos prácticos emergentes

Dado que todo estudio prospectivo del desarrollo de la IA generativa en el derecho tiende a desarrollar las líneas teóricas, vamos a enfocar algunos casos prácticos:

- Prometea (Argentina): Sistema de IA que automatiza tareas judiciales en el Ministerio Público Fiscal. Mejora la eficiencia, pero no reemplaza decisiones judiciales humanas.

- *Ross Intelligence* (EE.UU.): Herramienta que analiza jurisprudencia y ayuda en la redacción de argumentos legales. Fue retirada por conflictos de copyright, lo que abrió debate sobre límites legales de la IA.

- Cuatrecasas (España): Firma legal que usa IA para gestión del conocimiento y revisión documental. Reconocida por el Financial Times por innovación legal.

- IA generativa en departamentos legales corporativos: Estudios recientes muestran que hasta el 44% del trabajo legal puede ser automatizado. Sin embargo, se identifican riesgos como sesgos, alucinaciones y falta de comprensión contextual.

3.1 Cobotización [11] judicial y transformación institucional: Estudio de caso del sistema Prometea en Argentina

La irrupción de la inteligencia artificial (IA) en el ámbito jurídico ha generado profundas transformaciones en la forma en que se concibe, practica y administra el derecho. En América Latina, el sistema Prometea, desarrollado por el Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires, constituye un modelo pionero de aplicación de IA en procesos judiciales. Lejos de promover la automatización total de la justicia, Prometea opera bajo una lógica de cobotización inclusiva, entendida como la colaboración entre humanos y máquinas en tareas repetitivas que optimizan los tiempos institucionales sin desplazar el razonamiento jurídico. Este artículo propone un análisis crítico del funcionamiento de Prometea, su impacto en la estructura judicial argentina y las implicancias éticas asociadas a su implementación.

---

(11) La cobotización es un mayor trabajo colaborativo entre humanos y máquinas

Prometea es un sistema basado en procesamiento de lenguaje natural y aprendizaje automático que permite elaborar dictámenes jurídicos, resoluciones contravencionales y sentencias fiscales en tiempos drásticamente menores, con una tasa de acierto superior al 95% (Ministerio Público Fiscal CABA, 2023). Su interfaz conversacional y control de calidad.

Es un sistema de inteligencia artificial desarrollado por el Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires en colaboración con el Laboratorio de Innovación e Inteligencia Artificial de la Facultad de Derecho de la UBA. Su objetivo es automatizar tareas judiciales repetitivas, mejorar la eficiencia institucional y liberar tiempo para el análisis jurídico profundo documental. Según cifras oficiales ha reducido los errores de tipeo en un 99%, puede anticipar el resultado de un caso en menos de 20 segundos con una tasa de acierto del 96%. En 45 días elaboró 1000 dictámenes en forma digital sobre derecho a la vivienda, tarea que manualmente tomaría 174 días. Asistencia en procesos contravencionales: En casos de conducción en estado de ebriedad, puede generar 1000 resoluciones en 26 días, frente a los 110 días del método tradicional. En ejecuciones fiscales permite pasar de 255 sentencias mensuales a 1440.

Opera mediante comandos de voz o chat, similar a Siri o WhatsApp, reduce errores de tipeo en un 99% y verifica aspectos formales.

Costabel et al. (2021) destacan que “Prometea introduce una nueva lógica de colaboración inteligente en el proceso judicial, donde el razonamiento humano se ve asistido, no reemplazado”.

Según el Informe de Resultados institucionales (MPF CABA, 2023), “la implementación del sistema permitió aumentar en más del 400% la producción de sentencias mensuales sin incrementar la dotación de personal”.

La noción de cobotización se inscribe en la teoría de los sistemas adaptativos, que concibe el derecho como una estructura evolutiva sensible a transformaciones tecnológicas (Latrubesse, 2022). En esta línea, Prometea funciona como un agente auxiliar que optimiza tareas burocráticas para liberar tiempo del operador humano, manteniendo intacto su rol deliberativo. Esta perspectiva se aleja de modelos automatizados de justicia predictiva que intentan reemplazar el juicio humano mediante algoritmos.

4. Consideraciones éticas y desafíos regulatorios

La eficiencia de Prometea plantea interrogantes cruciales en torno a ¿cómo se explica una decisión sugerida por el sistema? O ¿qué sucede si los datos que alimentan el sistema reproducen desigualdades sociales o jurídicas? O aun ¿quién responde por una recomendación automatizada errónea?

Como subraya el informe del AI for Good Summit (ONU, 2023), "los sistemas de IA en justicia deben operar bajo marcos explicativos accesibles que permitan auditar su razonamiento y corregir sesgos".

Prometea ha sido implementado en más de 60 organismos públicos nacionales e internacionales, incluyendo la Corte Interamericana de Derechos Humanos, la Corte Constitucional de Colombia, el Registro Civil de CABA y aplicaciones en provincias argentinas como Mendoza, Chaco, Corrientes y Santa Fe.

Impacto medible

Tarea	Tiempo tradicional	Tiempo con Prometea	Reducción estimada
Dictámenes jurídicos	174 días	45 días	74%
Resoluciones contravencionales	110 días	26 días	76%
Sentencias fiscales	255/mes	1440/mes	+464%
Rectificatorias administrativas	8 meses	2 meses	75%

Prometea ofrece una vía prometedora para repensar la administración de justicia en clave de colaboración tecnológica, con impacto positivo en eficiencia y productividad institucional. No obstante, su consolidación como modelo replicable exige evitar una dependencia tecnocrática o una fetichización del algoritmo como garante de neutralidad. La justicia, al estar basada en principios de equidad, deliberación y contexto, no puede delegarse por completo en sistemas automatizados. En este sentido, Prometea constituye un laboratorio jurídico-tecnológico cuya legitimidad dependerá del equilibrio entre innovación y salvaguardas éticas. Prometea se basa en el principio de cobotización, es decir, colaboración entre humanos y máquinas. No reemplaza al profesional jurídico, sino que lo potencia, permitiéndole enfocarse en tareas complejas, creativas y empáticas.

Prometea ha sido presentado en la Cumbre Mundial de Inteligencia Artificial para el Bien (ONU), y reconocido por instituciones como Oxford, Sorbona, OEA y OCDE como un caso de éxito en justicia digital.

5. Conclusiones

Para tratar el futuro de la inteligencia artificial generativa en el derecho preferimos partir una prospectiva de la IA generativa en general, desarrollar lo que creemos sucede y sucederá con mayor vigor en los próximos años

de esta década [12] en el derecho, dar 4 casos de casos prácticos emergentes, para terminar analizando uno en particular que es Prometea, porque lo consideramos relevante y tenemos además la suerte de tenerlo cerca pues fue elaborado por desarrollado por el Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires, con la colaboración del laboratorio de inteligencia artificial de la Facultad de derecho de la UBA.

El cuadro emergente es complejo, con variaciones notables en el proceso judicial, así como el uso del derecho en la sociedad. Es fácil predecir que el proceso continuará con ritmo cada vez mayor, no pudiéndose ver bien los efectos que esto determinará en la sociedad y aun en el rol jurídico en sí mismo.

Creemos que la IA generativa será cada vez más como hoy es la electricidad y debiendo estar muy atentos a auditar las sucesivas aplicaciones para evitar los riesgos de tener cajas negras que empañen la transparencia o indeterminaciones que oscurezcan la aplicación de los derechos humanos fundamentales. Se irá acortando la distancia entre el actuar humano y el digital, sobre todo con el papel que irán tomando los agentes y avatares.

La esperanza es la colaboración y complementación, pero sería muy ingenuo pensar que ello será fácil y natural cuando es la primera vez en la historia humana que una creación que puede ser vista como un instrumento crea conocimiento y para ser eficaz irá exigiendo mayor autonomía.

Dos reflexiones generales que se aplican al futuro del derecho: siempre que construimos un programa de computación soñamos con prever todos los casos posibles. Esto no es ni teórica ni prácticamente posible, pero es difícil renunciar a esa aspiración. Tenerlo presente evita caer en los procesos reiterativos que terminan en la nada. En la automatización con IA generativa, cuando se sospeche que el sistema tiende a caer en alguno de esos procesos, lo mejor es derivarlo a un humano que ayuda a retomar el iter (camino o sendero).

La siguiente es hija de esta: en la IA generativa el sistema raramente dirá "no se nada sobre este tema" sino que buscará siempre una salida posible. Esto lleva a la fabulación o, dicho más crudamente, a groseros errores.

En el caso práctico repárese que ya el nombre del producto es disruptivo pues está basado en el mito de Prometeo, pero se llama Prometea.

---

(12) Creemos que la velocidad irá en aumento, por lo que pensar en los desarrollos de la década del treinta de este siglo no parece apresurado.

Bibliografía

AI for Good Global Summit.Seidor. (2024). Tendencias tecnológicas 2025: Cloud, IA generativa y experiencia total. https://www.seidor.com

BBC Mundo. (2025). Qué se espera en 2025 de la inteligencia artificial, el avance que marcó un antes y un después. Recuperado el 1 de agosto de 2025, de https://www.bbc.com/mundo/articles/c4gxzx0kpp6o

Computer Weekly. (2025). El futuro de la IA generativa: 10 tendencias a seguir en 2025. Recuperado el 1 de agosto de 2025, de https://www.computerweekly.com/es/cronica/El-futuro-de-la-IA-generativa-10-tendencias-a-seguir-en-2025

Costabel, M., García, L., & Fernández, R. (2021). Nuevas lógicas de colaboración inteligente en procesos judiciales. Revista Iberoamericana de Derecho Digital.

Datilab. (2025). IA Generativa en 2025: El Futuro de la Innovación Tecnológica. Recuperado el 1 de agosto de 2025, de https://datilab.com/blog/ia-generativa-en-2025-el-futuro.

deskbird. (2025). La IA y el futuro del trabajo: tendencias y estadísticas para 2025. Recuperado el 1 de agosto de 2025, de https://www.deskbird.com/es/blog/futuro-del-trabajo-con-ia

European Institute of Innovation and Technology (EIMT). (2025). The Future of Generative AI: Trends to Watch in 2025 and Beyond. Recuperado el 1 de agosto de 2025, de https://www.eimt.edu.eu/the-future-of-generative-ai-trends-to-watch-in-2025-and-beyond

Gartner. (2025). The 2025 Hype Cycle for Artificial Intelligence Goes Beyond GenAI. Recuperado el 1 de agosto de 2025, de https://www.gartner.com/en/articles/hype-cycle-for-artificial-intelligence

Harvard Business Review. (2025). How People Are Really Using Gen AI in 2025. Recuperado el 1 de agosto de 2025, de https://hbr.org/2025/04/how-people-are-really-using-gen-ai-in-2025

Herrera, F. (2025). El Futuro de la Inteligencia Artificial: Tendencias Claves 2025. Recuperado el 1 de agosto de 2025, de https://fabianherrera.net/el-futuro-de-la-inteligencia-artificial-tendencias-claves-2025

Informe de Resultados Institucionales 2023. Naciones Unidas (ONU). (2023). Recomendaciones éticas para sistemas de IA en justicia.

Latrubesse, D. (2022). Sistemas jurídicos adaptativos y transformaciones tecnológicas. Editorial Jurídica Continental.

Ministerio Público Fiscal CABA. (2023). Memoria técnica del sistema Prometea v4.1. https://prometea.com.ar

Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires (MPF CABA). (2023).

MIOTI Tech & Business School. (2025). Futuro de la IA Generativa: Desafíos, oportunidades y transformación social. Recuperado el 1 de agosto de 2025, de https://mioti.es/es/blog-futuro-de-la-ia-generativa-desafios-oportunidades-y-transformacion-social

Stanford Institute for Human-Centered Artificial Intelligence (HAI). (2025). Artificial Intelligence Index Report 2025. Recuperado el 1 de agosto de 2025, de https://hai.stanford.edu/ai-index/2025-ai-index-report

Teubner, G. (1988). Derecho como sistema autopoiético. Universidad de Florencia.

La construcción de *datasets* jurisprudenciales en el sistema judicial y los colaboradores expertos. Características y retos de un nuevo rol

Por Helga María Lell (*)

Resumen: Este trabajo se centra en los *datasets* jurisprudenciales ya que son relevantes para entrenar sistemas de inteligencia artificial y se centra en la relación entre los conjuntos de datos y un nuevo rol para los expertos que acompañan o deben acompañar en el sistema judicial. Aquí se sostiene que la calidad, estructura e integridad de los *datasets* jurisprudenciales (como colecciones de decisiones judiciales) determinan la utilidad de la IA. Se destaca el papel novedoso de los expertos interdisciplinares en la construcción, la anotación y la validación de los conjuntos de datos, pero también se enfatizan los retos. De este modo, los expertos garantizan la exactitud jurídica y técnica, detectan sesgos, salvaguardan la privacidad de los datos mediante la anonimización y explican los resultados de la IA a los actores judiciales. Para ello, en primer lugar, se definirán y caracterizarán los conjuntos de datos y, a continuación, se mencionarán los diferentes grados de complejidad de los *datasets* jurisprudenciales. En segundo lugar, se presentan las funciones y los retos de los expertos en el diseño, la conservación y la publicación de los conjuntos de datos. Se concluye que la adopción de la IA en el Poder Judicial no puede prescindir de la intervención de expertos en la construcción de *datasets* jurisprudenciales.

Palabras clave: *Datasets;* jurisprudencia; desafíos; expertos.

Abstract: This paper examines datasets in judicial contexts as they may be relevant for training artificial intelligence systems used for different purposes, focusing on the relationship between datasets and a new role for experts in the judicial system. It argues that the quality, structure, and

---

(*) Conicet; UNLPam, FCEyJ, CICJ, Santa Rosa, Argentina. helgalell@conicet.gov.ar. Profesora titular de Filosofía del Derecho y de Argumentación Jurídica y Debate en la Facultad de Ciencias Económicas y Jurídicas de la Universidad Nacional de La Pampa. Investigadora adjunta de Conicet. Posdoctora, doctora en Derecho, magíster en Filosofía, magíster y especialista en Estudios Sociales y Culturales, abogada. Diploma Superior en Construcción de Proyectos y Metodología en Ciencias Sociales, diploma en Humanidades Digitales y experta en Minería de Textos con Python.

integrity of jurisprudential datasets (such as collections of court decisions) determine the usefulness of AI in judicial proceedings. It highlights the novel role of interdisciplinary experts in the design, preservation, annotation, and validation of datasets, but also emphasizes the challenges. In this way, experts ensure legal and technical accuracy, detect and identify biases, safeguard data privacy through anonymization, and explain AI results to judicial actors. To this end, first, the datasets will be defined and characterized, and then the different degrees of complexity of case law datasets will be mentioned. Second, the roles and challenges of experts in the design, preservation, and publication of datasets are presented. It is concluded that the adoption of AI in the judiciary is inseparable from the involvement of experts in the construction of jurisprudential datasets.

Key words: Datasets; jurisprudence; challenges; experts.

1. INTRODUCCIÓN

La inteligencia artificial (en adelante, "IA") y los sistemas de procesamiento de grandes volúmenes de datos se han convertido en herramientas presentes en diversos aspectos de la vida cotidiana y en diversos contextos profesionales (Dignum 2017). Esto no es ajeno al ámbito judicial, donde la IA puede ser útil para el tratamiento masivo de datos, la evaluación de perfiles de posibles delincuentes, la asistencia en la toma de decisiones, la aceleración de causas ejecutivas e incluso la agilización de algunas decisiones procesales, entre otras.

Sin embargo, el funcionamiento de una IA depende de la calidad de los conjuntos de datos o *datasets* utilizados para su entrenamiento y especialización. Los conjuntos de datos jurisprudenciales permiten a los sistemas de IA aprender de decisiones pasadas, identificar patrones y presentar los resultados requeridos por un *prompt* adecuado. En ese contexto, este trabajo se centra en la relevancia de los conjuntos de datos y el papel de los expertos en torno a ellos en el marco del sistema judicial.

La construcción y publicación de conjuntos de datos jurisprudenciales forma parte de las iniciativas de gobierno abierto que se están adoptando actualmente en diferentes Estados [(1)]. En este contexto, la elaboración y la publicación abierta de *datasets* jurisprudenciales no es solo una contribución en términos de transparencia institucional, sino también para el análisis académico y científico del derecho, de su interpretación y del quehacer judicial [(2)]. Para ello, el papel de los expertos que colaboran con el sistema

(1) Sobre el concepto de justicia abierta y los parámetros para evaluar su cumplimiento, ver Mercado 2023.

(2) Existen diversos trabajos académicos que tratan las estadísticas tanto de las Cortes Supremas como de otros juzgados en los diferentes Estados. Aquí no se referenciarán en virtud de mantener

judicial en diferentes momentos —especialmente aquellos con formación en derecho, análisis de datos y programación— se torna esencial para asegurar la calidad, la consistencia y la utilidad de los datos.

Un ejemplo de cómo se lleva a cabo la publicación de *datasets* jurisprudenciales lo proporciona el caso de *ItalgiureWeb* [3], gestionado en colaboración entre el Ministerio de Justicia y el Consejo Superior de la Magistratura de Italia. Este producto ha requerido equipos con conocimientos técnicos informáticos y jurídicos para llevar a cabo tareas como la clasificación temática de las sentencias, la identificación de principios jurisprudenciales, entre otras. Un caso similar se observa en el proyecto *Judilibre* [4] (Francia), cuyo objetivo es mejorar la reutilización y el análisis automático de las resoluciones judiciales. En este contexto, los expertos colaboran en la anonimización para garantizar la protección de los datos personales, la curación de corpus y la validación de metadatos. Por su parte, *DataJud* [5], desarrollado por el Consejo Nacional de Justicia de Brasil, ha promovido diversos *datasets*, políticas de anonimización y tratamiento de datos y una API para facilitar la interoperabilidad. Otro ejemplo es la Corte Constitucional de Colombia que ha implementado criterios de clasificación temática y publicación de metadatos normativos. [6] En Argentina, la Corte Suprema, aunque cuenta con un sistema de publicación de estadísticas y también publica las sentencias en diferentes formatos, aún no pone a disposición *datasets* jurisprudenciales. En Estados Unidos, iniciativas de organizaciones sin fines de lucro, como *CourtListener* [7], muestran la colaboración de académicos/as y expertos/as en la construcción de *datasets* jurisprudenciales.

La utilidad de estos conjuntos de datos como herramientas no depende exclusivamente de su volumen o de su disponibilidad, sino sobre todo de que estén correctamente anotadas, estructuradas y validadas por expertos/as que entiendan no solo de datos, sino también de derecho (Constanzo, Malaret, Vega, Cistoldi, Di Iorio 2018). Conforme con ello, aquí se propone que existe un nuevo papel y nuevos retos para los expertos y las expertas que pueden colaborar con el sistema judicial en el diseño y validación de las

---

la brevedad. No obstante, a título ejemplificativo se pueden mencionar: Hamann, 2019, Ash, Goel, Li, Marangon, and Sun, 2023, Ngige, Ayankoya, Balogun, Onuiri, Agbonkhese, Sanusi, 2023; Boyd, Kim, Schlanger, 2020, Cui, Shen, Wen, 2023, Marković and Gostojić 2020, Weinshall, Epstein 2020, Delen, Zolbanin, Crosby, Wright 2021, Pina-Sánchez, Grech, Brunton-Smith, Sferopoulos 2019, Mahari, Stammbach, Ash, Pentland, 2024.

(3) Corte Suprema di Cassazione. *ItalgiureWeb*. https://www.italgiure.giustizia.it/index_en.asp?lang=en.

(4) Cour de cassation. Judilibre. https://www.courdecassation.fr/acces-rapide-judilibre

(5) Conselho Nacional de Justiça. *DataJud*. https://www.cnj.jus.br/sistemas/datajud/

(6) Ministerio de Tecnologías de la Información y las Comunicaciones (Colombia). Sentencias proferidas por la Corte Constitucional 1992-2025. https://www.datos.gov.co/Justicia-y-Derecho/Sentencias-proferidas-por-la-Corte-Constitucional-/v2k4-2t8s/about_data. Un estudio interesante sobre este tópico y cómo un software puede ser utilizado para el análisis de grandes volúmenes de sentencias puede ser encontrado en Ordoñez, Ordoñez, Ordoñez, and Urbano 2020

(7) The Free Law Project. *CourtListener*. RECAP Archive. https://www.courtlistener.com/recap/

herramientas tecnológicas que pueden intervenir en la toma de decisiones o en la transparencia de ellas, principalmente, en este caso, en el diseño de los *datasets* jurisprudenciales.

Las experiencias antes mencionadas, por lo tanto, muestran la relevancia de centrarse en los dos elementos mencionados para entender cómo entrenar y especializar una IA que pueda relacionarse con los procesos judiciales: los conjuntos de datos jurisprudenciales y los expertos que intervienen en lo que a ellos concierne. Para ello, la estructura de este trabajo responde a los dos elementos interrelacionados mencionados: en primer lugar, se definirán y caracterizarán los conjuntos de datos y, a continuación, se mencionarán los diferentes grados de complejidad de los conjuntos de datos jurisprudenciales. En segundo lugar, se presentan las funciones y los retos de los expertos en el diseño, la conservación y la publicación de los conjuntos de datos.

Este trabajo se basa en una revisión bibliográfica y teórica. De acuerdo con esto, el enfoque del tema pretende presentar ideas y argumentos sobre el papel de los conjuntos de datos para la IA y los/as colaboradores/as expertos/as. La tesis principal de este trabajo es que si la IA es o puede llegar a ser importante para el sistema judicial, entonces, los *datasets* también lo son. Por lo tanto, la contribución de los expertos que saben cómo construir los conjuntos de datos de alta calidad también es relevante para el sistema judicial.

2. DATOS Y *DATASETS*

La IA, además de funcionar mediante algoritmos, necesita datos para alimentar el proceso de entrenamiento, especialización y producción de resultados. Los datos primarios son aquellos que se recogen de primera mano, es decir, directamente de la fuente original y que no han sido previamente procesados o interpretados. Los datos procesados, en cambio, son aquellos sobre los que se ha realizado algún tipo de tratamiento (digitalización, anonimización, anotación, etc.). Así, un conjunto de datos es cualquier grupo de datos recogidos, observados o creados con el fin de generar resultados originales (Grupo de Trabajo de "Depósito y Gestión de datos en Acceso Abierto" del proyecto 2012). La relevancia de estos *datasets* radica en que, una vez explorados y procesados, pueden permitir validar un descubrimiento y establecer relaciones entre propiedades sobre la base de algo más que especulaciones, así como cuantificar el peso de estas relaciones (Marradi 2007). Este proceso puede incluso permitir elaborar modelos predictivos dentro de un contexto determinado. Así pues, cualquier dato es una entrada potencialmente relevante para determinados fines. Para ello, los datos deben poder rastrearse, conservarse y, sobre todo, reutilizarse. Además, hay que garantizar su calidad e integridad.

Los *datasets* deben ser precisos, exactos y descriptivos; deben contener las variables y la información necesarias para el análisis y deben estar actualizadas; deben ser consistentes, es decir, coherentes y veraces. Deben ser representativos o completaos, es decir, deben tener un volumen suficiente, en el caso de una muestra, o abarcar todo el universo. Deben ser legibles, es decir, deben estar estructuradas de forma comprensible. Deben respetar el anonimato y la privacidad en el caso de datos sensibles. Por último, deben ser accesibles, es decir, estar en un formato disponible para su reutilización (Scott 2016).

Los conjuntos de datos pueden presentarse en distintos formatos según su naturaleza y los fines para los que estén pensados. Algunos que en función de los datos que podría aportar el Poder Judicial parecen más apropiados, son: CSV (*Comma Separated Values*), JSON (*JavaScript Object Notation*), XML (*Extensible Markup Language*), SQL (*Structured Query Language*) y TXT (*Plain Text*). [8]

Una de las principales ventajas de publicar *datasets* en formatos adecuados es la interoperabilidad que permite a distintas aplicaciones compartir, acceder y utilizar datos independientemente de sus plataformas o formatos. Entre las ventajas podemos mencionar que permite que distintas instituciones e incluso particulares utilicen los datos de forma más dinámica. Esto conduce a formas más eficientes de intercambiar datos, reducir errores y contradicciones y, por supuesto, optimizar tiempo y recursos en la producción de información (Cubides 2018).

La interoperabilidad se puede lograr fácilmente a través de una API. Se trata de una interfaz de programación de aplicaciones que es un conjunto de protocolos que permite que diferentes aplicaciones se comuniquen entre sí. Las API funcionan como un punto intermedio para que un software acceda a los datos de otro. Es decir, solicita, envía o recibe datos de otro sistema de manera estructurada (Rosas Toro, Herrera, Velasco Salinas, Suzette, Martínez Chávez 2023; Plaza Estévez, Ramírez Lamela, Acosta Morales 2016). En este caso, un algoritmo dotado de inteligencia artificial podría conectarse a través de una API para obtener datos directamente de los *datasets* para los fines para los que está programado y para los que están orientados los respectivos *datasets* jurisprudenciales.

Además, la reutilización de datos primarios es relevante en términos de construir conocimiento colaborativo, evitar la duplicación de datos o la contradicción entre ellos. Esto conduce a un uso eficiente de los recursos ya disponibles. Si quien tiene los datos es una institución pública, las razones para hacer públicos los datos son aún mayores en clave de las iniciativas de

(8) Aunque no son formatos propios de bases de datos, es posible encontrar, en ocasiones, datos en XLS/XLSX o DOC/DOCX (Open Knowledge International, s/d).

gobierno abierto y transparencia. Así, por ejemplo, si los datos están disponibles, los investigadores, académicos, periodistas, abogados litigantes, legisladores, etc., pueden tomar la información abierta y pública y concentrarse en el análisis sin tener que repetir el proceso de recopilación.

A pesar de que hasta ahora se ha hecho referencia solo a algunos aspectos técnicos de los *datasets* en general, cabe considerar que, al contextualizarlos en el ámbito judicial, el papel de los expertos se centra en la validez técnica y jurídica de los procesos implicados en la elaboración y reutilización de los conjuntos de datos jurisprudenciales. Los expertos especializados en informática, análisis de datos o documentación digital están en condiciones de garantizar que los formatos utilizados por los *datasets* jurisprudenciales responden a criterios de calidad y legibilidad técnica, para facilitar así su correcta utilización tanto en contextos judiciales como en la investigación aplicada. En este sentido, su intervención no solo fortalece el proceso de toma de decisiones asistido por IA, como se mencionará más adelante, sino que también contribuye a reforzar el carácter público del derecho a través de datos fiables, reutilizables y abiertos.

3. LOS *DATASETS* JURISPRUDENCIALES COMO BASE DE LA INTELIGENCIA ARTIFICIAL

Como se ha mencionado anteriormente, los *datasets* son los insumos para entrenar la IA. Por lo tanto, cualquier sistema de IA que se utilice o pueda utilizarse en diferentes instancias del ámbito jurisdiccional debe basarse en *datasets* para funcionar. Los conjuntos de datos jurisprudenciales —que son conjuntos de resoluciones judiciales— son la materia prima con la que se pueden entrenar y especializar sistemas de IA enfocados en el procesamiento de decisiones judiciales. Estos *datasets* constituyen el input y el contexto en el que los algoritmos "aprenden" a replicar el proceso de razonamiento jurídico. Por ello, la calidad, estructura e integridad de estos *datasets* son preocupaciones relevantes para la toma de decisiones jurídicas asistida por IA.

Lo que aquí denominamos *datasets* o "conjuntos de datos jurisprudenciales" son colecciones estructuradas de decisiones judiciales, principalmente sentencias y resoluciones dictadas por jueces y juezas, que se recopilan, procesan y formatean para su uso en la investigación jurídica, el análisis de datos y las aplicaciones de inteligencia artificial. En otras palabras, son conjuntos de datos jurisprudenciales de grandes volúmenes. A través de ellos, los algoritmos aprenden cómo se estructura el razonamiento jurídico, qué factores se correlacionan con resultados específicos y qué normas doctrinales o procesales se aplicarían en diversos contextos jurídicos (Santosh, Haddad, Grabmair, 2024, Mielnik y Altszyler 2023).

Los conjuntos de datos jurisprudenciales pueden variar en cuanto a formato, especificidad y finalidad en función del uso al que se destinen y del marco institucional en el que se produzcan. A continuación, se presentan tres tipos de *datasets* en función de su complejidad y funcionalidad:

- *Datasets* jurisprudenciales sin procesar o con un procesamiento básico: se trata de colecciones de resoluciones judiciales sin procesar o mínimamente procesadas y se refieren a archivos a menudo en su formato original (.doc o formato similar o texto sin formato). Estos corpus pueden obtenerse de sitios web de tribunales, bases de información jurídicas o archivos institucionales. Aunque son útiles como sistematización, a menudo no son suficientes para las aplicaciones de IA sin una estructuración adicional. Se caracterizan por la falta de metadatos para la recolección automática (por ejemplo, número de caso o expediente, materia jurídica, jerarquía judicial, año de emisión, etc.), estandarización limitada en el formato y las citas, entre otros.

- *Datasets* anotados: se trata de conjuntos de documentos acompañados de metadatos y etiquetas añadidos manual o semiautomáticamente que enriquecen el texto bruto de las resoluciones judiciales. Estas anotaciones pueden indicar el tipo de asunto jurídico (por ejemplo, derecho contractual, derecho penal, responsabilidad civil, derecho administrativo), la instancia judicial (por ejemplo, primera instancia, apelación, casación), el papel de las partes, el resultado final (por ejemplo, declaración de responsabilidad, condena, confirmación de la instancia anterior, revocación, desestimación de una demanda) y las normas jurídicas o precedentes citados.

La anotación es esencial para entrenar modelos de aprendizaje, que se basan en ejemplos etiquetados para "aprender" patrones y asociaciones. Sin embargo, la anotación requiere un trabajo laborioso y preciso, y la calidad del *dataset* depende normalmente de los conocimientos jurídicos y técnicos de los anotadores de datos o de quienes los seleccionan.

- *Datasets* jurisprudenciales estructurados o altamente etiquetados: representan una fase avanzada en la organización y anotación de la información contenida en las resoluciones judiciales. A diferencia de los textos jurídicos brutos o mínimamente procesados mencionados anteriormente, estos *datasets* pretenden ser analíticos de las resoluciones judiciales a partir de su descomposición en componentes menores y semánticamente significativos. Esta estructuración permite a los sistemas de IA no solo leer, sino también "aprender" la lógica interna del razonamiento jurídico.

Normalmente, los conjuntos de datos estructurados segmentan las resoluciones judiciales en elementos como: resumen de los hechos, normativa aplicada, argumentos de las partes, jurisprudencia citada*, ratio decidendi, obiter dicta*, entre otros. Estas categorías pueden identificarse mediante di-

versas metodologías, aunque se destacan principalmente dos. En primer lugar, las técnicas de extracción automatizada, basadas en el procesamiento del lenguaje natural (PLN) permiten el análisis sintáctico a gran escala mediante el entrenamiento de algoritmos para reconocer patrones lingüísticos en los textos jurídicos (Ruiz 2021 [9]). Estos métodos, aunque útiles para agilizar las tareas, pierden precisión al enfrentarse a la complejidad retórica, las formulaciones ambiguas o los argumentos implícitos, entre otros. En segundo lugar, la estructuración manual, llevada a cabo por expertos y expertas (ya sean del área del derecho o de otras áreas de la ingeniería lingüística o las humanidades digitales) garantiza mayores niveles de precisión y fidelidad semántica, pero consume muchos recursos y tiempo, ya que requiere infraestructura humana y tecnológica. De ahí que sea deseable combinar la escalabilidad de la PNL con la precisión específica de la supervisión experta como forma de curaduría de los resultados.

Dado que la calidad de los conjuntos de datos jurisprudenciales condiciona la legitimidad de los sistemas de IA aplicados al ámbito judicial, los expertos y las expertas —en particular los/as que tienen formación en derecho, informática, análisis de datos, lingüística computacional— pueden desempeñar un papel fundamental en la validación técnica y semántica de estos *datasets*. Su intervención tanto en la estructuración y anotación de los datos como en la evaluación de los resultados generados por los modelos de IA garantiza que estos respeten los principios jurídicos aplicables y reflejen adecuadamente los patrones argumentativos del derecho.

4. ACTORES INSTITUCIONALES IMPLICADOS EN LA CONSTRUCCIÓN Y SUPERVISIÓN DE LOS *DATASETS* JURISPRUDENCIALES

La creación, publicación y control de calidad de los conjuntos de datos jurisprudenciales es un esfuerzo institucional e interdisciplinar que implica a una pluralidad de actores. En primer lugar, el Poder Judicial, a través de los jueces, es la fuente de datos, ya que el corpus está compuesto por resoluciones judiciales. Es a partir de los datos que proporcionan los propios jueces y juezas que se debe trabajar en la descripción de los metadatos, la estructuración y el marcado de las secciones que componen las decisiones y el proceso de anonimización (cuando corresponda).

En segundo lugar, los y las profesionales de la informática suelen ser responsables de la infraestructura técnica necesaria para almacenar, procesar y publicar los *datasets*. Esto incluye diseñar plataformas, desarrollar

(9) En este estudio el autor se enfoca en el diseño de un modelo que puede detectar similitudes entre casos.

formatos estructurados, integrar conjuntos de datos y garantizar la interoperabilidad entre sistemas.

En tercer lugar, otro grupo esencial de profesionales es el de los/as científicos/as del derecho que aportan conocimientos metodológicos y doctrinales para garantizar que la estructuración de los datos se ajuste a la lógica del derecho y a los posibles usos previstos, además de desarrollar estudios a partir de los *datasets*. Así, su papel se centra en la definición de categorías, etiquetas, campos relevantes y formas de segmentar los textos jurídicos.

Además, la participación de científicos/as de datos y expertos/as en procesamiento del lenguaje natural (PLN) puede ser muy útil, como ya se ha mencionado. Estos profesionales se encargan de desarrollar algoritmos capaces de extraer, clasificar y organizar la información contenida en las resoluciones judiciales. Transforman grandes volúmenes de texto en modelos de datos estructurados y reutilizables y en validar los métodos para obtener resultados fiables.

En este proceso, el papel de los expertos técnicos consiste en verificar si los datos cumplen las normas técnicas y reglamentarias, si existen sesgos o distorsiones significativas y si el uso de estos *datasets* en sistemas de IA podría afectar negativamente los derechos individuales y colectivos consagrados en la normativa. En muchos casos, también se les puede pedir que expliquen el funcionamiento de los modelos a la judicatura, lo que refuerza su papel de mediadores entre el derecho y la tecnología (Scherer 2019).

Por último, otros actores como los comités de ética, las autoridades de protección de datos, las universidades y las organizaciones de la sociedad civil son relevantes en la supervisión del proceso. Estas entidades pueden actuar para velar por la transparencia, pueden producir información a partir de los datos y pueden emitir recomendaciones para evitar tratos discriminatorios, lesivos, entre otros.

5. Relevancia del trabajo de los/as expertos/as para los conjuntos de datos jurisdiccionales

Los expertos y las expertas —especialmente aquellos con competencias interdisciplinares en derecho, ciencia de datos, programación y ética— desempeñan una serie de funciones que garantizan la integración de los sistemas de IA en el razonamiento jurídico. En cuanto a las instancias en las que la intervención de expertos/as es relevante, podemos mencionar las siguientes: [10]

(10) Esta sección se basa en los siguientes trabajos; Fürst, El-Assady, Keim, and Fischer, 2024, Ovádek, Schroeder, Zglinski, 2024, Wang, Zhang, Hoo, Shao, and Zhang 2024, Habernal, Faber, Recchia, Bretthauer, Sebastian Bretthauer, Spiecker, and Burchard 2024.

- Intervención en la anotación: la identificación y clasificación de elementos como *ratio decidendi, obiter dicta* o argumentación jurídica, citas de jurisprudencia o textos normativos presupone un conocimiento experto del derecho, del contexto procesal y de cuestiones lingüísticas. Aunque los sistemas PLN pueden detectar patrones textuales o rasgos estadísticamente destacados, a menudo carecen de la competencia interpretativa necesaria para distinguir, por ejemplo, entre una razón aducida para decidir un litigio y un argumento complementario o supletorio. Los expertos y las expertas, con conocimientos tanto jurídicos como técnicos, actúan como mediadores para determinar cómo se operativizan las categorías jurídicas abstractas en los conjuntos de datos.

Además, las personas expertas suelen ser los garantes de la coherencia conceptual en los esfuerzos de anotación a gran escala. En ausencia de un protocolo interpretativo, los proyectos de anotación corren el riesgo de introducir problemas semánticos, clasificaciones contradictorias o etiquetas jurídicamente inadecuadas. La participación de expertos puede mitigar estos riesgos.

Otra función clave de los expertos y las expertas es la evaluación de los resultados generados por los sistemas de IA. Cuando los modelos de PLN generan automáticamente *datasets* estructurados, a menudo se encarga a expertos/as que lleven a cabo un control de calidad mediante muestreo, validación y auditorías (Pezzini 2024). Estos procedimientos permiten detectar resultados y clasificaciones erróneas u omisiones relevantes. De este modo, los y las expertos/as son fundamentales para dotar de calidad a los resultados del uso de herramientas automatizadas en entornos judiciales.

- Auditoría y validación de *datasets*: este proceso implica revisar el origen de los datos, los supuestos metodológicos para su construcción y la coherencia interna de su estructura. Estas auditorías tienen por objeto evaluar no solo la integridad técnica del *dataset*, sino también su exactitud jurídica para garantizar que los datos reflejan realmente las categorías jurídicas.

- Detección de sesgos: dado el riesgo de que los conjuntos de datos incorporen o perpetúen sesgos —como la sobrerrepresentación de determinadas decisiones, resultados procesales o perfiles demográficos—, los/as expertos/as desempeñan un papel clave en la identificación de estas distorsiones. El análisis permite determinar si un sistema de IA puede reforzar patrones discriminatorios o sesgar la toma de decisiones de forma que comprometa la justicia y la legalidad.

- Intervención en la explicabilidad de los resultados de la IA: a medida que los sistemas de IA se utilizan cada vez más para apoyar a los actores judiciales —ya sea a través de análisis predictivos, apoyo a la toma de decisiones o clasificación de casos—, los expertos pueden explicar cómo dichos

sistemas producen resultados específicos. Esta función es relevante cuando un juez o una jueza se basa en herramientas de IA para tomar decisiones o para inspirarse en relación con determinadas vías de toma de decisiones y debe comprender, y posiblemente justificar, sus decisiones.

- Anonimización: cuando los conjuntos de datos incluyen información personal o sensible —especialmente en casos de familia o penales— los expertos evalúan la solidez de los protocolos de anonimización y el cumplimiento de estándares éticos y legales de protección de datos.

6. Retos para los expertos en la construcción de *datasets* (11)

La construcción de conjuntos de datos no es una tarea sencilla. La intervención de expertos especializados —que combinan conocimientos jurídicos, lingüísticos y tecnológicos, entre otros— es esencial en todas las fases del proceso, desde la recolección y anotación de los datos hasta su validación y auditoría, como se ha mencionado en la sección anterior. El trabajo de estos expertos influye directamente en la calidad, representatividad y ética de los modelos de IA generados a partir de dichos datos (Hildebrandt 2017; 2019).

En primer lugar, la normalización entre jurisdicciones representa uno de los retos más complejos. La diversidad y heterogeneidad en la redacción y estructuración de las resoluciones judiciales, tanto entre países como dentro de los sistemas nacionales (por ej., entre las provincias), dificultan la construcción de *datasets* interoperables. En este sentido, los expertos desempeñan un papel relevante en el diseño de pautas de anotación que respeten la diversidad normativa y de estructura sin perder coherencia y funcionalidad. Por ejemplo, un experto puede crear un protocolo de clasificación las sentencias penales de diferentes provincias, regiones o Estados de forma homogénea.

En segundo lugar, la anonimización y protección de datos personales requiere un alto grado de precisión y rigor, ya que las resoluciones contienen información muy sensible: nombres de partes, víctimas, testigos, menores, así como datos relacionados con la salud, el empleo, cuestiones financieras o antecedentes penales, entre otros. Los juristas deben garantizar que la anonimización cumple con marcos éticos y legales estrictos sobre protección de datos sensibles y la privacidad. Esta tarea es especialmente compleja cuando se trabaja con datos no estructurados como, por ejemplo, largos

(11) Los desafíos que aquí se incluyen se inspiran en los siguientes trabajos: Ferreira, Gromova, 2023, Eguiluz Castañeira 2020, Buchholtz, 2020, Ugwudike, 2021, Di Iorio, Cistoldi, Constanzo, Giaccaglia, Giordano Lerena, Greco, Iturriaga, Lamperti, Núñez, Malaret, Mollo, Podestá, Trigo, and Vega, 2019.

textos de sentencias en los que la información personal puede estar dispersa o implícita. Un experto o una experta puede desarrollar protocolos específicos para detectar y anonimizar este tipo de información, minimizando el riesgo de violación de la privacidad (Hidebrandt 2017; 2019), como en los casos de violencia de género o cuando hay menores implicados, en los que la protección de la identidad es vital.

Otro reto está relacionado con la exhaustividad y representatividad de los datos. Los *datasets* pueden carecer de cobertura territorial o judicial o excluir algún tipo de casos (por ejemplo, los casos de ejecución de pequeñas cuantías o títulos valores como un pagaré). Esto crea un sesgo en la información disponible que puede limitar la eficacia e imparcialidad de los resultaos de los sistemas de IA. Los expertos deben identificar estas lagunas y trabajar para incluir una gama amplia y diversa de fuentes, con el fin de garantizar que el conjunto de datos refleje la realidad jurídica.

Además, la presencia de sesgos históricos en los datos es una cuestión que requiere la atención de quienes son expertos/as. Las sentencias judiciales no son meros textos neutrales, sino que están inmersas en trasfondos sociales, culturales y temporales específicos que pueden incluir razonamientos no adecuados para otros contextos. La tarea del experto de consiste en detectar estos sesgos y recomendar estrategias correctoras. Por ejemplo, en el entrenamiento de modelos predictivos para casos penales, un experto podría identificar patrones que reflejen discriminación racial o socioeconómica y sugerir mecanismos para evitar que la IA perpetúe el sesgo.

En cuanto al formato, la accesibilidad y la legibilidad, los expertos técnicos y jurídicos colaboran para superar la barrera de los documentos en formatos no legibles por máquina (por ejemplo, PDF escaneados). Mediante el control de tecnologías como el reconocimiento óptico de caracteres (OCR), los expertos garantizan que la digitalización preserve la integridad de los datos y minimice los errores. Además, su participación es clave para diseñar e implementar API que faciliten el acceso estructurado y masivo a los datos judiciales al promover no solo la construcción eficiente de *datasets*, sino también la democratización del conocimiento jurídico. Por ejemplo, un experto en programación puede desarrollar una API para el acceso en tiempo real a conjuntos de datos judiciales, facilitando a investigadores, académicos y desarrolladores la integración de esta información en herramientas de análisis y sistemas de IA.

Por último, la función colaborativa y ética de los expertos es relevante y desafiante. Su papel, como señala Dignum (2019) contribuye a hacer de la construcción de estos *datasets* jurisprudenciales un proceso transparente, alineado con los valores del Estado de derecho.

7. Conclusión

El desarrollo y la utilización de la inteligencia artificial en el ámbito judicial dependen de la disponibilidad, la calidad y la estructura de los conjuntos de datos jurisprudenciales que alimentan los sistemas de IA. Estos *datasets* no son meros insumos técnicos, sino que condensan prácticas institucionales, opciones metodológicas y criterios jurídicos que deben ser cuidadosamente diseñados, auditados y validados para garantizar la legitimidad de los resultados que producen.

Este trabajo muestra que los conjuntos de datos jurisprudenciales presentan diferentes grados de complejidad. Para alcanzar niveles de calidad y utilidad, la intervención de expertos es esencial. Lejos de ocupar un papel auxiliar, los expertos son actores relevantes en los procesos de construcción, supervisión y validación de los conjuntos de datos. Los expertos intervienen en tareas fundamentales como la anotación y curaduría del corpus jurisprudencial, la definición de metadatos jurídicamente relevantes, la auditoría de los resultados generados por los modelos, la validación de protocolos de anonimización, la evaluación de sesgos y la explicación técnica a jueces o autoridades.

Por último, se ha argumentado que los expertos no solo garantizan la viabilidad técnica de los conjuntos de datos, sino que también contribuyen a su legitimidad institucional y científica. Al intervenir en el diseño y evaluación de los conjuntos de datos, su papel está directamente vinculado a la calidad de las decisiones jurídicas que tales sistemas producen o en las que en ellos se basan, además de la de la información que se produce a partir de los datos.

En conclusión, la incorporación de la inteligencia artificial al ámbito judicial no puede entenderse sin tener en cuenta el papel de los expertos frente a los *datasets* y el de la calidad y disponibilidad de los respectivos conjuntos de datos jurisprudenciales.

Referencias bibliográficas

Ash, E., Goel, N., Li, N., Marangon, C., & Sun, P. (2023). Curated large dataset of criminal cases from Wisconsin circuit courts. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, & S. Levine (Eds.), *Advances in neural information processing systems* (pp. 12626–12643). Curran Associates, Inc.

Asociación para el Gobierno Abierto, Ministère de Justice. (s.d.). *Open data of court decisions*. https://www.justice.gouv.fr/documentation/open-data-court-decisions

Boyd, C., Kim, P., & Schlanger, M. (2020). Mapping the iceberg: El impacto de las fuentes de datos en el estudio de los tribunales de distrito. *Journal of Empirical Legal Studies, 17*(3), 466–492.

Buchholtz, G. (2020). Inteligencia artificial y tecnología jurídica: Challenges to the rule of law. In T. Wischmeyer & T. Rademacher (Eds.), *Regulating artificial intelligence* (pp. 175–198). Springer. https://doi.org/10.1007/978-3-030-32361-5_8

Comisión Europea para la Eficacia de la Justicia. (2018). *Carta ética europea sobre el uso de la inteligencia artificial en los sistemas judiciales*. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c

Conselho Nacional de Justiça. (s.d.). *DataJud*. https://www.cnj.jus.br/sistemas/datajud/

Constanzo, B., Malaret, P., Vega, P., Cistoldi, P., & Di Iorio, A. H. (2018). Arquitectura y organización de laboratorios de informática forense. *Red Congreso Iberoamericano de Docentes e Investigadores en Derecho e Informática*.

Corte Suprema di Cassazione. (s.d.). *ItalgiureWeb*. https://www.italgiure.giustizia.it/index_en.asp?lang=en

Cubides, J., Herrera, F., Gaona García, P. A., & Sánchez Alonso, S. (2018). Linked data: Qué sucede con la heterogeneidad y la interoperabilidad. *Scientia et Technica, 23*(2), 230–240.

Cui, J., Shen, X., & Wen, S. (2023). A survey on legal judgment prediction: Datasets, metrics, models and challenges. *IEEE Access, 11*, 102050–102071.

Delen, D., Zolbanin, H., Crosby, D., & Wright, D. (2021). Encarcelar o no encarcelar: Un modelo analítico para los tribunales de drogas. *Anales de Investigación Operativa, 303*(1–2), 101–124.

Di Iorio, A. H., Cistoldi, P., Constanzo, B., Giaccaglia, M. F., Giordano Lerena, R., Greco, F., Iturriaga, J., Lamperti, S., Núñez, L., Malaret, P., Mollo, M., Podestá, A., Trigo, S., & Vega, P. (2019). *Guía técnica para el diseño, implementación y gestión de laboratorios de informática forense*. Universidad FASTA.

Dignum, V. (2017). Inteligencia artificial responsable: Diseñando IA para los valores humanos. *Revista de la UIT: Descubrimientos TIC, número especial*, (1). https://www.itu.int/dms_pub/itu-s/opb/journal/S-JOURNAL-ICTF.VOL1-2018-1-P01-PDF-E.pdf

Dignum, V. (2019). *Inteligencia artificial responsable. Inteligencia artificial: Fundamentos, teoría y algoritmos*. Springer. https://doi.org/10.1007/978-3-030-30371-6

Eguiluz Castañeira, J. A. (2020). Desafíos y retos que plantean las decisiones automatizadas y los perfilados para los derechos fundamentales. *Estudios de Deusto: Revista de Derecho Público, 68*(2), 325–367.

Ferreira, D., & Gromova, E. (2023). Jurisprudencia hiperrealista: La era digital y la (in)certeza del análisis del juez. *International Journal of Semiotics of Law, 36*, 2261–2281. https://doi.org/10.1007/s11196-023-10015-0

Fürst, D., El-Assady, M., Keim, D., & Fischer, M. (2024). Challenges and opportunities for visual analytics in jurisprudence. *arXiv preprint arXiv:2412.06543*. https://doi.org/10.48550/arXiv.2412.06543

Grupo de Trabajo de "Depósito y Gestión de datos en Acceso Abierto" del proyecto RECOLECTA. (2012). *La conservación y reutilización de los datos científicos en España: Informe del Grupo de Trabajo de Buenas Prácticas*. Fundación Española para la Ciencia y la Tecnología (FECYT).

Habernal, I., Faber, D., Recchia, N., Bretthauer, S., Gurevych, I., Spiecker, I., & Burchard, C. (2024). Mining legal arguments in court decisions. *Artificial Intelligence and Law, 32*, 1–38. https://doi.org/10.1007/s10506-023-09361-y

Hamann, H. (2019). The German federal courts dataset 1950–2019: From paper archives to linked open data. *Journal of Empirical Legal Studies, 16*, 671–688.

Hildebrandt, M. (2017). Regulación algorítmica y Estado de derecho. *Philosophical Transactions of the Royal Society A, 376*, 20170355. https://doi.org/10.1098/rsta.2017.0355

Hildebrandt, M. (2019). La privacidad como protección del yo incomputable: Del aprendizaje automático agnóstico al agonístico. *Theoretical Inquiries in Law, 20*(1), 83–121. https://doi.org/10.1515/til-2019-0004

Mahari, R., Stammbach, D., Ash, E., & Pentland, A. (2024). LePaRD: A large-scale dataset of judicial citations to precedent. In L.-W. Ku, A. Martins, & V. Srikumar (Eds.), *Proceedings of the 62nd annual meeting of the Association for Computational Linguistics* (Vol. 1, pp. 9863–9877). Association for Computational Linguistics.

Marković, M., & Gostojić, S. (2020). Datos judiciales abiertos: Un análisis comparativo. *Social Science Computer Review, 38*(3), 295–314.

Marradi, A., Archenti, N., & Piovani, J. I. (2007). *Metodología de las ciencias sociales*. Emecé.

Mercado, J. G. (2023). Justicia abierta y participación pública para un enfoque de justicia centrada en las personas. *Reflexión Política, 25*(52), 40–48.

Mielnik, D., & Altszyler, E. (2023). Inteligencia artificial aplicada al estudio del derecho: Análisis computacional de la jurisprudencia de casación penal. In J. Corvalán (Ed.), *Tratado de inteligencia artificial y derecho* (2nd ed.). La Ley.

Ministerio de Tecnologías de la Información y las Comunicaciones (Colombia). (s.d.). *Sentencias proferidas por la Corte Constitucional 1992–2025*. https://www.datos.gov.co/Justicia-y-Derecho/Sentencias-proferidas-por-la-Corte-Constitucional-/v2k4-2t8s/about_data

Ngige, O. C., Ayankoya, F. Y., Balogun, J. A., Onuiri, E., Agbonkhese, C., & Sanusi, F. A. (2023). A dataset for predicting Supreme Court judgments in Nigeria. *Data in Brief, 50*. https://doi.org/10.17632/ky6zfyf669.1

Open Knowledge International. (s.d.). Formato de archivos. *Open Data Handbook*. https://opendatahandbook.org/guide/es/appendices/file-formats/

Ordoñez, H. A., Ordoñez, C., Ordoñez, J. A., & Urbano, F. (2020). Búsqueda de jurisprudencia en Colombia basada en procesamiento de lenguaje natural (PLN) y linked data. *Inge CuC, 16*(2), 277–284.

Ovádek, M., Schroeder, P., & Zglinski, J. (2024). Where law meets data: A practical guide to expert coding in legal research. *European Law Open, 3*(4), 820–848. https://doi.org/10.1017/elo.2024.23

Pezzini, M. C. (2024). *Inteligencia artificial explicable: Análisis de metodologías y aplicaciones* (Trabajo final de la Especialización en Ingeniería de Software). Facultad de Ingeniería, Universidad Nacional de La Plata. http://sedici.unlp.edu.ar/handle/10915/174328

Pina-Sánchez, J., Grech, D., Brunton-Smith, I., & Sferopoulos, D. (2019). Explorando el origen de las disparidades de sentencias en la Corte de la Corona: Usando técnicas de minería de textos para diferenciar entre disparidades entre el tribunal y el juez. *Social Science Research, 84*. https://doi.org/10.1016/j.ssresearch.2019.102343

Plaza Estévez, S., Ramírez Lamela, N., & Acosta Morales, C. (2016). *API de servicios web orientados a accesibilidad* (Tesis de Ingeniería Informática). Universidad Complutense de Madrid.

Rosas Toro, M. A., Herrera, H., Velasco Salinas, J., Leslie Suzette, & Martínez Chávez, J. M. (2023). Construyendo una web API con la arquitectura N capas. *Multidisciplinas de la Ingeniería, 11*(18), 14–25.

Ruiz, R. A. (2021). *Búsqueda de jurisprudencia basada en la similitud de hechos utilizando técnicas de PLN y ML* (Tesis doctoral). Universidade de São Paulo. https://doi.org/10.11606/D.3.2021.tde-14022022-122906

Santosh, T. Y. S., Haddad, R. G., & Grabmair, M. (2024). TEDH-PCR: A dataset for precedent understanding and prior case retrieval in the European Court of Human Rights. *arXiv:2404.00596*. https://doi.org/10.48550/arXiv.2404.00596

Scherer, M. (2019). Inteligencia artificial y toma de decisiones jurídicas: ¿Abierto de par en par? Estudio sobre el ejemplo del arbitraje internacional. *Revista de Arbitraje Internacional, 36*(5), 539–573. https://doi.org/10.54648/joia2019028

Scott, M., Stark, I., Byatt, D., & Cox, S. (2016). *Presentación de los datos de investigación*. Universidad de Stanford.

The Free Law Project. (s.d.). *CourtListener. RECAP archive*. https://www.courtlistener.com/recap/

Ugwudike, P. (2021). Data-driven technologies in justice systems: Intersections of power, data configurations, and knowledge production. In A. Lavorgna & T. Holt (Eds.), *Researching cybercrimes*. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-74837-1_5

Wang, X., Zhang, X., Hoo, V., Shao, Z., & Zhang, X. (2024). LegalReasoner: A multi-stage framework for legal judgment prediction via large language models and knowledge integration. *IEEE Access, 12*, 166843–166854. https://doi.org/10.1109/ACCESS.2024.3496666

Weinshall, K., & Epstein, L. (2020). Developing high-quality data infrastructure for legal analytics: Introducing the Israeli Supreme Court database. *Journal of Empirical Legal Studies, 17*(2), 416–434.

Los usos de la IAGen como objeto de casos judiciales [*]

Por Débora Marhaba [**] *y Nicolás Pérez Felicioni* [***]

Resumen: El uso cada vez más extendido de la IAGen en las profesiones jurídicas trae aparejado que jueces y juezas ahora también deban tomar intervención y resolver acerca de la legalidad de aquella práctica. A partir del análisis de sentencias recientes, con especial atención a las "alucinaciones" no declaradas por los abogados en procesos judiciales, el presente trabajo examina cómo la IAGen impacta en el análisis judicial y en la responsabilidad profesional, desde la perspectiva de la "Justicia digital" y de las técnicas propias del derecho.

Palabras clave: Tecnología; ejercicio de la abogacía; inteligencia artificial.

Abstract: The increasingly widespread use of AI-generated content (IAGen) in legal professions entails that judges must now also intervene and make decisions regarding the legality of such practices. Based on the analysis of recent rulings, with particular attention to unreported "hallucinations" by lawyers in judicial proceedings, this work examines how IAGen impacts judicial analysis and professional responsibility, from the perspective of "Digital Justice" and the techniques inherent to the field of Law.

Key words: Technology; practice of law; Artificial Intelligence.

1. Planteo

El poder judicial, pero especialmente los jueces y las juezas, tienen por lo menos tres grandes desafíos en lo que respecta a la inteligencia artificial generativa (en adelante IAGen): 1) analizar y sopesar la implementación de esta tecnología en el sistema de justicia; 2) intervenir en casos donde los/

(*) Este trabajo fue elaborado en el contexto del Proyecto Universidad de Buenos Aires Ciencia y Técnica (UBACyT) 20020220100260BA, Programación Científica 2023, línea dedicada al estudio del tema que lleva el título de esta investigación. Contó con la colaboración de las correcciones de la Lic. Daniela Donni, a quien estamos agradecidos.

(**) Abogada (UNLaM). Magister en Magistratura (UBA). Subdirectora del Instituto de Derecho del Usuario y Consumidor del Colegio Público de la Abogacía de la Capital Federal.

(***) Abogado (UBA). Profesor adjunto interino de la materia "Genealogía de la Justicia" (UBA). Maestrando en Cultura Jurídica (Univ. de Girona).

as operadores/as jurídicos/as la utilizan; y 3) resolver casos en los que esta genera controversias, daños, delitos, entre otros.

A partir de las contribuciones doctrinarias de los últimos años, es posible identificar mayoritariamente discursos teóricos y divulgativos sobre los posibles beneficios y los riesgos de la IAGen y, en menor medida, análisis de experiencias e investigaciones empíricas sobre su uso e impacto en los diferentes ámbitos jurídicos. (1)

El hecho de abordar "por casos" los usos de la IAGen en el ámbito jurídico, es decir, sobre la base de acontecimientos que llegan a la órbita judicial, posibilita no solo describir el estado de cosas actuales sino que, de manera más profunda, refleja decisiones institucionales frente a los desarrollos de índole técnico y comercial. El rol de los poderes judiciales en la organización social es el que se ve desafiado con la irrupción de la IAGen, ya sea tanto por parte de los/as distintos/as sujetos participantes del mundo jurídico así como del derecho en general.

En este trabajo exploratorio abordaremos algunas recientes sentencias acerca de las "alucinaciones" de los sistemas de IAGen no detectadas o corregidas por los/as abogados/as al momento de hacer presentaciones judiciales desde una perspectiva de técnica jurídica como la que nos ofrece la propuesta de "Justicia digital".

2. "Justicia digital" como punto de partida

Nos resulta oportuno recordar cuáles son los presupuestos técnico-digitales de los modelos de inteligencia artificial. Para ello, y con fundamento en un trabajo anterior (2) que sirve de puntapié, tomamos como referencia la obra del jurista Antoine Garapon que, en colaboración con el investigador Jean Lassègue, asumieron –de forma precursora– la tarea de analizar las implicancias jurídicas del fenómeno que tituló "Justicia digital". (3)

En los capítulos iniciales del libro los autores señalan, en primer lugar, que para operar la transformación digital de la realidad es necesario "codificar en forma de números acontecimientos del mundo físico" de manera tal que "la realidad externa resulta así codificada en el modo de un 'dato', término generalmente empleado en plural". (4) En segundo lugar, para darle sentido a esos datos, un programa los organiza a través de un algoritmo que

(1) Llorens, María Pilar. (2025). "Apoyo a los sectores de América Latina en la adopción responsable de IA Generativa. Argentina. Informe Nacional". CETyS, UDESA.

(2) Wierzba, Sandra; Cosoy, Diego; Perez Felicioni, Nicolás y Marhaba, Débora (2024). "Justicia digital, inteligencia artificial y profesiones jurídicas". Diario La Ley, 25/09/24.

(3) Garapon, Antoine y Lassègue, Jean. (2024). Justicia Digital. Revolución gráfica y ruptura antropológica. Editorial Jusbaires.

(4) Garapon, Antoine y Lassegue, Jean, *op. cit.*, p. 39.

"adopta la forma de instrucciones de manera de clasificar, contar, descifrar los datos cuya estructura repetitiva es posible aislar". Es en este paso donde se arriba a la posibilidad de automatizar y delegar la ejecución de las instrucciones.

Existe otra cuestión importante –además de la automatización– que es la posibilidad de "agregación estadística de cantidades gigantescas de datos [que] apunta a establecer correlaciones que hasta entonces pasaron inadvertidas que permiten dibujar tendencias en los fenómenos naturales y las interacciones sociales, y al hacerlo, establecer previsiones por lo que respecta al futuro". (5) Esta última parte nos resulta relevante ya que a partir de esta lógica es que podemos decir que, frente a una consulta que le hagamos a un sistema de IAGen, se nos brindará una respuesta a modo de interpretación o sugerencia de cómo deberíamos argumentar o resolver cuestiones jurídicas.

Esta práctica de consulta a sistemas de IAGen consolida lo que Garapon llama "un nuevo derecho natural interno al mundo". (6) Esto es así ya que, a partir de la delegación de las respuestas y a lo que el sistema pueda correlacionar a raíz de la información en forma de datos que viene "estandarizado de fábrica", no solo aceptamos lo que el sistema nos arroja –por falta de un criterio propio– sino que lo tomamos como un referente de evaluación del conocimiento jurídico. Desde este punto de vista, no resulta difícil advertir que los problemas que ya estaban presentes en la cultura del *copy-paste* se profundizan de manera significativa.

El uso generalizado de IAGen, que por un lado redunda en la falta de actualización de conocimientos, no solo impacta en el desarrollo profesional de la abogacía sino que también comienza a permear en la formación de los/as estudiantes de derecho. Respecto a lo primero, es preciso advertir que en los estudios jurídicos, a las primeras tareas asignadas se le suma –generalmente de manera gratuita– el entrenamiento de diferentes sistemas de IAGen, convirtiendo las fuentes del derecho en datos tratables para la creación de agentes "no humanos" que realicen las mismas tareas que los/as recién iniciados/as. En cuanto a los/as estudiantes universitarios/as, se vislumbra un importante incremento en el uso de IAGen en la producción y redacción de escritos en instancias evaluatorias. Esto último instó a las autoridades universitarias a analizar y generar propuestas sobre el futuro de la educación, en especial sobre el modo de evaluar a los/as alumnos/as; incluso se llegó a sugerir establecer la evaluación oral obligatoria en diferentes instancias. De tales situaciones, cabe preguntarse sobre la calidad del aprendizaje –que ya contaba con sus propias limitaciones previas– y el incremento de la precarización en un área que ya mostraba dicha tendencia.

---

(5) Ibídem, p. 43.
(6) Ibídem, p. 103.

3. LOS CASOS ANALIZADOS

En Argentina comienzan a vislumbrarse diferentes resoluciones judiciales donde se advierte que los/as letrados/as y magistrados/as utilizan diferentes aplicaciones de IAGen para munirse de fundamentos jurídicos que avalen sus argumentos.

En el caso "Giacomino, César Adrián y otros c/ Monserrat, Facundo Damián y otros s/ daños y perjuicios", [7] los jueces de la Sala II de la Cámara de Apelaciones en lo Civil y Comercial de Rosario, Santa Fe, evaluaron la conducta de un letrado que para contestar la expresión de agravios de su contraria utilizó citas jurisprudenciales inexistentes extraídas de un programa de IAGen.

Específicamente, al advertir la inexistencia de las citas utilizadas por el letrado, los jueces dictaron una medida para mejor proveer para que las identifique. Frente a ello, el letrado aclaró que *"*... utilizó un sistema de inteligencia artificial y que volcó el resultado encontrado de buena fe, de modo que resulta claro que las citas mencionadas no fueron chequeadas por tal curial".

Luego, consideraron que el accionar del abogado, aun de buena fe, comprometía su responsabilidad profesional no solo con el tribunal ante el que litigaba, sino también con su cliente. Algo interesante para destacar de la sentencia es que se indicó que incluso ante la existencia de un formulario de consentimiento informado entre el letrado y su cliente, sobre la utilización de sistemas de chats conversacionales, ello no releva la obligación del profesional de cotejar de manera fehaciente las fuentes jurídicas que utiliza para fundar la pretensión de su cliente.

Asimismo, resaltaron que en los propios términos y condiciones del tipo de *chatbots* utilizado por el letrado, se destaca que se trata de un "... sistema experimental y que sus resultados deben ser cotejados...", ya que "... los resultados que ofrecen, en pos de satisfacer los requerimientos del consultante, incluyen aun y en muchos casos las denominadas 'alucinaciones' y por ello, como ocurrió en el caso, aparecen fuentes lisa y llanamente inventadas".

Finalmente, más allá de la solución del recurso de apelación en concreto, los jueces decidieron hacer un llamado de atención al letrado en cuestión y, además, a fin de prevenir que este tipo de situaciones se generalice en el fuero, anoticiaron al Colegio de Abogados de la jurisdicción a fin de que adopte las medidas que correspondieren.

---

(7) Cámara de Apelaciones en lo Civil y Comercial de Rosario, Santa Fe, Sala II, "Giacomino, César Adrián y otros c/ Monserrat, Facundo Damián y otros s/ daños y perjuicios", 20/08/2025.

El caso descripto no fue el único que abordó la utilización de herramientas de IAGen en el contexto de procesos judiciales. Es tentador sostener que frente a la proliferación de capacitaciones, artículos de doctrina, disertaciones sobre la aplicación de la IAGen en las profesiones jurídicas, diversos profesionales hasta parecieran querer forzar algunos pronunciamientos para destacar o poner el tema en agenda, cuando en los pocos trabajos de relevamiento en el ámbito nacional hay silencio por parte de las máximas autoridades sobre el uso de esta tecnología y pocos recursos para su aplicación de manera responsable.

Seguidamente señalaremos algunas sentencias donde la utilización de herramientas de IAGen fue irrelevante para la solución del caso en concreto o donde no fue acreditada su utilización por parte de los/as abogados/as intervinientes.

En la primera causa, "A., N. B. y otros c/ Aguas Santafesinas S.A. s/ medida autosatisfactiva", [8] la parte actora reclamó para que se ordene a la demandada la inmediata adecuación y prestación del servicio de agua potable en su domicilio. El juez interviniente, luego de un importante desarrollo sobre el derecho al acceso de agua potable, fundado en normativa nacional e internacional, consideró:

> "A mayor abundamiento, y recurriendo a Meta IA, el botón de la inteligencia artificial de WhatsApp, ante la pregunta '¿Se cumple el derecho al acceso al agua con una presión menos a 1 mca?' la respuesta brindada fue que no, entendiendo que el derecho al acceso al agua implica no solo tener una cantidad suficiente, sino que el agua sea segura, aceptable, asequible y accesible, siendo la presión un factor importante para garantizar el acceso equitativo y seguro al agua".

Si ya la solución se había encontrado por métodos tradicionales y estaba probado lo reclamado, ¿qué sentido tuvo citar en un documento público, como lo es una sentencia, una respuesta de un sistema de IA que no contiene pautas precisas sobre el tema discutido?

Una posible respuesta al interrogante es que este modo de utilizar la IAGen podría deberse a un intento de legitimar lo decidido por el juez y proponer como estándar de verdad la respuesta brindada por dicha herramienta, asimilándola a una cita de autoridad extrajurídica.

Por otro lado, cabría la posibilidad de preguntarse si el magistrado consultó al chat conversacional de manera previa al análisis del caso. Dicha situación implicaría una delegación no tradicional, ya que en los expedientes judiciales –como el analizado– se cuenta con peritos que son los expertos auxiliares del juez en una materia determinada, quienes ante el mal desempeño de sus tareas podrían ser sancionados.

---

(8) Juzgado Civil y Comercial de 14ª Nominación de Rosario (Santa Fe), "A., N. B. y otros c/ Aguas Santafesinas S.A. s/ medida autosatisfactiva", CUIJ 21-02975994-1.

Por ello, entendemos que en el intento de legitimar lo decidido por el juez –en un caso donde la sentencia ya cumplía con el deber de motivación suficiente– a su vez se legitimó la propia herramienta.

En otro caso, la Sala I de la Cámara de Apelaciones en lo Civil y Comercial de Morón, Provincia de Buenos Aires, en ocasión de analizar si un recurso de apelación cumplía con las exigencias procesales de admisibilidad, se advirtió que el abogado apelante citaba casos jurisprudenciales que no pudieron ser hallados. (9)

Específicamente, se indicó que "[a]quí no puedo afirmar, con certeza, que la cita de fallos que no existen (o al menos no están localizables) provenga de un inadecuado uso de herramientas de inteligencia artificial. Puede deberse a múltiples causas".

Sin perjuicio de no tener por acreditado el uso de IAGen, en virtud de las disposiciones referidas a las facultades instructorias y ordenatorias de los magistrados, hicieron algunas valoraciones. Por ejemplo, afirmar que frente al uso de IAGen no se está a favor de la sanción del letrado si utiliza de manera inadecuada dicha herramienta; se puso en conocimiento del letrado que en caso de hacer uso de esta tecnología herramienta (porque nada lo impide) es menester, en todos los casos, revisar, chequear y corroborar que la información incluida en sus planteos responda a fuentes reales. Además, se solicitó la colaboración del Colegio de Abogados para que se difunda entre los/as matriculados/as la importancia de verificar cuidadosamente las citas y referencias empleadas en el caso de utilizarse herramientas de IAGen para preparar presentaciones.

Por último, en un sentido similar, la Cámara de Apelaciones en lo Civil, Comercial, Familia y Minería de General Roca, (10) Río Negro, tras advertir la inexistencia de las citas jurídicas utilizadas para fundar una expresión de agravios, afirmó que era muy probable que los letrados hayan recurrido al uso de la IA. Por ello, y ante la reiteración de este tipo de conductas, llamó a la reflexión a los letrados y remarcó que aún:

> "… actuando de buena fe, compromete la responsabilidad profesional del/a letrado/a no solo ante el tribunal sino, especialmente, respecto de su cliente –existiera o no consentimiento informado que autorice el uso de *chatbots* con inteligencia artificial generativa– pues no puede haber consentimiento válido alguno que releve a un/a letrado/a de su deber de cotejar las fuentes en las que basa sus posiciones jurídicas".

---

(9) Cámara de Apelaciones en lo Civil y Comercial de Morón, Pcia. de Buenos Aires, "Acevedo Gerardo Gabriel c/ Cáceres Mareco Illian Arsenio y Agrosalta Cooperativa de Seguros Limitada s/ daños y perjuicios", MO 19435-2020, 15/09/2025.

(10) Cámara de Apelaciones en lo Civil, Comercial, Familia y Minería de General Roca, "M.J.L. c/ Peugeot Citroen Argentina S.A. y otra s/ sumarísimo", Sentencia 199, 18/09/2025.

En estas dos últimas causas, los jueces identificaron –de modos diferentes entre sí– a la IAGen a pesar de no tener por reconocido su uso en el caso concreto. Uno de los puntos más importantes, a nuestro entender, del litigio de la Cámara de General Roca es que los jueces partieron de una presunción para instituir [11] el uso indebido de sistemas de IAGen. A partir de esta herramienta propiamente jurídica, se resolvió llamar la atención a los letrados; no regular honorarios profesionales en la segunda instancia por considerar inoficiosas sus tareas; dar intervención al Colegio profesional para que investigue un eventual incumplimiento al Código de Ética. Además, otra cuestión saliente en términos jurídicos, es la calificación de la actuación de los letrados que señalaron como actitud culposa, separándola de lo que hubiera sido una actividad de tipo dolo.

Distinto sería, tal como resalta el tribunal, que

> "... si no fuera así y las citas y fallos fueran de su propia autoría, es decir llanamente inventados por los propios abogados en cuanto a las carátulas y contenido, esa conducta resultaría mucho más grave aún (estafa procesal), pues se intentaría inducir a error al Tribunal citando nada menos que precedentes del STJ que, como se sabe, constituyen doctrina legal obligatoria".

En cambio, en la sentencia de la Cámara de Morón, frente a la incertidumbre respecto de uso de IAGen -que tampoco había sido alegada- no se utilizó el mecanismo legal de presunción o una medida para mejor proveer, sino que directamente hicieron valoraciones sobre lo que podría representar un "buen uso" de las herramientas de inteligencia artificial.

En tal aspecto, si quien juzga no pudo tener por reconocido que quien expresó agravios en una causa utilizó un sistema de IA, ¿qué sentido tiene hablar de ello en una sentencia?

Podríamos arriesgar que el valor de la sentencia de la Cámara de General Roca está en que, a partir de una técnica propia y exclusiva del derecho, como lo son las "presunciones", intentó controlar el uso de una técnica ajena al derecho, como lo es la técnica digital.

4. Normas éticas de la abogacía frente a la IA

Tras el análisis de las sentencias abordadas, advertimos que en su mayoría se efectuaron llamados de atención a los letrados y, asimismo, se dio intervención a los colegios profesionales correspondientes. Por ello, abordaremos ahora algunos aspectos relevantes de las normas éticas que rigen el ejercicio de la abogacía vinculados a lo analizado en este trabajo.

---

(11) Thomas, Yan. (1999). *Los artificios de las instituciones. Estudios de derecho romano*, Buenos Aires, EUDEBA-Rojas, p. 9.

Uno de los principales deberes del/de la abogado/a es el de la "probidad". En tal sentido, se ha indicado que dicho concepto:

> "... presenta especiales perfiles por cuanto comporta conductas de variado contenido no solo referidas a la honorabilidad, a la honradez y a la integridad, sino que también abarca el sentido más simple del vocablo referido a la cualidad de las cosas y personas y solo en un sentido figurado se la utiliza para calificar moralmente a estas últimas". (12)

De hecho, como se ha destacado en una de las sentencias analizadas, el deber de probidad:

> "... no importa tan solo corrección desde el punto de vista pecuniario: requiere además lealtad personal, veracidad, buena fe. Así, por ejemplo, no debe [...] formular afirmaciones o negaciones inexactas, efectuar en sus escritos citaciones tendenciosamente incompletas, aproximativas o contrarias a la verdad...".

El servicio de los/las profesionales jurídicos/as se debe a la obligación de custodiar los intereses del/de la cliente utilizando las técnicas jurídicas que se tengan al alcance en la prestación del servicio, con un fuerte apego a las normas éticas.

Vinculado al deber de probidad, encontramos otro que es el de "decir la verdad al/a la cliente", no crearle falsas expectativas, ni magnificar las dificultades, o garantizar el buen resultado de su gestión profesional y atender los intereses confiados con celo, saber y dedicación. Específicamente, y en lo que refiere a los casos elegidos para este trabajo, el artículo 24 del Código de Ética del Colegio Público de la Abogacía de la Capital Federal dispone: "Falsedad de citas: Es falta ética efectuar citas doctrinarias o jurisprudenciales inexistentes, o exponerlas en forma tal que falseen la opinión o el fallo invocados, o realizar falsas transcripciones de resoluciones judiciales o escritos del contrario". (13)

No caben dudas de que muchas de las faltas éticas dentro del ejercicio profesional implican reproches jurídicos. En tal sentido, fundar un escrito judicial en citas falsas que son el resultado de una herramienta de IA y que no han sido verificadas por el/la letrado/a, como se constató en el caso de la Cámara de Apelaciones de Rosario, o como presumió la Cámara de Apelaciones de General Roca, darían lugar a diferentes tipos de reproches vinculados entre sí, como por ejemplo: a) la declaración de deserción del recurso de apelación por no cumplir con los requisitos dispuestos en el Código Procesal respecto a la crítica concreta y razonada de la sentencia de

---

(12) Tribunal de Ética Forense, "G. V., A.", 03/03/1983, LA LEY 1984-A, 314 - LLP 1984, 457, Cita: TR LALEY AR/JUR/1624/1983.

(13) Colegio Público de la Abogacía de la Capital Federal. (2023). *Normas para el ejercicio profesional*, Buenos Aires, Editorial Jusbaires, p. 40.

primera instancia; b) el eventual análisis de la actitud del/de la abogado/a por parte del Comité de Ética del Colegio Profesional por no cumplir los deberes fijados al respecto; c) dependiendo de las circunstancias del caso, el/la cliente/a que se vio perjudicado/a por la tarea llevada adelante por su abogado/a podría reclamar por los daños y perjuicios ocasionados en virtud de la pérdida del recurso de apelación.

Estas diferentes sanciones o llamados de atención a los/as letrados/as que hacen un "uso indebido" de las herramientas de IAGen también constituyen un empoderamiento de la técnica jurídica con el fin de controlar el uso de la IA en las profesiones jurídicas que podría ser extensivo a otras profesiones.

5. La abogacía y la IAGen

En el contexto de lo planteado para este trabajo, nos resulta de interés reflexionar sobre los motivos por los cuales los/as profesionales jurídicos/as hacen uso de las herramientas de IAGen.

El primero de ellos es la amabilidad con la cual se presentan este tipo de herramientas, muchas veces bajo la inevitable frase de "llegó para quedarse". (14) Es cierto que esto es un poco difícil de discutir, cuando la profesión jurídica se ha beneficiado de diversas tecnologías a través del tiempo: desde la máquina de escribir hasta internet, la firma digital, el expediente electrónico, las audiencias virtuales, entre otras tantas. Tal vez, ante la irrupción de la IAGen la pregunta que cabría hacerse hoy es la inversa: ¿cuánto se aprovechan las "tecnologías digitales" de la profesión jurídica?

Así, si bien es absurdo intentar evitar el uso de la inteligencia artificial –por el simple motivo de que no está prohibida y, de intentar hacerlo, nos encontraríamos en un conflicto de otra escala–, lo cierto es que también lo es desapegarla de nuestra propia cultura jurídica, del control jurisdiccional y de su puesta al servicio (de justicia) a la comunidad.

Por otro lado, si bien no existen estadísticas a nivel nacional sobre cuántos/as profesionales utilizan este tipo de herramientas en la abogacía, de una simple búsqueda en cualquier portal de los diferentes colegios profesionales se mantiene la tendencia de una gran oferta de cursos de capacitación sobre herramientas de IA aplicadas a la profesión. El hecho de que los propios colegios profesionales ofrezcan este tipo de capacitaciones legitima de algún modo su uso. Sin embargo, en paralelo no se ofrecen similares espacios para discutir las implicancias de sus aplicaciones concretas.

---

(14) Llorens, María Pilar, *op. cit.*

Del relevamiento de diferentes trabajos académicos surge que el principal consenso sobre la utilización de las herramientas de IAGen es la de mejorar el servicio de justicia tomando como parámetro el factor tiempo. Si bien los índices de confianza en el sistema judicial –en general– no son del todo buenos, lo cierto es que la profesión del/de la abogado/a tampoco cuenta con gran admiración en nuestro país. Entonces, las herramientas de IA para "usos consensuados" se presentan como un medio para mejorar la imagen de un sistema obsoleto, con recursos limitados, fundado en el ahorro de tiempo o, incluso, poder destinarlo a temas "más importantes": acelerar tareas repetitivas, realizar resúmenes o síntesis de casos o de argumentos, identificar aquellos relevantes o corregir textos. En esencia, estas herramientas son más ágiles que los humanos y permiten ahorrar tiempo en procesos que, sin embargo, no se acompañan con reformas procesales que vayan en el mismo sentido.

Ahora bien, los beneficios de la IAGen que se imponen no descartan los temores de los/as profesionales jurídicos/as. Dentro de las preocupaciones ante la llegada de estas herramientas, la primera es la del reemplazo. (15) Hoy los/as abogados/as, además de lidiar con el fenómeno que ocurre cuando el/la cliente/a "googlea" determinadas inquietudes y desafía al/a la profesional, conviven con la afirmación que sostiene que, en pocos años, las tareas de un/a abogado/a junior serán realizadas de manera completa por sistemas de IA. Además, se suma la preocupación acerca de las modificaciones, en sus propios fundamentos y técnicas –como lo hemos reseñado al principio del trabajo–, de las formas de recepción, trámite y resolución de los conflictos sociales que llegan a la órbita del derecho.

Otras preocupaciones en el uso de estas herramientas por parte de los/as abogados/as, en tanto ponen en juego su ética profesional, fueron abordadas en la "Guía para el uso de Inteligencia Artificial para Abogados", que si bien no es un documento oficial, constituye una serie de pautas diagramadas por el Laboratorio Fintech y Legaltech Unidad de Innovación y Transformación Digital del Colegio Público de la Abogacía de la Capital Federal. En este documento "pro uso de la IA" se proponen una serie de directrices genéricas que intentan dar respuesta a una serie de preocupaciones que comparte la comunidad académica en general, tales como la veracidad del contenido que arrojan las herramientas; la protección de los datos de la causa y de sus clientes/as; y programas de capacitación.

Se sostiene como primer objetivo que un uso responsable de los sistemas de IA implica que el humano mitigue los riesgos y, para ello, se recomienda validar todo el contenido de las respuestas de estas herramientas, en especial jurisprudencia y doctrina.

---

(15) Wierzba, Sandra; Cosoy, Diego; Perez Felicioni, Nicolás y Marhaba, Débora. "Justicia digital, inteligencia artificial…", *op. cit*.

Entendemos que este punto podría ser el de más fácil control por parte de los/as abogados/as, mediante una búsqueda en bases de datos públicas de jurisprudencia –en su mayoría disponibles en los portales de los propios tribunales–, a través del chequeo en algún servicio de suscripción tradicional o por medio de una consulta a las bibliotecas jurídicas disponibles y gratuitas.

Por otro lado, si un tercero quisiera chequear la información presentada por un/a letrado/a en un escrito judicial, podría hacer lo mismo: utilizar alguna herramienta que advierte que el contenido se realizó con IA o intentar consultar a un agente conversacional sobre jurisprudencia acerca del tema debatido, para conocer qué resultados arroja. Sin embargo, esto último acaso sea dificultoso porque estas herramientas constituyen en esencia un sistema guiado, customizado para cada usuario/a. Por lo tanto, conforme la utilización, preguntas, instrucciones que cada usuario/a provea en la interacción, serán los resultados que obtendrá.

Es importante remarcar a los/as profesionales jurídicos/as la existencia de las "alucinaciones" de estos sistemas que, *a priori*, consideramos más preocupantes que el tradicional *copy-paste*. Se conocen sanciones a magistrados/as o asimismo, se han declarado nulas sentencias que han abusado de ello. (16) Pero la diferencia es que lo que se copiaba mediante simples funciones de un teclado, sí existía en la realidad. También consideramos que difieren de la estandarización de algunas resoluciones judiciales. Es que la utilización de "plantillas" para resoluciones judiciales que guardan similitudes (regulaciones de honorarios, deserción de recursos de apelación, alimentos provisorios) y que hasta pueden encontrarse en los propios sistemas informáticos de los poderes judiciales, también cuentan con fundamentos existentes, además de que se alojan en sistemas informáticos creados y controlados por y desde los poderes judiciales.

Entonces, frente al riesgo que presenta esta situación, los/as abogados/as cuentan con diversas opciones: cotejar ellos/as mismos/as la veracidad de los contenidos utilizando fuentes públicas y de libre acceso; contar –si poseen los recursos económicos necesarios– con alguna suscripción a un sistema comercial de IA aplicado a la profesión jurídica; o utilizar alguna herramienta creada por iniciativas individuales de acceso libre y gratuito. (17)

Sin embargo, todo ello no obsta a la revisión del/de la abogado/a toda vez que el deber de probidad recae sobre la persona, no sobre un sistema

---

(16) Cámara de Apelaciones en lo Civil y Comercial, Sala Segunda, Mar del Plata, "Iezzi, Jorge Ángel y ot. c/ Auto Club Balcarce y ot. s/ daños y perjuicios" e "Ibarra, Pablo Ariel y ot. c/ Auto Club Balcarce y ot. s/ daños y perjuicios", 15/03/16.

(17) Fundar, Recursero para implementar nuevas tecnologías digitales a la justicia. Disponible en: https://fund.ar/publicacion/recursero-para-implementar-nuevas-tecnologias-digitales-a-la-justicia/

de IAGen. De allí que entendemos que no alcanzaría con un consentimiento informado que deslinde de responsabilidad al/a la profesional, ya que cualquier cláusula en tal sentido sería contraria a la legislación vigente y abusiva.

Esta última valoración se vincula con el segundo objetivo planteado en la guía citada, "Control humano y prohibición de delegación", que pretende dejar en claro que la responsabilidad profesional –jurídica y ética– es del/de la abogado/a. Pesa sobre este/a el control de los resultados producidos por el sistema de IAGen y la decisión de incluirlos o no en un escrito judicial. Es decir, la estrategia legal no podría delegarse en un sistema de IA ya que la eventual responsabilidad recaerá sobre el/la profesional. Sin perjuicio de lo expuesto, dicha afirmación acaso resulta extraña frente a la práctica habitual de pedir recomendaciones a los diferentes agentes conversacionales hasta para los aspectos más íntimos de la vida humana.

Así, por un lado nos referimos a una habitualidad en los usos de estas herramientas de IAGen para la profesión jurídica; por otro, advertimos sobre las responsabilidades que podrían acarrear estos sistemas por sus propios riesgos. Pero en lo que no se indaga de manera suficiente es sobre el hecho de que para hacerse responsable del uso de estas herramientas se necesita contar con un conocimiento del derecho, que en definitiva no estamos seguros de que pueda ser garantizado por el mero uso de las herramientas digitales.

Respecto a las recomendaciones sobre la protección de los datos de las partes en un proceso judicial y la seguridad de las herramientas de IA utilizadas, justificadas en el deber de confidencialidad, a nuestro entender son las que más escollos encuentran en la actualidad. Ello, porque tienden a abordar cuestiones que son sumamente sensibles, pero que en la realidad no tienen una protección efectiva.

Día a día cedemos nuestros datos personales, muchos de ellos sensibles, a distintas aplicaciones que poseen herramientas de IA para sus diferentes procesos. Esta circunstancia crea una tensión para la profesión jurídica ya que el deber de confidencialidad es tal vez el más importante.

Por ello, las recomendaciones de no emplear datos sensibles, anonimizar los que se utilizan en distintas herramientas de IAGen y evaluar sus políticas de privacidad, es una tarea que amerita más que una mera recomendación.

Poner en cabeza de un/a abogado/a analizar y evaluar las políticas de privacidad o términos y condiciones de las distintas herramientas de IAGen es continuar con una lógica abusiva de lo desarrollado hasta el momento. En especial, en lo que atañe a la responsabilidad del/de la usuario/a de las plataformas digitales que continúan considerándose ajenas a cualquier tipo de

responsabilidad. Tras la lectura de los "términos y condiciones" de distintas herramientas de IA que se nos ofrecen para mejorar nuestra vida humana, es posible observar, desde las primeras cláusulas, que se trata de contratos de adhesión, repletos de cláusulas abusivas, muchas veces en idioma extranjero, con términos que poseen grandes dificultades técnicas y lingüísticas y donde el/la único/a responsable de lo que sucede por el uso de ellas es el/la usuario/a, en este caso un/a abogado/a. También, es dable señalar que estas "pautas de convivencia" no permiten modificaciones por parte de sus usuarios/as aunque, por el contrario, sí pueden cambiar sin previo aviso y solo por decisión de sus propietarios/as quienes, además, se rigen por leyes y en una jurisdicción foránea. (18)

Nos cuestionamos sobre la efectividad de la firma de un formulario de consentimiento informado. La recomendación en sí podría resultar valiosa, en general, para una eventual defensa del/de la abogado/a respecto de la información brindada a su cliente/a.

Sin perjuicio de ello, en el caso concreto de la responsabilidad profesional por el uso de IAGen por parte de abogados/as, tal como se ha señalado con anterioridad, la firma de un formulario de consentimiento informado no podría constituirse en un eximente por haber citado fundamentos legales inexistentes –alucinaciones–.

Por otro lado, si la información brindada al/a la cliente/a en dicho formulario de consentimiento informado fuera sobre la mera utilización de estas herramientas de IAGen tampoco lo exoneraría de su responsabilidad en virtud del deber de confidencialidad, más aún en casos donde el acceso a la información del expediente no es pública en razón de la materia (familia, derecho penal).

También acaso deberían considerarse las características propias del empleo de un documento como el formulario de consentimiento informado en nuestro país, toda vez que si bien es posible su utilización en todas las profesiones liberales, en el caso local es mayor en la medicina por expresa disposición legal. (19)

Entendemos que sería un valioso aporte que las instituciones que correspondan elaboraran un documento como formulario de consentimiento informado en conjunto, a partir del análisis de las prácticas reales y sus eventuales conflictos, con el fin de alcanzar una respuesta consensuada que proteja con mayor énfasis al/a la profesional.

---

(18) Wierzba, Sandra; Cosoy, Diego; Pérez Felicioni, Nicolás y Marhaba, Débora. "Justicia digital, inteligencia artificial…", *op. cit*.

(19) Cámara Nacional de Apelaciones en lo Civil, "Maddaleno de Diego, Daniel Alberto c/ Zynger, Fernando s/ daños y perjuicios", 14/05/2025.

De todos modos, advertimos que en los casos analizados se le impusieron más cargas y responsabilidades a quien litiga y utiliza estas herramientas que a los propios funcionarios judiciales. Si tomamos como ejemplo la sentencia "Giacomino" observamos que los magistrados sugieren que aun si el abogado hiciera firmar un formulario de consentimiento informado a su cliente defendido sobre la posibilidad del uso de sistemas de IAGen, igual debería chequear los resultados arrojados por dicho sistema. No cabe duda de las responsabilidades de los/as profesionales, pero no se ha visto en ningún caso que un tribunal advierta que estas herramientas son utilizadas por sus integrantes. En otras palabras, se le exige a una parte un apego a normas jurídicas y éticas que no son cumplidas por los propios tribunales donde, a pesar de los avances e intentos de transparencia, sus sistemas continúan siendo opacos.

En tal sentido, parecería ser más acertada la propuesta de informar a los colegios profesionales para que se trabaje y aborde el tema en cuestión de manera conjunta, ética y responsable. Pero, para ello, todas las partes involucradas deberían aclarar y advertir cuando se utilizan estos sistemas.

En lo que refiere a la capacitación en este tipo de herramientas de IAGen, más allá de las valoraciones que efectuamos en este trabajo, entendemos que cuando se sugiere la capacitación del/de la abogado/a todavía se hace referencia solo a usos genéricos, sin detenerse en abordajes prácticos más específicos que permitan detectar los riesgos que se advierten en los diferentes documentos de "buenas prácticas" o en normativa comparada que hasta prohíbe ciertos usos.

6. Conclusiones

A modo de cierre, en este trabajo exploratorio quisiéramos realizar algunas valoraciones y dejar planteadas algunas inquietudes que acaso servirán para futuros aportes.

En primer lugar, respecto a las "alucinaciones" de los sistemas de IAGen utilizadas en distintos escritos judiciales por parte de los/as abogados/as y que han generado respuesta de la magistratura, proponemos que sean entendidas como correlaciones no jurídicas, ya que no se han interiorizado las pautas éticas, ni la causalidad de las sanciones que ellas acarrean, sino que se cumplen órdenes para las cuales fueron programadas. Ello no implica de modo alguno que las herramientas de IAGen no puedan afinarse y hacerse más certeros los patrones a reconocer, pero distan de poseer crítica y un criterio acorde.

Por otro lado, y en consonancia con lo anterior, hemos observado cómo la responsabilidad por el uso de estas herramientas está centrada en los/as abogados/as, por lo que por más perfeccionados que estén lo sistemas

de IAGen, si en algo fallan, quienes serán responsables son quienes lo han utilizado. En este sentido, hemos notado cómo el uso termina por precarizar el empleo por parte de los/as abogados/as en general, en especial los/as jóvenes, pero sobre todo concentrando el aprendizaje en cuestiones prácticas de nivel general, que no solo no garantizan que estemos aprendiendo y aprehendiendo derecho, sino que no está garantizada la formación de un uso seguro para la profesión y para quienes deciden sobre los daños ocasionados con relación a ese uso.

Si bien todavía resta explorar, entre otras tantas cuestiones, el régimen de responsabilidad de los fabricantes y cómo sortear los obstáculos jurisdiccionales para un eventual reclamo o pedido de explicaciones a empresas que están radicadas en el exterior y que se rigen por normas foráneas, pudimos vislumbrar a partir del análisis de la mayoría de los casos seleccionados la importancia que todavía tiene asignada el derecho. Es decir, ya sea para legitimidad judicial de estándares provistos por la IAGen –cuestión con la que hay que ser cuidadosos con los usos intuitivos que pueden derivar en legitimación inconsciente– pero sobre todo por cómo todavía existe la posibilidad de apelar a técnicas exclusivamente jurídicas (como la de presunción o las medidas para mejor proveer) para controlar y amoldar el uso al derecho.

También, de algún modo, continuar con la utilización de las técnicas propias del derecho, evitar la privatización de la argumentación jurídica o la tercerización del control, situación que sería paradójica si se delega la detección de utilización de inteligencia artificial en casos que trajeran alguna consecuencia negativa en otros sistemas de inteligencia artificial.

Por último, y realmente importante, queremos hacer hincapié en cómo se podría revalorizar el rol de las colegiaturas, para llamar la atención a sus colegiados/as, orientarlos/as y proyectar prácticas y capacitaciones sobre el uso de IAGen, herramientas con las que tendrían la oportunidad de hacer del ejercicio profesional una actividad responsable y de calidad.

Virtual labs en la educación jurídica: una propuesta de transformación pedagógica mediante inteligencia artificial

Por Jorge A. Cerdio Herrán (*)

Resumen: Este capítulo presenta un nuevo modelo pedagógico, el *Virtual Lab*, como respuesta a la obsolescencia de la educación jurídica tradicional. Partiendo del paradigma de los laboratorios científicos multiagente, la propuesta no busca que la IA genere derecho de forma autónoma, sino aumentar la capacidad humana para crearlo. Se articula una arquitectura donde estudiantes, guiados por un 'profesor-orquestador', colaboran con agentes de IA especializados a través de un ciclo de descubrimiento en cinco fases para formular soluciones jurídicas innovadoras. Tras analizar los desafíos inherentes, se concluye que el *Virtual Lab* transforma el propósito de las facultades de derecho, pasando de la transmisión de doctrina a la formación de arquitectos de futuros normativos.

Palabras clave: educación jurídica, inteligencia artificial, *Virtual Lab*s, innovación pedagógica, conocimiento original.

Abstract: This chapter introduces a new pedagogical model, the *Virtual Lab*, as a response to the obsolescence of traditional legal education. Drawing from the paradigm of multi-agent scientific laboratories, the proposal does not aim for AI to autonomously generate law, but rather to augment the human capacity to create it. It articulates an architecture where students, guided by a 'Professor-Orchestrator,' collaborate with specialized AI agents through a five-phase discovery cycle to formulate innovative legal solutions. After analyzing the inherent challenges, the article concludes that the *Virtual Lab* transforms the purpose of law schools, shifting them from the transmission of doctrine to the training of 'architects of normative futures.'

(*) Profesor de tiempo completo de Teoría y Filosofía del Derecho; director de los grados y del *college*, Facultad de Derecho, Empresa y Gobierno, Universidad Francisco de Vitoria, Madrid. Ha sido miembro del CAHAI (Comité Ad Hoc para la IA y los Derechos Humanos) del Consejo de Ministros de Europa. E-mail. jorge.cerdio@ufv.es.

Key words: legal education, artificial intelligence, *Virtual Lab*s, pedagogical innovation, original knowledge.

1. INTRODUCCIÓN

La educación jurídica atraviesa una crisis existencial. La irrupción de la inteligencia artificial generativa ha tornado obsoletas las competencias tradicionales que por siglos han definido al buen jurista, obligándonos a preguntar: ¿qué valor humano aportamos cuando una máquina redacta y analiza con mayor eficiencia? Este capítulo sostiene que esta crisis no es una amenaza, sino una oportunidad para una transformación radical: abandonar el paradigma de la transmisión de doctrina para adoptar uno centrado en la creación de conocimiento jurídico original.

Para materializar esta visión, se presenta el modelo del *Virtual Lab* jurídico. Inspirado en los laboratorios científicos multiagente, esta propuesta pedagógica no busca que la IA reemplace al jurista, sino aumentar la capacidad de los estudiantes para que, guiados por un profesor-orquestador, colaboren con agentes especializados y aprendan a construir soluciones inéditas para los dilemas de nuestra era. Se trata de un cambio de propósito: formar no a meros operadores del sistema, sino a los futuros arquitectos del derecho.

Para desarrollar esta tesis, la exposición seguirá una estructura progresiva. La sección 1 establecerá el imperativo de transformación, diagnosticando la crisis del modelo educativo y presentando el paradigma del *Virtual Lab*. La sección 2 describirá la arquitectura del sistema, detallando los roles de los agentes especializados y del profesor-orquestador. La sección 3 se centrará en la metodología de implementación, explicando el ciclo de descubrimiento jurídico y las métricas de originalidad. La sección 4 abordará los indispensables desafíos éticos e institucionales que esta innovación conlleva. Finalmente, la sección 5 presentará las conclusiones, reflexionando sobre el nuevo contrato educativo que esta propuesta implica y el rol del jurista como arquitecto de futuros normativos.

2. EL IMPERATIVO DE TRANSFORMACIÓN EN LA EDUCACIÓN JURÍDICA

Toda propuesta de transformación radical debe partir de un diagnóstico claro. Por ello, esta primera sección se dedica a establecer el 'porqué' de la innovación aquí presentada. El recorrido argumental comenzará (2.1) por analizar la crisis existencial que la IA generativa impone al modelo tradicional de enseñanza del derecho. Una vez establecido el problema, se introducirá (2.2) el paradigma del *Virtual Lab*, una solución inspirada en la ciencia que servirá como pilar para el resto del capítulo.

2.1. La crisis del modelo tradicional ante la IA generativa

La irrupción de los grandes modelos de lenguaje ha acelerado una crisis que venía gestándose desde hace décadas (McEntee & Koenig, 2012; Rhode, 2013; Young, 2021). Cuando ChatGPT puede identificar en segundos todos los precedentes relevantes para un caso, cuando Claude analiza contratos con una exhaustividad que supera a la mayoría de los asociados junior, cuando los sistemas de IA redactan escritos procesales técnicamente impecables (Goswami, 2025; Prakash, 2025), surge una pregunta existencial: ¿qué valor único aporta el jurista humano?

La respuesta no está en competir con las máquinas en su terreno —la reproducción y combinación de información existente— sino en reclamar lo que permanece irreductiblemente humano: la capacidad de formular las preguntas jurídicas que nadie ha planteado antes, de imaginar soluciones normativas para problemas que emergen en los intersticios del ordenamiento jurídico, de traducir anhelos sociales difusos en arquitecturas legales concretas. Sin embargo, nuestro modelo pedagógico actual, anclado en la transmisión vertical de doctrina establecida y la memorización de códigos, forma precisamente las competencias que la IA ha vuelto redundantes, un punto de fricción sobre el que la literatura de la reforma educativa lleva años insistiendo (Kleeger, 2019).

Esta crisis demanda algo más radical que integrar herramientas digitales en las aulas o enseñar *Legal Tech* como una asignatura más. Requiere reimaginar fundamentalmente qué significa pensar jurídicamente en el siglo XXI. Y es aquí donde emerge una posibilidad fascinante e inexplorada: ¿qué pasaría si transformáramos las facultades de derecho en laboratorios de descubrimiento jurídico?

2.2. El paradigma del *Virtual Lab*: de la ciencia al derecho

El punto de partida es la traslación de un paradigma revolucionario de la ciencia a la *pedagogía jurídica*: el *Virtual Lab*. Estudios como los de Zou et al. (2025) demuestran que la orquestación de agentes de IA puede generar conocimiento científico nuevo. La pregunta que se deriva no es si la IA puede descubrir derecho por sí sola, sino si podemos usar un modelo análogo para transformar la educación, enseñando a los estudiantes a *generar conocimiento original y aumentar así sus capacidades creativas*.

Esta propuesta educativa se basa en una traducción sistemática de los conceptos clave. Mientras el laboratorio científico valida sus hallazgos contra la realidad física, el laboratorio jurídico lo hace contra principios de justicia y coherencia sistémica. Así, la *replicabilidad experimental* se transforma en *consistencia doctrinal y la falsabilidad popperiana en resistencia al escrutinio jurisprudencial*.

Este trabajo presenta, por primera vez, una traducción sistemática del paradigma del laboratorio virtual científico al ámbito de la educación jurídica. No se trata de otra propuesta más sobre IA en el derecho —terreno ya abundantemente explorado— sino de algo cualitativamente distinto: la transformación de las facultades de derecho en laboratorios donde estudiantes y agentes de IA colaboran para generar conocimiento jurídico original.

La analogía entre descubrimiento científico y descubrimiento jurídico es poderosa pero requiere traducción cuidadosa. Mientras el laboratorio científico busca verdades empíricas verificables contra la realidad física, el laboratorio jurídico busca construcciones normativas validables contra principios de justicia y coherencia sistémica. La replicabilidad experimental se transforma en consistencia doctrinal; la falsabilidad popperiana se convierte en resistencia al escrutinio jurisprudencial; el *peer review* científico muta en deliberación jurídica intersubjetiva.

Pero hay una diferencia fundamental que hace esta traducción aún más fascinante: mientras las moléculas existen en un universo de leyes físicas inmutables, el derecho existe en un universo de significados sociales en constante evolución. Un *nanobody* funciona o no funciona según criterios objetivos; una interpretación jurídica "funciona" según su capacidad de persuadir, de resonar con los valores de una época, de articular aspiraciones colectivas. Esta diferencia no debilita la analogía; la enriquece. Sugiere que el *Virtual Lab* jurídico podría ser incluso más transformador que su contraparte científica, porque no solo descubre lo que es, sino que ayuda a imaginar lo que podría y debería ser.

La propuesta que desarrollaremos no es, por tanto, una aplicación mecánica de métodos computacionales al derecho. Es una reconceptualización radical de la educación jurídica que abraza la IA no como amenaza sino como catalizador de una nueva forma de creatividad jurídica. En este nuevo paradigma, el valor del jurista no reside en lo que sabe, sino en las preguntas que es capaz de formular; no en su capacidad de reproducir argumentos, sino en su habilidad para orquestar perspectivas diversas hacia síntesis originales; no en su memoria de precedentes, sino en su imaginación para concebir futuros normativos posibles. Para materializar este potencial, es necesario definir tanto la arquitectura de este laboratorio como el método que guiará el descubrimiento en su interior.

3. ARQUITECTURA DEL *VIRTUAL LAB* JURÍDICO

El *Virtual Lab* jurídico se concibe como un ecosistema con dos componentes fundamentales: los instrumentos de análisis, que son los agentes de IA especializados, y el director de la investigación, que es el profesor-

orquestador. Ambos elementos redefinen sus roles tradicionales para crear un nuevo entorno de descubrimiento. Una vez establecido el paradigma, es necesario definir sus componentes y funcionamiento. Esta sección detallará, en primer lugar (3.1), el ecosistema del laboratorio, compuesto por los agentes especializados y el profesor-orquestador. A continuación (3.2), se expondrá el ciclo de descubrimiento jurídico, el método de cinco fases que estructura el aprendizaje. Finalmente (3.3), se presentarán las métricas de evaluación diseñadas para medir la originalidad jurídica.

3.1. Los agentes especializados

El corazón del *Virtual Lab* son sus agentes especializados, que no son repositorios de información, sino perspectivas epistemológicas computacionales. Su diseño no busca la armonía, sino el conflicto productivo y la dialéctica, forzando a los estudiantes a encontrar síntesis originales a partir de sus interacciones. Cada agente encarna una tradición y metodología jurídica específica:

- Agente de derecho constitucional: evalúa toda propuesta desde la jerarquía normativa y el prisma de los derechos fundamentales, manteniendo una sospecha sistemática hacia el poder.

- Agente de derecho civil: opera con la lógica de la autonomía privada, el equilibrio contractual y la justicia conmutativa entre iguales formales.

- Agente de derecho penal: porta la paradoja del sistema punitivo, confrontando la dogmática del delito con los fines filosóficos del castigo (retribución, prevención, rehabilitación).

- Agente de derecho administrativo: materializa la racionalidad burocrática (procedimientos, competencias), pero también sus tensiones con las garantías ciudadanas y la eficiencia.

- Agente de filosofía del derecho: activa diversos metalenguajes (iusnaturalista, positivista, realista, crítico) para desnaturalizar las premisas que los otros agentes dan por sentadas.

- Agente de teoría del derecho: funciona como guardián de la coherencia sistémica, distinguiendo entre contradicciones destructivas y aquellas que impulsan la evolución del derecho.

La interacción deliberadamente disonante entre estos agentes es lo que genera el descubrimiento. Ante un problema como la "muerte digital", cada agente revela una dimensión distinta del conflicto (propiedad, delito, derechos personalísimos, jurisdicción), obligando a los estudiantes a buscar una solución jurídica que integre todas las perspectivas en una síntesis novedosa.

3.2. El profesor-orquestador como director del laboratorio

Si los agentes son los instrumentos, el profesor-orquestador es el director de esta sinfonía disonante. Pero la metáfora musical, aunque sugerente, es insuficiente. Un director de orquesta busca armonía; el profesor-orquestador busca fricción productiva. Su rol no es resolver las tensiones entre agentes sino amplificarlas hasta que los estudiantes no tengan más remedio que inventar síntesis originales.

Esta figura representa la transformación más radical del rol docente desde la invención de la universidad medieval. El profesor tradicional era el portador del saber, el intermediario entre el texto canónico y el estudiante ignorante. Esa función ha muerto: cualquier estudiante con un smartphone accede instantáneamente a más información jurídica de la que cualquier profesor podría memorizar. El profesor-orquestador no transmite información sino que diseña experiencias de descubrimiento.

Consideremos un ejemplo concreto. El tema de la semana es "personalidad jurídica de entidades artificiales". El profesor tradicional explicaría la doctrina de la persona jurídica, desde la teoría de la ficción hasta la realidad técnica. El profesor-orquestador hace algo radicalmente distinto: lanza una provocación al *Virtual Lab*: "Una DAO (Organización Autónoma Descentralizada) basada en *blockchain* ha 'decidido' —mediante *smart contracts*— donar fondos a una ONG. La transacción falló por un error de código y los fondos se perdieron. ¿Quién responde?".

Los agentes explotan en actividad contradictoria. El civil no encuentra un sujeto de imputación clara. El penal busca desesperadamente un autor. El administrativo no sabe qué autoridad tiene jurisdicción. El constitucional se pregunta si las DAOs tienen derechos. El filosófico cuestiona si "decidir" requiere consciencia. El teórico advierte que cualquier solución crea precedentes sistémicos peligrosos.

El profesor-orquestador no resuelve este caos; lo cultiva. Guía a los estudiantes para que vean que están ante una genuina laguna jurídica, un espacio donde el derecho existente simplemente no llega. Pero no los deja en la perplejidad paralizante. Los desafía: "Diseñen un marco regulatorio que resuelva esto sin destruir la innovación tecnológica ni abandonar la seguridad jurídica."

Aquí surge la dimensión más sutil del rol: calibrar el desafío al nivel exacto de desarrollo de cada estudiante. Para novatos, puede simplificar activando solo dos agentes. Para avanzados, puede introducir complicaciones: "La DAO opera simultáneamente en 15 jurisdicciones." Esta calibración no es solo técnica sino profundamente humana. El profesor-orquestador debe leer la frustración productiva (que impulsa) y distinguirla de la destructiva (que paraliza).

Crucialmente, este rol preserva e intensifica la dimensión humana de la educación jurídica. Los agentes pueden procesar información y generar argumentos, pero no pueden sentir la angustia ética de enviar a alguien a prisión, la satisfacción estética de una construcción jurídica elegante, o la responsabilidad existencial de decidir sobre derechos fundamentales. El profesor-orquestador asegura que estas dimensiones afectivas y éticas no se pierdan en la fascinación tecnológica.

Un momento crítico ocurre cuando los estudiantes comienzan a sobredepender de los agentes. El profesor-orquestador debe entonces realizar lo que podríamos llamar "el giro socrático": apagar momentáneamente el *Virtual Lab* y preguntar: "¿Qué piensan ustedes, no los agentes? ¿Qué sienten sobre esta solución? ¿Dormirían tranquilos si su interpretación enviara a alguien a prisión por 20 años?" Es en estos momentos cuando la educación jurídica recupera su dimensión más profunda: formar no solo técnicos del derecho sino custodios de la justicia.

4. Metodología de implementación en el contexto jurídico

Habiendo presentado la arquitectura del *Virtual Lab*, es crucial detallar su motor operativo: la metodología de implementación. Este apartado se centrará en el 'cómo' se genera el conocimiento. Para ello, se expondrá en primer lugar (4.1) el ciclo de descubrimiento jurídico, un proceso de cinco fases que estructura el aprendizaje creativo. A continuación (4.2), se definirán las métricas de originalidad jurídica, las herramientas diseñadas para evaluar la calidad de las innovaciones producidas.

4.1. El ciclo de descubrimiento jurídico

El método es el mensaje. Marshall McLuhan lo intuyó para los medios de comunicación, pero su *insight* cobra urgencia existencial en la educación jurídica algorítmica. No podemos enseñar innovación jurídica mediante pedagogías de reproducción, como no podemos enseñar natación en un aula sin agua. El *Virtual Lab* jurídico requiere una metodología que no solo tolere la incertidumbre sino que la cultive, que no solo permita el error sino que lo celebre como estación necesaria hacia el descubrimiento.

El ciclo que proponemos no es lineal sino helicoidal: cada vuelta retorna a preguntas similares pero en un nivel de complejidad superior. Se inspira en el método científico pero lo traiciona creativamente, porque el derecho no descubre leyes naturales preexistentes.

Fase 1 - Identificación del problema: El arte de ver lagunas

Todo comienza con aprender a ver lo que no está. Los estudiantes de derecho tradicionales son entrenados para encontrar la norma aplicable; nosotros los entrenamos para detectar donde no hay norma aplicable, o peor aún, donde las normas aplicables generan consecuencias absurdas o injustas. Esta es una inversión gestáltica de la mirada jurídica: en lugar de ver el ordenamiento como un sistema completo con algunas lagunas excepcionales, comenzamos a verlo como un archipiélago de islas normativas rodeadas por un océano de indeterminación.

El profesor-orquestador no presenta problemas; presenta síntomas. "Una *influencer* virtual generada por IA firmó un contrato de publicidad y ahora la marca quiere demandarla por incumplimiento." Los estudiantes deben excavar hasta encontrar el problema real: ¿puede un ente sin consciencia tener responsabilidad contractual? ¿Quién responde por las "decisiones" de una IA? ¿Es esto siquiera un contrato?

Los agentes asisten en esta excavación, pero de manera socrática. El agente de teoría del derecho mapea las presuposiciones conceptuales que fallan. El filosófico pregunta: "¿Qué teoría de la agencia estamos asumiendo?" El civil busca desesperadamente analogías: "¿Es como un menor? ¿Como una empresa? ¿Como un esclavo romano?" Cada analogía falla de manera instructiva.

Fase 2 - Exploración multiperspectiva: La polifonía del derecho

Identificado el problema genuino, se desata lo que Grimes (2015) llama "exploración salvaje" pero que nosotros preferimos denominar "proliferación controlada de perspectivas". Cada agente ataca el problema desde su tradición disciplinar, pero —y esto es crucial— sin intentar resolverlo prematuramente. El objetivo es multiplicar las dimensiones del problema, no reducirlas.

Tomemos el caso real que exploraron estudiantes de tercer año: la regulación del "derecho al olvido algorítmico" no solo en buscadores sino en modelos de lenguaje. El agente constitucional ve un conflicto entre privacidad y libertad de expresión. El administrativo ve un problema de implementación técnica: ¿cómo "olvidar" información en un modelo entrenado? El penal ve potencial para encubrimiento de delitos. El civil ve derechos de personalidad y daño moral. El filosófico pregunta si el olvido selectivo es compatible con la verdad histórica. El teórico advierte que regular el "olvido" presupone una teoría de la "memoria" jurídica que no tenemos.

Esta fase es deliberadamente caótica. Casey (2014) la describe como "desorientación productiva". Los estudiantes experimentan lo que los juristas profesionales rara vez admiten: que ante problemas genuinamente nue-

vos, el derecho balbucea antes de articular. Esta humildad epistemológica es esencial para la innovación real.

Fase 3 - Síntesis creativa: El salto al vacío

Aquí ocurre la magia y el terror del descubrimiento jurídico. Los estudiantes deben proponer soluciones que no derivan deductivamente de normas existentes ni inductivamente de casos anteriores, sino que emergen abductivamente de la imaginación jurídica disciplinada. Como señalan Burton (2015) y Pascoe (2017), este es el momento donde la educación jurídica tradicional falla más catastróficamente: nunca enseñamos a inventar.

Pero la invención jurídica no es arbitraria. Opera bajo restricciones severas: coherencia sistémica, viabilidad práctica, legitimidad democrática, justicia material. Los estudiantes no pueden simplemente decretar: "Declaremos a las IAs personas jurídicas". Deben construir un puente argumentativo desde el derecho existente hacia su propuesta, un puente lo suficientemente sólido para soportar el peso del escrutinio jurídico pero lo suficientemente flexible para adaptarse a realidades cambiantes.

El *Virtual Lab* asiste mediante lo que llamamos "provocaciones generativas". Los agentes no ofrecen soluciones sino que sugieren híbridos conceptuales: "¿Y si combinamos la figura del *trust* anglosajón con la teoría del órgano alemana?" "¿Qué pasaría si aplicamos la doctrina del levantamiento del velo corporativo a las IAs?" Estas provocaciones no son respuestas sino catalizadores para que los estudiantes generen sus propias síntesis originales.

Un grupo desarrolló el concepto de "personalidad jurídica gradual" para IAs: entidades que adquieren capacidades jurídicas incrementales según su nivel de autonomía decisional verificable. No es una persona ni una cosa, sino una categoría intermedia nueva, como lo fue en su momento la empresa. La propuesta es imperfecta, controversial, quizás inviable. Pero es genuinamente original.

Fase 4 - Simulación de consecuencias: El laboratorio del futuro

Una molécula diseñada computacionalmente debe probarse en el mundo físico. Una innovación jurídica debe probarse en el mundo social, pero no podemos experimentar con sociedades reales como con ratones de laboratorio. El uso de simulaciones para este fin no es nuevo en el derecho administrativo (Botein, 1974), pero aquí el *Virtual Lab* ofrece una capacidad única: simular las consecuencias sistémicas de innovaciones jurídicas antes de su implementación.

Los agentes, trabajando en conjunto, modelan escenarios: ¿Qué pasaría si se implementa la "personalidad jurídica gradual"? El agente civil simula miles de contratos potenciales. El penal identifica nuevas formas de crimi-

nalidad posible. El administrativo mapea los cambios regulatorios necesarios. El constitucional evalúa potenciales violaciones de derechos. El teórico traza las ondas de propagación conceptual: si aceptamos esto aquí, ¿qué más debemos aceptar por consistencia?

Pero la simulación más importante es la humana. Los estudiantes *role-play* diferentes actores sociales: ¿Cómo usaría Amazon esta nueva figura? ¿Cómo la resistirían los sindicatos? ¿Qué harían los hackers? Esta combinación de simulación computacional y humana revela consecuencias no intencionales que ningún método aislado detectaría.

Un descubrimiento perturbador emergió cuando simulamos el "derecho al olvido algorítmico": implementarlo completamente haría imposible la investigación histórica digital. El remedio sería peor que la enfermedad. Los estudiantes tuvieron que volver a la Fase 3, pero ahora con una comprensión más profunda de las restricciones del problema.

Fase 5 - Validación doctrinal: El tribunal de la razón jurídica

La validación en el *Virtual Lab* jurídico no es binaria (válido/inválido) sino dimensional. Evaluamos las propuestas en múltiples ejes que capturan diferentes aspectos de la excelencia jurídica, siguiendo los principios de evaluación auténtica que proponen Vlachopoulos (2024) y Bahar & Maker (2025).

Primero, la *coherencia interna:* ¿La propuesta es consistente consigo misma o contiene contradicciones que la harían inaplicable? El agente de teoría del derecho realiza un análisis lógico despiadado, identificando circularidades, regresiones infinitas, y paradojas autorreferenciales.

Segundo, la *coherencia sistémica:* ¿Cómo se integra la propuesta con el ordenamiento existente? No buscamos armonía perfecta —eso sería conservadurismo disfrazado de rigor— sino fricción productiva. Las mejores innovaciones jurídicas crean tensiones que obligan al sistema a evolucionar.

Tercero, la *elegancia conceptual:* ¿La solución es innecesariamente compleja? El derecho, como las matemáticas, valora la simplicidad elegante sobre la complejidad barroca. Una solución que requiere 47 excepciones probablemente está mal concebida desde la raíz.

Cuarto, la *robustez práctica:* ¿Sobrevive la propuesta al contacto con la realidad social? Aquí intervienen evaluadores externos: jueces, abogados practicantes, reguladores. Su escepticismo profesional es invaluable. "Esto es hermoso en teoría, pero ¿han considerado los costos de transacción?"

Quinto, y más elusivo, la *resonancia normativa:* ¿La propuesta articula valores que la sociedad reconoce como propios, incluso si no los había arti-

culado antes? Esta es la diferencia entre innovación jurídica genuina y mero tecnicismo *clever*. Las grandes innovaciones jurídicas —el habeas corpus, los derechos sociales, el principio precautorio— no solo resuelven problemas técnicos sino que expresan aspiraciones morales colectivas.

4.2. Métricas de originalidad jurídica

Medir la originalidad en el derecho es como medir la belleza en el arte: necesario pero peligroso. Necesario porque sin criterios evaluativos no hay aprendizaje dirigido. Peligroso porque los criterios pueden osificarse en fórmulas que matan precisamente la creatividad que buscamos cultivar. Siguiendo a Siems (2008), proponemos métricas que son herramientas heurísticas, no algoritmos determinísticos.

Novedad interpretativa: El coeficiente de sorpresa

No toda diferencia es innovación. Cambiar "debe" por "deberá" en una norma no es innovación; reconceptualizar la obligación jurídica sí lo es. Medimos la novedad no por la distancia superficial respecto a interpretaciones existentes sino por lo que llamamos "coeficiente de sorpresa": ¿cuántos juristas competentes no habrían pensado en esta solución?

Pero la sorpresa sola no basta. Las propuestas absurdas también sorprenden. La novedad valiosa es la que, una vez articulada, parece retrospectivamente inevitable. "¿Cómo no lo vimos antes?" es la reacción que buscamos. Como la figura del trust, que ahora parece obvia pero tardó siglos en inventarse.

El *Virtual Lab* mide esto mediante análisis comparativo: ¿existe esta solución o similar en algún ordenamiento jurídico? ¿En algún periodo histórico? ¿En alguna propuesta doctrinal? Mientras más vacío el espacio conceptual donde aterriza la propuesta, mayor su novedad. Pero también evaluamos la "fertilidad conceptual": ¿cuántos problemas adicionales podría resolver esta innovación?

Coherencia sistémica: La tensión óptima

La coherencia absoluta es parálisis; la incoherencia absoluta es caos. La innovación jurídica habita en la tensión óptima entre ambas. Una propuesta que encaja perfectamente en el sistema existente no innova; una que lo destruye completamente no es derecho sino revolución.

Medimos esto mediante lo que denominamos "análisis de ondas de choque": cuando insertamos la propuesta en el sistema jurídico simulado, ¿qué normas deben ajustarse? ¿Qué doctrinas deben reinterpretarse? ¿Qué instituciones deben reformarse? Buscamos propuestas que generen ondas

de magnitud media: suficientes para transformar, insuficientes para destruir.

Un estudiante propuso que los *smart contracts* fueran considerados "actos jurídicos automatizados" sujetos a un régimen especial. La propuesta requería ajustar la teoría del consentimiento, la doctrina del error, y los mecanismos de impugnación, pero no destruía el derecho contractual sino que lo extendía creativamente. Esa es la tensión óptima.

Aplicabilidad práctica: El test del lunes por la mañana

El derecho no es filosofía aplicada ni ingeniería social; es ambas y ninguna. Una innovación jurídica brillante pero inaplicable es un ejercicio académico; una aplicable pero torpe es un parche temporal. Buscamos el punto dulce: elegancia conceptual con viabilidad práctica.

El "test del lunes por la mañana" pregunta: ¿puede un juez de primera instancia, sobrecargado de trabajo, aplicar esta innovación sin un doctorado en filosofía del derecho? ¿Puede un abogado explicársela a su cliente en 10 minutos? ¿Pueden las instituciones existentes implementarla sin rediseño total?

Pero no confundimos simplicidad de aplicación con simplicidad de concepción. Los derechos humanos son conceptualmente sofisticados pero prácticamente aplicables: "no torturar" es claro aunque el fundamento filosófico sea complejo. Buscamos innovaciones con interfaces simples sobre arquitecturas profundas.

La aplicación del ciclo que venimos de presentar transforma el aprendizaje en un acto de creación, capacitando a los estudiantes para ir más allá de la doctrina existente. Es precisamente este método el que les permitiría abordar problemas emergentes y prototipar soluciones innovadoras, como la formulación de un 'derecho a la discontinuidad digital' para identidades en el metaverso o el diseño de un 'Código de Procedimiento Algorítmico' que garantice la transparencia de la IA en la administración pública. Sin embargo, una propuesta tan transformadora no está exenta de importantes desafíos éticos e institucionales que deben ser abordados.

5. DESAFÍOS Y CONSIDERACIONES ÉTICAS

Tras haber delineado la arquitectura y la metodología del *Virtual Lab*, es imperativo adoptar una perspectiva crítica. La innovación tecnológica, especialmente en un campo tan normativo y conservador como el derecho, no es un proceso neutral; genera tensiones, cuestiona estructuras de poder consolidadas y plantea dilemas éticos profundos. Ignorar estos aspectos sería caer en un optimismo tecnológico ingenuo. Por tanto, un análisis rigu-

roso de los posibles riesgos es fundamental, no para invalidar la propuesta, sino para fortalecerla y asegurar su implementación responsable.

Este apartado analiza los principales riesgos y desafíos. Se abordarán, primero (5.1), los desafíos pedagógicos e institucionales, como el 'solucionismo tecnológico' y la resistencia al cambio. Posteriormente (5.2), se examinarán los riesgos sistémicos relacionados con la gobernanza de los datos educativos y la equidad.

5.1. Desafíos pedagógicos e institucionales

Existe un peligro seductor en el *Virtual Lab* jurídico: creer que hemos encontrado la solución definitiva a la crisis de la educación jurídica. Esta tentación del solucionismo tecnológico —la creencia de que todo problema social tiene una solución técnica elegante— es particularmente perniciosa en el derecho, donde las respuestas "correctas" son a menudo indeterminadas y los valores en conflicto irreducibles a algoritmos (Morozov, 2015).

El *Virtual Lab* no es, ni puede ser, un oráculo jurídico que dispensa verdades normativas. El derecho, a diferencia de la química molecular que inspiró el modelo de Zou et al., no busca descubrir leyes naturales preexistentes sino construir ficciones operativas que organizan la convivencia humana. Una molécula funciona o no funciona según criterios empíricos objetivos; una norma jurídica "funciona" según su capacidad de articular valores compartidos, resolver conflictos sociales y preservar la dignidad humana. Ningún algoritmo, por sofisticado que sea, puede resolver definitivamente las tensiones entre libertad y seguridad, entre eficiencia y equidad, entre innovación y tradición.

El riesgo más profundo no es que el *Virtual Lab* falle técnicamente, sino que tenga demasiado éxito en sus propios términos. Si los estudiantes llegan a creer que toda pregunta jurídica tiene una respuesta óptima computable, habremos creado tecnócratas del derecho, no juristas. Este es uno de los múltiples desafíos éticos que la IA plantea tanto a la práctica profesional como a la educación superior (Bizzuka Legal, 2025; García-López et al., 2025; Stockholm University, 2020). La tecnología debe amplificar la capacidad de análisis y síntesis, pero no puede sustituir el juicio prudencial —esa sabiduría práctica aristotélica— que distingue al buen jurista del mero operador legal.

La resistencia al *Virtual Lab* no vendrá principalmente de la tecnofobia sino de algo más profundo: el cuestionamiento de qué constituye conocimiento jurídico legítimo y una resistencia institucional al cambio bien documentada en la literatura (Polden, 2017; Robson & Huffman, 2019; Giddens & Siliquini-Cinelli, 2023). El derecho es un campo profundamente conservador, no por accidente sino por diseño. La estabilidad jurídica, la seguridad

normativa, el respeto al precedente —todos estos valores fundamentales del derecho están en tensión con la innovación disruptiva que el *Virtual Lab* promueve.

Imaginemos el escepticismo de un magistrado del Tribunal Supremo ante una propuesta de "responsabilidad civil fractal" generada en colaboración con agentes de IA. ¿Es esto derecho o es un juego académico? ¿Tienen estos "descubrimientos" la gravitas, la prudencia, la sabiduría acumulada que esperamos del conocimiento jurídico?

La batalla por la legitimidad no se ganará con argumentos técnicos sobre la sofisticación de los algoritmos o la validez de las simulaciones. Se ganará demostrando que el *Virtual Lab* no destruye la tradición jurídica sino que la revitaliza, que no reemplaza el juicio humano sino que lo amplifica, que no trivializa el derecho sino que lo toma tan en serio que se atreve a reimaginarlo.

Esto requiere *diplomacia epistémica*, pero que podríamos expandir como "traducción jurídico-cultural". Las innovaciones del *Virtual Lab* deben ser presentadas no como rupturas radicales sino como evoluciones naturales de principios jurídicos establecidos. La "personalidad jurídica gradual" para IAs no es una fantasía ciberpunk sino la aplicación contemporánea del mismo ingenio jurídico que creó la persona jurídica corporativa en el siglo XIX.

5.2. Gobernanza de datos y riesgos sistémicos

Aquí llegamos a una paradoja fundamental del *Virtual Lab*: el concepto del "dato propietario educativo" y su manifestación específica como "dato propietario jurídico". En la economía del conocimiento algorítmico, donde los modelos de IA pueden replicar instantáneamente cualquier información pública, el valor se desplaza hacia lo genuinamente original e irreplicable: las trazas únicas del proceso de descubrimiento.

Cada interacción en el *Virtual Lab* genera datos de valor extraordinario. Cuando un estudiante formula una pregunta genuinamente original sobre la personalidad jurídica de las DAOs, cuando un profesor diseña un escenario de simulación que revela una laguna normativa previamente invisible, cuando un grupo de estudiantes itera hacia una solución innovadora a través de múltiples fracasos productivos —todos estos momentos generan *datos propietarios jurídico-educativos* que no existen en ningún corpus público.

Estos datos tienen una triple dimensión de valor:

Valor pedagógico inmediato: Las trazas de comportamiento revelan patrones de aprendizaje únicos. ¿Cómo navegan los estudiantes más creativos

entre doctrina establecida e innovación? ¿Qué secuencias de provocación del profesor-orquestador generan los *breakthrough* más significativos? Estos patrones, agregados y analizados, permiten refinar continuamente la arquitectura pedagógica del *Virtual Lab*.

Valor institucional estratégico: Una facultad que acumula años de datos sobre cómo sus estudiantes generan innovación jurídica posee un activo cognitivo único. Puede identificar talentos excepcionales tempranamente, diseñar programas ultra-personalizados, y demostrar empíricamente su capacidad de formar juristas innovadores. En un mundo donde la educación jurídica tradicional se vuelve un *commodity*, estos datos propietarios se convierten en la ventaja competitiva definitiva.

Valor epistemológico colectivo: Agregados a gran escala, estos datos revelan cómo emerge el conocimiento jurídico original. ¿Existen patrones universales en la innovación normativa? ¿Cómo difieren los procesos de descubrimiento entre tradiciones jurídicas (*common law* vs. derecho continental)? Esta meta-información sobre la génesis del pensamiento jurídico constituye un nuevo campo de conocimiento: una *ciencia del descubrimiento jurídico*.

Pero este valor conlleva vulnerabilidades éticas profundas. Si cada pregunta, cada hipótesis, cada iteración hacia una solución queda registrada, ¿no estamos creando un panóptico pedagógico? El documento fuente sugiere que los datos se analizan anónimamente, pero la anonimización es frágil en *datasets* ricos. Los patrones de pensamiento jurídico de un estudiante pueden ser tan únicos como su huella digital.

Surge entonces la necesidad de lo que podríamos llamar "gobernanza del dato propietario jurídico-educativo":

- *Propiedad compartida:* Los datos no pertenecen solo a la institución sino también a quienes los generan. Los estudiantes deben tener derechos sobre las trazas de su proceso creativo.

- *Transparencia algorítmica:* Los estudiantes y profesores deben entender cómo sus datos alimentan el sistema y qué inferencias se extraen de ellos.

- *Derecho al olvido selectivo:* Debe existir la posibilidad de que ciertas exploraciones —especialmente aquellas que resultaron en fracasos— puedan ser eliminadas del registro si el estudiante lo solicita.

- *Beneficio distributivo:* Si estos datos generan valor económico (por ejemplo, si se licencian a otras instituciones o se usan para desarrollar productos educativos comerciales), los beneficios deben compartirse con la comunidad que los generó.

El modelo del *Virtual Lab* describe un ciclo virtuoso: estudiantes interactúan → profesores refinan escenarios → comportamientos se registran → sistema mejora → estudiantes se benefician. Es una visión *compelling* de mejora continua, pero debemos examinar también las sombras de este ciclo.

La homogeneización sutil: A medida que el sistema *aprende* qué funciona, puede comenzar a canalizar sutilmente a todos los estudiantes hacia patrones «óptimos» de descubrimiento. La diversidad cognitiva —esas formas idiosincráticas y aparentemente ineficientes de pensar que a veces producen los *insights* más radicales— podría ser inadvertidamente podada en favor de la eficiencia.

La performatividad del dato: Cuando los estudiantes saben que sus interacciones alimentan el sistema, pueden comenzar a «actuar» para el algoritmo. En lugar de explorar genuinamente, podrían intentar generar los patrones que creen que el sistema valora. El dato propietario se contamina con performatividad, perdiendo su autenticidad.

La dependencia sistémica: A medida que el *Virtual Lab* se vuelve más sofisticado gracias a los datos acumulados, la brecha entre quienes tienen acceso a él y quienes no se amplifica exponencialmente. Las facultades con *Virtual Lab*s maduros, alimentados por años de datos propietarios, podrían volverse inalcanzables para las instituciones que recién comienzan.

Abordar estos desafíos no es un obstáculo, sino una parte integral del diseño de esta nueva pedagogía. Es precisamente al confrontarlos que podemos redefinir el propósito de la educación jurídica para el siglo XXI.

La reflexión desarrollada a lo largo de este capítulo nos conduce a una constatación decisiva: la educación jurídica ya no puede sostenerse en el paradigma tradicional de transmisión doctrinal y repetición de precedentes. La irrupción de la inteligencia artificial ha alterado de raíz el contrato implícito entre las facultades de derecho y la sociedad, obligándonos a repensar no solo los métodos de enseñanza, sino el sentido mismo de formar juristas en el siglo XXI.

El *Virtual Lab* jurídico se presenta, en este contexto, como una respuesta más profunda que la simple modernización tecnológica. No es una herramienta adicional ni un recurso metodológico accesorio, sino la encarnación de un *nuevo contrato educativo:* la facultad de derecho deja de ser un repositorio de respuestas heredadas y se convierte en un laboratorio de preguntas inéditas, un espacio donde se ensayan las arquitecturas normativas que el porvenir demanda.

De esta transformación emerge una nueva figura de jurista, ya no concebido como custodio del pasado, sino como *arquitecto de futuros normativos.* El entrenamiento en el ciclo de descubrimiento jurídico, la interac-

ción con agentes especializados y la guía del profesor-orquestador forman deliberadamente las competencias necesarias para este papel: una *visión sistémica* que conecta áreas dispares del ordenamiento; una *imaginación normativa* capaz de generar categorías inéditas; un *foresight ético* que anticipa consecuencias humanas y sociales; y una *resiliencia epistemológica* que abraza la incertidumbre como ocasión de creación.

Así preparados, los graduados no temerán a la inteligencia artificial, porque habrán aprendido a colaborar con ella en la frontera misma del conocimiento. No se verán reducidos a técnicos desplazados por la automatización, sino reconocidos por lo irreductiblemente humano que aportan: la formulación de preguntas significativas, el juicio prudencial y la imaginación moral.

La implementación de *Virtual Lab*s no constituye una opción estratégica, sino un *imperativo institucional* para cualquier facultad de derecho que aspire a permanecer relevante. Asegura la empleabilidad de los futuros juristas, revitaliza la función social de la universidad y preserva el carácter humanista del derecho como herramienta fundamental de la humanidad para orientarse en un mundo incierto. La cuestión ya no es solamente qué es el derecho, sino en qué debemos transformarlo. Y es precisamente en el laboratorio jurídico donde podemos atrevernos a formular y ensayar esas respuestas.

Bibliografía

Allbon, E., & Perry, M. (Eds.). (2023). *Design in legal education*. Routledge.

Bahar, K., & Maker, C. J. (2025). A new perspective on scoring children's originality in educational assessment. *Frontiers in Psychology, 16*, 1545396. https://doi.org/10.3389/fpsyg.2025.1545396

Bizzuka Legal. (2025). *Ethical issues of using AI in legal practice: Professional responsibility in the age of automation*. Legal Technology Review.

Botein, M. (1974). Simulation and roleplaying in administrative law. *Journal of Legal Education, 27*(3), 364–389.

Burton, K. (2015). *Measuring and enhancing the authenticity of an examination in legal education*. University of Southern California Research.

Casey, T. (2014). Reflective practice in legal education: The stages of reflection. *New York City Law Review, 20*(2), 327–389.

Centro de Estudios Políticos y Constitucionales. (2025). *Master's degree in constitutional law: Case-based learning methodology*. CEPC Academic Programs.

Columbia Law School. (2025). *Constitutional law program: Experiential learning and clinical applications*. Columbia Law Academic Programs.

Corrada, R. L. (2022). Teaching a hybrid administrative law simulation class using 'Jurassic Park'. *Journal of Legal Education, 71*(2), 245–290.

Darling-Hammond, S. (2015). Creating wise classrooms to empower diverse law students: Lessons in pedagogy from transformative law professors. *Berkeley Journal of African-American Law & Policy, 17*(1), 47–112.

García-López, I. M., et al. (2025). Ethical and regulatory challenges of Generative AI in education: A comprehensive analysis. *Frontiers in Education, 10*, 1565938. https://www.google.com/search?q=https://doi.org/10.3389/feduc.2025.1565938

Ghafarollahi, A., & Buehler, M. J. (2024). *SciAgents: Automating scientific discovery through multi-agent intelligent graph reasoning*. Nature Machine Intelligence.

Ghareeb, A. E., Chang, B., Mitchener, L., et al. (2025). *Robin: A multi-agent system for automating scientific discovery*. arXiv.

Giddens, T., & Siliquini-Cinelli, L. (Eds.). (2023). *Biopolitics and resistance in legal education*. Routledge.

Goswami, D. P. (2025). *Revolutionizing legal education: The role of artificial intelligence in shaping the future of law teaching and learning*. SSRN.

Grimes, R. (2015). Problem-based learning and legal education – a case study in integrated experiential study. *REDU. Revista de Docencia Universitaria, 13*(1), 361–375.

Harvard Law School. (2025). *The case study teaching method in legal education*. Harvard Law School Academic Resources.

IAALS. (2024). *Administrative law: Dinosaur Park simulation - Innovative teaching methods*. Institute for the Advancement of the American Legal System.

IE Law School. (2025). *Future-proofing legal education in the age of AI*. Lawtomation Educational Research.

Kam, L., Ruyters, M., Coburn, C., & Toohey, M. (2012). Get real! A case study of authentic learning activities in legal education. *Murdoch University Law Review, 19*(2), 17–35.

Kleeger, J. (2019). Implementing a meaningful & effective legal education reform. *Journal for Clinical Legal Learning & Education, 13*(1), 1–25.

Kong, X., Gao, M., Xu, T., Pan, S., & Wang, S. (2025). Examining human–AI collaboration in hybrid intelligence learning environments. *Nature Humanities & Social Sciences Communications, 12*, 97. https://www.google.com/search?q=https://doi.org/10.1038/s41599-025-05097-z

Law Society UK. (2024). *How AI is reshaping the future of legal practice*. Legal Practice Innovation Review.

Maharg, P., & Nicol, E. (2024). *Simulation and technology in legal education: A systematic review and future research programme*. Legal Education Review.

McEntee, K. P., & Koenig, P. (2012). The crisis in legal education: Dabbling in disaster planning. *Michigan Journal of Law Reform, 46*(1), 225–270.

Morozov, E. (2015). *La locura del solucionismo tecnológico (Ensayos)*. Katz-Clave Intelectual.

Pascoe, D. (2017). How can legal education speak to the discovery-based learning revolution? *Legal Education Review, 27*(1), 89–112.

Peng, J. (2023). Construction and application of legal education platform in colleges and universities under the background of governing schools by law in Internet Plus. *Atlantis Press*. https://www.google.com/search?q=https://doi.org/10.2991/978-94-6463-230-9_25

Polden, D. J. (2017). Leading institutional change in legal education: Overcoming faculty and administrative resistance. *Tennessee Law Review, 84*(4), 897–936.

Prakash, G. A. (2025). Integrating generative AI into legal education: Transforming teaching methods and student learning. *Legal Technology and Higher Education Journal*.

Pu, Y., Lin, T., & Chen, H. (2025). *PiFlow: Principle-aware scientific discovery with multi-agent collaboration*. arXiv.

Rhode, D. L. (2013). Legal education: Rethinking the problem, reimagining the solutions. *Pepperdine Law Review, 40*(2), 437–472.

Robson, N. A., & Huffman, J. L. (2019). Disruptive leadership in legal education: Institutional resistance and reform challenges. *University of Hawaii Law Review, 22*(3), 563–598.

Siems, M. (2008). Legal originality. *Cambridge Law Journal, 67*(1), 78–108.

Stanford Law School. (2025). *Constitutional Law Center: Innovative pedagogical approaches*. Stanford Law Academic Publications.

Stockholm University. (2020). *Ethical and legal challenges in relationship to AI-driven practices in higher education*. Research Project Report.

Su, H., et al. (2025). *Improved scientific idea generation by a LLM-based multi-agent system with scientific collaboration*. Proceedings of the ACL.

The Colleges of Law. (2025). *7 ways artificial intelligence is already changing law school and legal careers*. Legal Education Innovation Report.

Toussaint, E. (2022). The miseducation of public citizens. *Georgetown Journal on Poverty Law & Policy, 29*, 287–325.

University of Toronto. (2004). *The future of law lab: Technology, innovation, and legal practice evolution*. Future of Law Research Initiative.

Verdú Baeza, J. (2023). Case studies in legal document drafting: Simulation-based learning approaches. In *Modernising legal education*.

Vlachopoulos, D. (2024). A systematic literature review on authentic assessment in higher education: Trends, challenges, and future directions. *Learning and Individual Differences, 104*, 102297. https://www.google.com/search?q=https://doi.org/10.1016/j.lindif.2024.102297

Wilkins, D. (2024). *How AI may transform the legal profession in 2024*. Harvard Law School Today.

World Lawyers Forum. (2025). *Future of virtual law firms: Remote practice rising*. Legal Industry Analysis.

Young, K. M. (2021). What the access to justice crisis means for legal education. *George Washington Law Review, 89*(3), 811–876.

Zou, J., Swanson, K., Wu, W., Bulaong, N. L., & Pak, J. E. (2025). The *Virtual Lab* of AI agents designs new SARS-CoV-2 nanobodies with experimental validation. *Nature, 634*, 312–318.

La inteligencia artificial como clave para recuperar la legitimidad de la justicia argentina

Por Lara Morquecho [*]

Resumen: La legitimidad del sistema de justicia atraviesa una profunda crisis de confianza, especialmente en Iberoamérica y en Argentina, donde los índices de credibilidad resultan alarmantemente bajos. En este contexto, la inteligencia artificial (IA) emerge como una herramienta estratégica para repensar el derecho procesal y recuperar la eficacia, eficiencia y transparencia de la función jurisdiccional. Inspirado en el legado humanista de Augusto M. Morello, este trabajo propone articular la aplicación de la IA bajo un enfoque centrado en la persona, con normas procesales que operen como barreras de contención en resguardo de los derechos humanos. La propuesta se orienta a consolidar un modelo de *Human-Centered AI* en la justicia, que permita optimizar los tiempos procesales, fortalecer la tutela judicial efectiva y garantizar un debido proceso en plazo razonable, sin sustituir la valoración judicial. La IA puede convertirse en una oportunidad histórica para reconstruir la confianza ciudadana en la justicia, siempre que se la implemente con principios claros de transparencia, oficiosidad, supervisión humana y seguridad técnica.

Palabras clave: Inteligencia Artificial; legitimidad de la justicia; derecho procesal; acceso a la justicia; oficiosidad; transparencia; supervisión humana; eficiencia procesal; tutela judicial efectiva.

Abstract: The legitimacy of the justice system is undergoing a profound crisis of trust, particularly in Latin America and Argentina, where confidence levels are alarmingly low. In this context, artificial intelligence (AI) emerges as a strategic tool to rethink procedural law and restore the efficiency, effectiveness, and transparency of judicial functions. Inspired by the humanist legacy of Augusto M. Morello, this paper proposes to articulate

(*) Profesora de Derecho Procesal en la Facultad de Ciencias Sociales y Jurídicas (UNLP) y en Elementos de Derecho Procesal la Facultad de Derecho de la Universidad de Buenos Aires (UBA). Correo de contacto: laramorquecho@gmail.com. Abogada (UNLP). Especialista en Derecho Procesal Profundizado (UNA). Maestranda en Derecho Procesal (UNLP). Diplomada en Derecho Universidad Austral. Codirectora del Laboratorio de Inteligencia Artificial de la Asociación Argentina de Derecho Procesal. Investigadora de la Universidad Nacional de La Plata sobre Modernización de la Justicia.

the application of AI under a human-centered approach, with procedural rules acting as safeguards for the protection of human rights. The proposal aims to consolidate a model of *Human-Centered AI* in the justice system, enabling the optimization of procedural timelines, strengthening effective judicial protection, and ensuring due process within a reasonable time, without replacing judicial assessment. Ultimately, AI can become a historical opportunity to rebuild public trust in justice, provided it is implemented under clear principles of transparency, judicial initiative, human oversight, and technical security.

Keywords: Artificial Intelligence; legitimacy of justice; procedural law; access to justice; judicial initiative; transparency; human oversight; procedural efficiency; effective judicial protection.

1. INTRODUCCIÓN

En su libro de ensayos *Otras inquisiciones*, Jorge Luis Borges nos recuerda que todo conocimiento es una relectura infinita del pasado y nos desafía con una verdad inspiradora: *'Nada se edifica sobre la piedra, todo sobre la arena, pero nuestro deber es edificar como si fuera piedra'*. Una invitación a construir, aún en la incertidumbre.

Estas palabras condensan con precisión el desafío que hoy enfrenta la justicia: edificar certezas en un tiempo de cambios vertiginosos, donde las bases parecen moverse constantemente. En este breve trabajo, intentaré esbozar algunas propuestas derivadas de uno de los mayores desafíos de nuestra historia: la incorporación de la inteligencia artificial en el sistema de justicia.

El punto de partida será construir desde la base de *lo hecho*, repensando —a la luz de las nuevas tecnologías y, en particular, de la IA— las ideas de Augusto M. Morello sobre el derecho procesal moderno, con la intención de volver a los clásicos [(1)] como una forma de homenaje y, al mismo tiempo, de renovación.

El profesor citaba entre sus enseñanzas como mensaje de esperanza para proyectar la justicia hacia el futuro la importancia de *porfiar por un nuevo equilibrio entre la tradición y cambios sin quedarnos en el ayer.* Esto es lo que, modestamente, intentaré realizar en esta propuesta.

Morello fue un pionero en advertir la necesidad de incorporar la dimensión humana en el derecho procesal y en la justicia, postulando de manera

(1) Entendidos en el sentido que lo explica Calvino en Por qué leer los clásicos: "como los libros que nunca terminan de decir lo que tienen que decir" y que no puede serte indiferente y sirve para definirte a ti mismo en relación y quizás en contraste con él.

constante la humanización del quehacer jurisdiccional. Dicho ideario, impregnado de un fuerte componente ético y solidarista, se materializó en la consolidación de un derecho procesal de orientación social, cuyo eje radica en la tutela judicial efectiva de la persona humana, particularmente de los sectores más vulnerables, en línea con el movimiento universal por el acceso a la justicia iniciado por Mauro Cappelletti.

En su legado, resulta especialmente fecunda la vinculación del derecho procesal con la evolución del derecho constitucional y convencional y estos son los ejes fundamentales que tendremos como premisas para la que considero —debida— aplicación de la IA en nuestro sistema de justicia. Intentaré, modestamente, recordar al profesor en cuya obra y trayectoria se advierte con nitidez un humanismo integral [(2)].

2. La legitimidad de la justicia en Iberoamérica: ¿podemos recuperarla con la aplicación de IA?

El tema que nos ocupa reviste un indudable interés tanto desde la óptica académica como práctica y de él asoman múltiples aristas que cabría analizar. Partiendo de la premisa de la necesidad de tomar acción ante la situación crítica de la justicia argentina especialmente [(3)].

En palabras de Berizonce, la concepción del servicio de justicia como modelo de eficacia y tempestividad estructural implica una visión más abarcadora del proceso civil, en el sentido del equilibrio en la protección de todos los intereses en conflicto..

La eficacia implica alcanzar el resultado pretendido, la eficiencia es más que eso. Cuando hablamos de eficiencia pensamos en la obtención de los resultados, pero con un elemento más que la diferencia de su hermana gemela: se habrá alcanzado la meta utilizando la mejor manera. Sin dudas estos conceptos se vinculan con el análisis económico del derecho y el derecho procesal en momentos históricos estratégicos como el presente debe conocer cuáles son los principales efectos y consecuencias de sostener un proceso judicial que continúe perteneciendo a las ideas del siglo XX o que avance junto con las nuevas tecnologías.

---

(2) Morello no fue solamente un exégeta del derecho vigente, sino un jurista creador, que supo elaborar categorías de proyección duradera en el pensamiento procesal. Entre sus aportes más significativos se encuentra la difusión y sistematización de la denominada justicia de acompañamiento o de protección, y la reivindicación de la jurisdicción protectora como herramienta interpretativa y operativa en el derecho argentino y en la jurisprudencia contemporánea.

(3) Índice de Confianza en la Justicia - Primer Semestre 2025 La Escuela de Derecho de la Universidad Torcuato Di Tella y el Foro de estudios sobre la Administración de Justicia (Fores), presentaron el Índice de Confianza en la Justicia (ICJ) correspondiente al Primer Semestre de 2025. Para ver el documento podrá hacer click aquí: https://www.utdt.edu/download.php?fname=_175035767014177400.pdf

Para comenzar el análisis, es dable atender a la distinción de la "eficiencia procesal" y de la "función del proceso". La finalidad de la primera debe ser que el sistema de administración estatal de Justicia tramite los casos concretando el valor justicia con máxima eficiencia y la de la segunda, obedece a la necesidad funcional de dirimir el conflicto jurídico. Coincidimos con que la práctica tribunalicia se ha concentrado esencialmente en la materialización de la función y se ha descuidado la eficiencia.

Índices actuales demuestran que, en todo el mundo, los poderes legislativo y judicial y la sociedad civil han perdido terreno en el control del poder ejecutivo. Aunque estas y otras tendencias autoritarias se habían ralentizado un poco el año pasado, se expandieron en 2024. Argentina se encuentra entre el 59% de los países en los que se redujeron los límites al poder gubernamental en 2023 y entre el 63% de los países en los que disminuyeron los derechos fundamentales en el 2023.

Estos parámetros inciden en la percepción del servicio de justicia de la sociedad existiendo un alto nivel de desconfianza. El índice de confianza en la Justicia en Argentina en un estudio efectuado por la Escuela de Derecho de la Universidad Torcuato Di Tella y el Foro de estudios sobre la Administración de Justicia (Fores) correspondiente a noviembre de 2023 arroja que el 91% de las personas entrevistadas percibe a la Justicia "nada" o "poco confiable" en términos de capacidad y eficiencia observando que la respuesta "nada confiable" coincide en niveles de educación más altos.Para evidenciarlo resulta interesante analizar indicadores estadísticos de los últimos tiempos demuestran un grave deterioro en el estado del derecho a nivel mundial. En las estadísticas de *World Justice Project* del año 2024 el *Rule of Law Index* revela que el estado de derecho ha descendido en todo el mundo por séptimo año consecutivo73 y en 2023 los derechos humanos disminuyeron en 3 de cada 4 países desde 2016. Los números son alarmantes, a ello se le adhiere que de la vista de países como Argentina, Chile, Uruguay y España podemos observar que desde el 2020, las estadísticas han descendido de forma trágica en comparación con otros años.

La confianza es esencial. Los operadores jurídicos debemos adaptarnos a los cambios que la realidad impone y principalmente garantizar el plazo razonable en los procesos. Si la inteligencia artificial ha logrado una presencia cotidiana en nuestras vidas, facilitando nuestras operaciones bancarias, nuestra forma de vincularnos con el mundo, la manera en la que hacemos las compras en el mercado e incluso en muchos países las formas en las que se hacen los reclamos al estado, ¿por qué nos resistimos en incorporar la inteligencia artificial en el ámbito de la justicia?

La IA genera inquietudes y dudas inagotables, sin embargo, me atrevo a afirmar que gran parte de esta desconfianza proviene del temor al error y a

lo desconocido: aún no hemos podido definir precisamente cómo controlar esta "ola imparable".

3. Las normas procesales como barrera de contención

La revolución traída por la IA que nos genera la necesidad de repensar el proceso como *framework* o barrera de contención. Estas normas procesales de "protección", vistas desde una perspectiva constitucional convencional en resguardo de los derechos humanos, lograrán reafirmar la confianza en la justicia al promover mayor celeridad —eficiencia del servicio— con la debida seguridad y protección para los usuarios del sistema.

Si bien debemos considerar a la tecnología nuestra aliada, la contención resulta necesaria para que exista un equilibrio de poder entre los humanos y estas herramientas, abarcando la regulación, seguridad técnica, nuevos modelos de gobernanza y propiedad, de responsabilidad y de transparencia. Se considera el primer paso crítico para construir el futuro [(4)].

La propuesta implica trazar un marco ético de aplicación de la IA de forma responsable y segura en los procesos judiciales, equilibrando los beneficios de precisión y celeridad de esta tecnología con el resguardo de los derechos de las personas involucradas. Necesitamos una IA confiable. La confianza ha de ser la piedra angular de las sociedades y para ello no podemos pensar en la sustitución del ser humano sino en la construcción de una herramienta a nuestro servicio.

En un estudio efectuado en el marco del Centro de Estudios de Justicia de las Américas ya se comprueban los impactos positivos en la mejora de los procesos y en la toma de decisiones de la justicia, concretamente en simplificar y automatizar tareas; reducir errores y generar documentos de manera asistida.

Sin embargo, debemos advertir que la mayoría de los cambios hasta ahora, han involucrado tecnología con un *mindset* procesal que replica

---

(4) Suleyman (CEO de Microsoft, cofundador de DeepMind, una empresa británica adquirida por Google en 2014) explica cómo algunos países reaccionan para contener riesgos catastróficos en forma de autoritarismo de corte tecnológico para ralentizar la expansión de estos nuevos poderes, lo que requerirá enormes niveles de vigilancia, además de intrusiones en nuestra vida privada. Así sucedió cuando la imprenta se extendió por Europa en el siglo XV y el imperio otomano decidió prohibirla. Sin embargo, prohibir el desarrollo tecnológico es un riesgo en sí mismo pues las sociedades estancadas en materia tecnológica son tradicionalmente inestables. Resistirse o rechazar la evolución de estas tecnologías de la información y la comunicación en gran medida explicó las diferencias de alfabetización entre los habitantes de la vieja Europa, de las colonias españolas e inglesas en América, por un lado, y del imperio otomano, el imperio mongol de la India y la dinastía Ming en China por otro, que continuaron empleando escribas. Hacia fines del siglo XVIII, en Inglaterra las personas tenían un 90% de alfabetización aunque existía una importante brecha de género. En la misma época, los habitantes de Hungría eran prácticamente analfabetos: solo un 5.5% eran capaces de leer y escribir

viejas prácticas. Susskind nos inspira a pensar en que debemos utilizar la tecnología, pero desde una óptica diferente a lo que Nunes denomina como una "revolución de los institutos procesales". En definitiva, las herramientas de la inteligencia artificial pueden ser muy beneficiosas, pero depende de los seres humanos —en este caso de los operadores de justicia— la manera en la implementaremos.

4. UNA NUEVA HERRAMIENTA PARA POTENCIAR LA OFICIOSIDAD DEL JUEZ

Couture hace varias décadas advertía que el proceso civil no puede ser concebido como un mero duelo privado entre partes, frente a un juez pasivo que solo certifica el resultado de la contienda. Por el contrario, el Estado es también sujeto interesado en el litigio. Mientras los particulares persiguen la satisfacción de intereses propios, el Estado procura que la función jurisdiccional se cumpla conforme a los principios y garantías consagrados en la Constitución.

En el mismo sentido, Morello se preguntaba retóricamente si resultaba posible la convivencia entre la democracia con la intolerable desigualdad. Pensaba como uno de los objetos del derecho procesal era acortar la brecha de las desigualdades de hecho, que menoscaban la dimensión social de la justicia.

De este modo, los maestros desmantelan la posibilidad de un juez espectador neutral, lo proyectan como director del proceso quien debe resguardar la eficacia y justicia de aquel. A partir de allí, se abre paso a un nuevo paradigma procesal: el del "activismo judicial", entendido como el ejercicio legítimo y necesario de los poderes-deberes que le confiere la Constitución para tutelar derechos y garantizar la efectividad del orden jurídico.

Las causas judiciales —sin importar la materia de que se trate— se tramitan con los recursos públicos del Estado. Desde esta óptica es que se propone pensar al impulso procesal de oficio como garante de la celeridad procesal y eficiencia del servicio. En los casos de vulnerabilidad o desigualdad procesal entre las partes, el deber del juez se ve reforzado para garantizar el cumplimiento y respeto efectivo de esos derechos, lo que resulta de un mandato constitucional específicamente en los artículos 75 inciso 23 de la Constitución Nacional.

No caben dudas que para lograr recuperar la confianza en los justiciables, es necesaria una reforma del proceso, especialmente considerando que del último Índice de Confianza en la Justicia (2025) el 40% de las personas consultadas respondió que no recurriría a la Justicia debido a su mal funcionamiento o lentitud en la justicia en mayo de 2025 y un 40% de los encuestados considera que los cambios más urgentes a fin de mejorar la confianza en la justicia son: menores demoras en los juicios.

Siguiendo los lineamientos del activismo judicial, mi propuesta consiste en que el principio de oficiosidad —junto con garantías tales como seguridad, transparencia, igualdad— sea un aliciente para aplicar IA, de forma contenida, mejorando la eficiencia de los procesos y así, los plazos procesales.

5. La IA como herramienta y parte de un sistema procesal

La clave será posicionar a la IA como parte del sistema procesal, como una herramienta más que facilite la oficiosidad del juez, logrando así resguardar debidamente el debido proceso en plazo razonable y la tutela judicial efectiva, siempre desde el concepto de *Human-Centered AI* es decir la inteligencia artificial centrada en el ser humano [5]. Una versión actualizada de los principios procesales desde esta óptica garantizará una aplicación responsable, segura y transparente de la IA en la justicia.

Para su buen funcionamiento, resultará imprescindible complementar el principio de oficiosidad con dos premisas fundamentales que surgen de los propios desafíos del sistema: el principio de transparencia como elemento basal del sistema democrático y la supervisión humana en resguardo de la seguridad jurídica y la confianza del sistema.

Morello planteaba que *la democracia es un proceso y no un producto y que debemos recrear día a día nuestras energías y empeño para materializarla en plenitud*. En la realidad digital, donde la democracia vive en un estado de crisis permanente, el control del uso de la tecnología se erige como presupuesto de garantía e igualdad lo que exige un ejercicio de transparencia y de responsabilidad.

Para lograrlo la reforma normativa debe ser integral, acompañada por programación de los sistemas informáticos desde el propio Poder Judicial a través de la programación de los sistemas de IA en forma interdisciplinaria, con debida protección de datos personales y evitando sesgos algorítmicos de empresas privadas. Tampoco puede soslayarse, la necesidad de una ardua alfabetización en IA para los ciudadanos, jueces, abogados y justiciables en promoción de una mayor confianza en el sistema para evitar resistencias.

En definitiva, resulta claro hasta aquí que las certezas sobre las que edificamos el derecho nunca son definitivas. La IA nos ha obligado a repensar el proceso en miras de garantizar un derecho fundamental que al mismo tiempo resulta ser un aspecto central para la democracia argentina: la justicia.

---

(5) Lo que pretende esta disciplina es aumentar las capacidades de los seres humanos y abordar sus necesidades sociales, creando herramientas eficaces para las personas.

6. PROPUESTA *DE LEGE FERENDA*: IMPULSO PROCESAL OFICIOSO CON IA

Una propuesta *de lege ferenda* en la normativa procesal podría ser redactada del siguiente modo: "Impulso procesal oficioso: promovido el proceso, el Tribunal tomará de oficio las medidas necesarias para efectivizar su trámite con la mayor celeridad posible. La IA debe ser utilizada para optimizar la gestión del proceso agilizando los tiempos".

La regulación también debería contener pautas de seguridad y garantías que afiancen la confianza del justiciable. La norma podría continuar pautando que "[t]odo ello será llevado a cabo a través de sistemas seguros y confiables cuya transparencia sea proporcionada por el poder judicial, evitando especialmente afectar la capacidad de las partes para ejercer su derecho de defensa. Se garantizará el cumplimiento de los requisitos esenciales de IA confiable, incluyendo: supervisión humana, seguridad y solidez técnica, y privacidad de datos personales".

Respecto de la supervisión humana, se considera prudente añadir que "[t]odo uso de IA en la justicia deberá contar con control y validación por parte de jueces y autoridades competentes" y, en pos de una mayor legitimación del sistema, debería regularse que estos sean: "resistentes a ataques, garantizando precisión, fiabilidad y privacidad de datos personales" (6).

Este principio debería complementarse con una norma clara que establezca la pronta y eficiente administración de justicia donde podría añadirse que se regule que "el Tribunal y sus auxiliares deberán garantizar la pronta y eficiente administración de justicia, maximizando la utilización de las herramientas tecnológicas confiables proveídas por el poder judicial que permitan optimizar tiempos procesales".

De este modo se promueve la automatización de tareas administrativas mediante IA siempre que no afecte el análisis jurisdiccional ni sustituya la valoración judicial.

Finalmente, el último —pero no menor— sobre acceso a la justicia: "deberá garantizarse promoviendo la inclusión de sistemas de IA en tribunales que agilicen la tramitación del proceso y el dictado de su sentencia en plazo razonable conforme los parámetros interpretados por la CIDH. Todo justiciable tendrá derecho a conocer si en su proceso se está utilizando IA, lo que deberá ser informado al comienzo de las actuaciones pudiendo impugnar cualquier resultado basado en decisiones automatizadas, lo que deberá ser revisado por el Tribunal personalmente. En los casos donde se utilice IA en

(6) Véase propuesta *de lege ferenda* completa de reforma del Código Procesal Modelo para Iberoamérica en ¿Un nuevo comienzo? El Código Procesal Civil Modelo para Iberoamérica en la era de la inteligencia artificial: un enfoque convencional en resguardo de los derechos humanos?

la gestión del proceso, se establecerán mecanismos de *evaluación continua de IA*, garantizando su actualización, supervisión y mejora constante para su correcto funcionamiento en la administración de justicia".

7. Conclusiones

La incorporación de la inteligencia artificial en la justicia no puede pensarse como un simple recurso técnico, sino como una oportunidad histórica para recuperar la legitimidad de un sistema en crisis. En este sentido, el legado humanista de Morello ofrece un marco inspirador: situar siempre a la persona en el centro del proceso, privilegiando la dimensión ética y social del derecho procesal.

La IA, correctamente regulada y enmarcada en principios procesales renovados —oficiosidad, transparencia, supervisión humana y eficiencia— puede transformarse en una herramienta para garantizar la tutela judicial efectiva y el plazo razonable, condiciones que la sociedad actual reclama con urgencia. Sin embargo, el uso responsable de esta tecnología exige tres pilares fundamentales: (i) normas procesales que funcionen como barreras de contención y garanticen los derechos humanos; (ii) sistemas diseñados y controlados por el poder judicial con criterios de seguridad, confiabilidad y rendición de cuentas; y (iii) una alfabetización digital masiva que involucre a jueces, abogados y ciudadanos.

En definitiva, la legitimidad de la justicia se reconstruirá con instituciones capaces de adaptarse a los desafíos tecnológicos reafirmando su esencia democrática. La inteligencia artificial, concebida bajo el paradigma de una *Human-Centered AI*, puede ser la clave para alcanzar un equilibrio renovado entre tradición y cambio, asegurando que la justicia sea al mismo tiempo más cercana, más rápida y más confiable. Esa es la apuesta: transformar la crisis de confianza en una oportunidad de innovación humanista para el derecho procesal del siglo XXI.

Como cierre, volviendo al comienzo de este trabajo, Borges nos recordaba que ningún conocimiento ni construcción humana se edifica sobre cimientos absolutamente firmes, sino sobre la arena del tiempo y de la incertidumbre. Sin embargo, debemos construir como si lo fueran y sostener la esperanza de que la justicia puede ser el pilar que —algún día— afiance la democracia.

Bibliografía

Barbosa Moreira, J. C. (2003). Reformas processuais e poderes do juiz. En *Temas de Direito Processual* (8ª série, pp. 62–63). Saraiva.

Barbosa Moreira, J. C. (2004). O direito em tempos de globalização. En *Temas de Direito Processual* (8ª série, pp. 275 y ss.). Saraiva.

Berizonce, R. O. (2011). Bases para actualizar el Código Modelo Procesal Civil para Iberoamérica. *Civil Procedure Review, 2*(2), 104–129.

Berizonce, R. O. (2019). Técnicas de tutela temprana y sistema de oralidad. *Revista de Derecho Procesal, 2019-1*. Rubinzal-Culzoni.

Berizonce, R. O., & Soto, A. A. (2013). El renacer del proceso sumario a partir de la gestión judicial y las nuevas tecnologías. *La Ley*, 765 y ss.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018). Towards human-centered explainable AI: A survey of user studies for model explanations. *ACM Computing Surveys, 51*(5). https://doi.org/10.1145/3236009

Calamandrei, P. (1953). Crisis de la justicia. En AA. VV., *La crisis del Derecho*. Ediciones Jurídicas Europa-América.

Calvo Soler, R., Rojas, J. A., & Salgado, J. M. (2022). *El proceso articulado* (1ª ed.). Rubinzal-Culzoni.

Carnelutti, F. (1953). La muerte del derecho. En AA. VV., *La crisis del Derecho*. Ediciones Jurídicas Europa-América.

Chayer, H. M., & Marcet, J. P. (2016). *Nueva gestión judicial: oralidad en los procesos civiles* (1ª ed.). Ediciones SAIJ.

Faggiani, V. (2022). El derecho a un proceso con todas las garantías ante los cambios de paradigma de la inteligencia artificial. *Revista Teoría y Realidad Constitucional, 50*. Universidad de Granada.

Fairén Guillén, V., Faria Kuger Thamay, R., & Tesheiner, J. M. R. (2016). *Teoría Geral do Processo*. Forense.

Falcón, E. (2023). La edad de la magia. En R. Arazi (Dir.), *Revista de Derecho Procesal 2023-2: Gestión e innovación en el proceso – II* (1ª ed. rev.). Rubinzal-Culzoni.

Falcón, E. M. (2006). *Tratado de Derecho Procesal Civil y Comercial* (T. I). Rubinzal-Culzoni.

Gelsi Bidart, A., Torello, L., & Véscovi, E. (1990). Hacia la preparación de Códigos Procesales Modelos: Las bases para la reforma. En *Códigos Procesal Civil y Procesal Penal. Modelos para Iberoamérica*. Ministerio de Justicia.

Gómez Fróde, C. (2017). *Manual de Teoría General del Proceso*. Tirant lo Blanch.

Hankovits, F. A. (2019). La interpretación procesal como herramienta de eficiencia de la tutela. Paradigma actual del debido proceso constitucional. *Revista de Derecho Procesal, 2019-1*. Rubinzal-Culzoni.

Maranielo, P. A. (s. f.). El activismo judicial, una herramienta de protección constitucional.

Marinoni, L. G., Pérez Ragone, A., & Núñez Ojeda, R. (2010). *Fundamentos del proceso civil: Hacia una teoría de la adjudicación*. Abeledo Perrot Legal.

McCarthy, J. (2007). The science and engineering of making intelligent machines. Computer Science Department.

McCarthy, J. (2007). What is artificial intelligence? Computer Science Department.

Morello, A. M. (1967). *La reforma procesal civil en Buenos Aires*. Ed. Platense.

Morello, A. M. (1996). *Lo hecho (1940-1996): Una retrospectiva no solo jurídica*. La Plata: Librería Editora Platense.

Navarro Wolkart, E. (2022). Inteligencia artificial e sistemas de justiça. *Thomson Reuters Revista dos Tribunais*.

Nunes, D. (2022). Virada tecnológica no direito processual e etapas do emprego da tecnologia no direito processual: seria possível adaptar o procedimento pela tecnologia? *Revista EJEF, 1*(1).

Oteiza, E. (2018). Punto de vista: MARC/ADR y diversidad de culturas: el ejemplo latinoamericano. En *Justiça multiportas: mediação, conciliação, arbitragem e outros meios de solução adequada para conflitos* (2ª ed., pp. 743–765). JusPodivm.

Palacio, L. E. (2003). *Manual de Derecho Procesal Civil*. Abeledo Perrot.

Suleyman, M. (2025). *The coming wave* (3ª reimpresión). Penguin Random House Grupo Editorial.

Taruffo, M. (2002). Dimensioni transculturali della giustizia civile. En *Sui confini. Scritti sulla giustizia civile*. Il Mulino.

Tavolari Oliveros, R. (2010). Discurso de inauguración de las XXII Jornadas Iberoamericanas de Derecho Procesal. *Revista Ius et Praxis, 16*(2), 385–390.

Valentín, G. (2019). La aplicación de las TIC a las categorías del proceso jurisdiccional: hacia una transformación del sistema procesal. *Revista Uruguaya de Derecho Procesal, 1–2*.

Véscovi, E. (1984). *Teoría general del proceso*. Temis.

Implicaciones éticas y pedagógicas de la inteligencia artificial como herramienta de apoyo en el ámbito educativo

Por Guadalupe Juárez Quezada [*]

Resumen: El presente trabajo analiza las implicaciones éticas y pedagógicas del uso de la inteligencia artificial (IA) en la enseñanza del derecho, especialmente a partir de la irrupción de la IA generativa en las aulas universitarias. Este fenómeno ha transformado la manera en que los docentes y estudiantes producen, asimilan y aplican el conocimiento jurídico. La investigación plantea que la IA puede convertirse en una poderosa herramienta de mediación cognitiva que potencie el aprendizaje activo, la investigación autónoma y el pensamiento crítico. No obstante, advierte los riesgos éticos asociados, como el plagio automatizado, la dependencia cognitiva y la pérdida de la autoría intelectual. A través de un enfoque transdisciplinario que integra la pedagogía, la ética, la filosofía del derecho y la comunicación, el texto propone estrategias institucionales para el uso responsable de la IA, enfatizando la transparencia, la autenticidad, la supervisión humana y la equidad como principios esenciales. Se subraya que el docente debe asumir un rol de guía y curador de información, y las universidades deben promover políticas de alfabetización digital ética y modelos didácticos multimodales. En conclusión, el trabajo sostiene que la IA no debe reemplazar el pensamiento jurídico, sino renovarlo, formando juristas digitales éticos capaces de utilizar la tecnología como medio para fortalecer los valores del derecho, la justicia y los derechos humanos digitales.

Palabras clave: inteligencia artificial; enseñanza del derecho; ética académica; innovación educativa; pedagogía digital.

Abstract: This article examines the ethical and pedagogical implications of integrating Artificial Intelligence (AI) into legal education, focusing on the disruptive emergence of generative AI in university classrooms.

(*) Licenciada en Derecho y doctoranda por la Facultad de Derecho de la Universidad Nacional Autónoma de México (UNAM). Maestría en Administración pública. Profesora de tiempo completo Facultad de Derecho de la UNAM. Editora de la Revista de la Facultad de Derecho de México.

The study argues that AI represents both a transformative learning tool and a source of ethical challenges that question academic integrity and intellectual authenticity. From a transdisciplinary approach that combines pedagogy, communication theory, legal philosophy, and applied ethics, the paper explores how AI can serve as a cognitive mediator to foster active learning, independent research, and critical thinking in law students. The author identifies potential risks —such as academic plagiarism, cognitive dependency, and the erosion of analytical skills— while highlighting the need for institutional strategies that promote responsible AI use. These include digital literacy programs, the creation of ethical governance frameworks, and the implementation of multimodal learning models grounded in transparency, authenticity, human supervision, and equity. The paper concludes that AI should not replace legal reasoning but rather renew it, contributing to the formation of ethical digital jurists capable of using technology to strengthen justice, human rights, and the ethical foundations of law.

Keywords: Artificial Intelligence; legal education; academic ethics; educational innovation; digital pedagogy.

1. INTRODUCCIÓN

Actualmente la irrupción de la inteligencia artificial (IA) ha generado un cambio paradigmático en la manera en que la humanidad produce, organiza y asimila el conocimiento. Las aulas universitarias, particularmente en las disciplinas jurídicas, no han permanecido ajenas a este fenómeno. El avance acelerado de las tecnologías digitales y el uso creciente de plataformas en IA generativa, como Chat GPT, Gemini o Copilot, han obligado a los docentes a iniciar un arduo camino para reconfigurar los métodos tradicionales de enseñanza y las prácticas de investigación jurídica.

La IA, como tecnología disruptiva, plantea una oportunidad sin precedentes para fortalecer los procesos formativos mediante estrategias didácticas innovadoras que fortalezcan el aprendizaje activo, la investigación autónoma y el pensamiento crítico del estudiante. Sin embargo, también abre la puerta a dilemas éticos y pedagógicos que cuestionan la autenticidad del conocimiento, la integridad académica y la responsabilidad del docente ante la automatización del aprendizaje.

Desde esta perspectiva, el propósito central de este artículo es analizar las implicaciones éticas y pedagógicas de la integración de la IA como herramienta de apoyo en la enseñanza del derecho, y proponer autorizar su uso responsable y formativo.

2. Antecedentes históricos de la inteligencia artificial y su vínculo con la educación

A lo largo de la historia podemos observar el profundo cambio que ha experimentado el ser humano a través del avance tecnológico, ese cambio se ha incrementado con el paso del tiempo, y se refleja en tres aspectos fundamentales en la vida del hombre y la mujer: la educación, la comunicación, y ambos se vinculan en la sociedad a través del proceso de la socialización.

De tal manera que la evolución de este ser está condicionada por estos tres factores, educación, comunicación, socialización.

López Durán (2025) sostiene respecto de la educación en su libro *Didáctica y Metodología de la Enseñanza superior I*:

> "La educación sí es una acción y un efecto de formar e instruir, sin embargo, no se queda en ello. También es un conjunto de medios que pretenden desarrollar en los individuos y grupos determinada instrucción u opiniones, sin embargo, es más que eso, en el concepto de educación se debe hacer referencia el origen o la fuente de ella, es decir, son medios desarrollados por la sociedad en su conjunto, o más propiamente dicho por los grupos que integran a la sociedad y que tiene como propósito orientar la conducta de los sujetos (que son los miembros de la comunidad, ya sea diferenciando a los que son del mismo grupo social, o en general de otros grupos sociales) hacia el cumplimiento de determinados valores, asíć como la adopción de determinados conocimientos generales o específicos que les permitirán desenvolverse en ciertos ámbitos sociales". (p. 30)

El desarrollo de la tecnología ha sido paulatino, pero la aparición del Internet es un parteaguas en la comunicación, de tal forma que la socialización se modificó. "El internet es una tecnología joven, menor a los cincuenta años (formalmente nace en 1965 con la primera comunicación, vía telefónica entre dos computadoras remotas, una situada en California y otra en Massachusetts)". (López Durán, 2019, p. 175)

La llegada del internet tuvo un impacto especifico en la comunicación de la sociedad y surgieron nuevas formas de comunicarse por medio de las redes sociales.

Gallardo Cano sostiene (2002):

> "La comunicación es un fenómeno complejo que se da en la sociedad y que es susceptible de ser estudiado específicamente por una rama científica. La calidad de ciencia de un estudio de esta naturaleza encuentra plena justificación a través de las disciplinas filosóficas". (p. 2)

Es importante resaltar el papel fundamental de la filosofía en el proceso de la comunicación y sobre todo en el proceso de enseñanza-aprendizaje, ya que de ello deriva la honestidad académica que tenga un estudiante, por

ejemplo: al elaborar un trabajo de investigación se verá reducido o nulificada la posibilidad de caer en el plagio, posiblemente solo se llegue a coincidencias por una mala praxis en las referencias.

Si los alumnos son conscientes de la importancia de la honestidad académica como un valor filosófico, luego entonces no corren el riesgo de caer en el plagio y ser sancionados por la autoridad universitaria; y lo mismo aplicaría para los académicos, en sus diversas labores, ya sea como académicos, investigadores, dictaminadores, etc.…

Derivado de este vínculo entre comunicación y educación, se deduce que también la cultura se ve influenciada por el avance tecnológico, de tal forma que hoy en día los jóvenes se comunican a través de las redes sociales, y el nivel de comunicación intrapersonal va cayendo en desuso, mientras el nivel de comunicación Mass Media tiene su auge.

En 2025, Trejo Medina señaló:

> "Así surgieron novedosas formas de expresión e interrelación para que las sociedad se exprese, comparta ideas; herramientas como *Gopher, Wide Area Information Services (*WAIS*), los Bulletin Board Systems (*BBS*), la World Wide Web (www*), redes sociales, blogs y foros fueron fundamentales para publicitar el conocimiento explícito de diversos investigadores a principios de los 1980 y 1990, generando movimientos sociales y políticos, que proporcionaron un espacio para la voz ciudadana, e incluso para la disidencia". (p. 37)

El desarrollo de la IA no puede entenderse sin considerar el extenso recorrido histórico que ha tenido. Desde los primeros autómatas de Herón de Alejandría y los silogismos de Aristóteles hasta los modelos matemáticos de Alan Turing y las redes neuronales de McCulloch y Pitts, la IA ha sido una extensión del anhelo del humano por replicar su propia inteligencia.

Si nos situamos en la segunda mitad del siglo XX, sabemos que la IA se consolidó como disciplina científica con la conferencia de Dartmouth (1956), donde John McCarthy acuñó el término. Desde entonces la IA ha transitado por distintas fases: la simbólica, basada en reglas razonamiento lógico; la conexionista, sustentada en redes neuronales artificiales; y la actual etapa generativa, caracterizada por su capacidad de aprender, crear y producir nuevos contenidos a partir de grandes volúmenes de datos (*big data*).

En el ámbito educativo, estos avances se traducen en sistemas de tutoría inteligente, entornos adaptativos de aprendizaje y, recientemente, en asistentes conversacionales capaces de interactuar en lenguaje "natural". La UNESCO (2023) advierte que el uso de la IA en la educación debe orientarse hacia la equidad, la inclusión y el respeto por los derechos humanos digitales, evitando reproducir sesgos o desigualdades estructurales.

Trejo Medina (2025) menciona:

> "La IA está incidiendo y tiene el potencial de transformar el ámbito jurídico de manera significativa, permitiendo obtener mejoras en términos de eficiencia, precisión y pronóstico, alineado e integrando la necesaria y pertinente consideración ética y metodológica". (p. 46)

Lo cual nos lleva a reflexionar en la necesidad de incluir a la IA, como una herramienta regulada de manera ética en el proceso de enseñanza aprendizaje de todas las materias impartidas en el plan de estudios de la Licenciatura de Derecho de cualquier universidad.

3. La inteligencia artificial generativa (IAG): una herramienta disruptiva en la enseñanza del derecho

La inteligencia artificial generativa (IAG) representa a la más reciente evolución de los sistemas de IA. A diferencia de los modelos tradicionales, que analizaban o clasificaban datos, la IAG es capaz de crear textos, imágenes, sonidos o códigos originales mediante redes neuronales profundas (*deep learning*) y modelos de lenguaje masivos (*Large Language Models*, LLM).

En el campo jurídico, esta capacidad de transformar se traduce en nuevas posibilidades de investigación, redacción, de documentos legales, análisis de jurisprudencia y simulación de escenarios normativos; ahora bien, su empleo sin orientación ética puede generar dependencias cognitivas y fomentar el plagio académico, como advierten Gallent, Zapata y Ortego (2023). Enfrentando así un nuevo reto: aprovechar el potencial creativo de la IAG como recurso pedagógico al igual que preservar la información ética, crítica y humana de los estudiantes. En este contexto, la IA no debe entenderse como sustituto del docente, sino como un mediador cognitivo, una herramienta que amplifica las capacidades de análisis, argumentación y reflexión del estudiante.

> "La enseñanza de fundamentos de tecnologías de la información e IA en la licenciatura de derecho es esencial, tanto desde el punto de vista educativo, como para el laboral. En el ámbito educativo, incorporar estos temas en los estudios de derecho ayuda a los futuros abogados a familiarizarse con herramientas tecnológicas que están cambiando la manera en que se práctica la abogacía". (Trejo Medina, 2025, p. 46)

Por lo tanto, es necesario que los académicos de las universidades cuenten con la capacitación necesaria para poder combinar la forma tradicional y la IA en la enseñanza del derecho y actualizar su pedagogía para que los alumnos encuentren un incentivo adicional y estén en posibilidad de explotar al máximo esas capacidades que por perteneces a las generaciones *Y, X, Baby Boomers,* tiene implícitas.

4. APLICACIONES PEDAGÓGICAS DE LA INTELIGENCIA ARTIFICIAL EN LA ENSEÑANZA DEL DERECHO

4.1. Transformación del proceso de enseñanza-aprendizaje

La incorporación de la IA en la enseñanza del derecho no se limita a un cambio tecnológico, también implica una redefinición epistemológica del aprendizaje. Según Simon (1983), el aprendizaje humano y el aprendizaje de máquina comparten una misma lógica: la mejora constante a partir del error; lo cual dentro del aula se traduce como la posibilidad de retroalimentación inmediata, simulaciones jurídicas, análisis automatizados de casos y la personalización de rutas de estudio según las competencias del estudiante.

Los sistemas de análisis predictivo, los *chatbots* de asesoría académica y los generadores de textos jurídicos son herramientas basadas en IA que permiten al docente amplificar su alcance didáctico a la par que monitorea el avance del alumnado y detecta áreas de mejora. Este tipo de recursos puede fortalecer las materias teóricas como *Metodología de la Investigación Jurídica*, y las prácticas, como *Argumentación Jurídica o Derecho Procesal,* al fomentar un aprendizaje más autónomo y reflexivo.

Es complejo hoy en día el proceso enseñanza-aprendizaje por el grado de desarrollo de la IA, ante tal situación los académicos debemos actualizarnos constantemente en el correcto uso y buen aprovechamiento de estas herramientas de innovación tecnológica, no sin antes enseñarles a nuestros alumnos sobre la ética y la honestidad académica.

> "A grado tal se ha desarrollado esta tecnología, que la inteligencia artificial generativa o IAG, ha encontrado un nicho de desarrollo industrial y comercial muy competitivo e impresionante: cada pocos meses, a partir del 2022, han aparecido nuevas APPS o servicios que permiten a los usuarios generar contenidos originales a partir de datos existentes (es decir, de bases de datos ya existentes, cada vez más interconectadas y más nutridas). Lo mismo se trate de textos, audios, videos o imágenes vivas. Las versiones son gratuitas de tales desarrollos ofrecen al usuario respuestas más precisas y coherentes, así como posibilidades de exploración 'inteligente' en la red, generar textos altamente documentados, detección, generación o modificación de imágenes y videos con una apariencia sumamente congruente". (Juárez Quezada, 2025, p. 188)

4.2. Modelos didácticos multimodales

El modelo metodológico propuesto por diversos autores en el contexto de nivel licenciatura —basado en la transdisciplina de Nicolesco (1996)— y

la pedagogía de la complejidad de Morin (2001) promueve la articulación entre lo técnico, lo comunicativo y lo ético.

En este marco, como bien se ha venido mencionando, la IA se convierte en un instrumento de mediación cognitiva: un puente entre el conocimiento disciplinar de derecho y las competencias digitales del siglo XXI.

Entre los modelos multimodales más eficaces se encuentran:

-*El aprendizaje colaborativo con IA*, donde estudiantes y docentes construyen conocimiento junto con la herramienta, validando colectivamente la información generada.

-*El aprendizaje adaptativo*, en el que los sistemas de IA ajustan la dificultad de las actividades según el desempeño individual.

-*La gamificación jurídica*, que permite simular audiencias, debates y procesos legales mediante entornos virtuales.

-*El laboratorio de IA aplicada al derecho*, una propuesta que busca crear espacios universitarios donde se integren la tecnología, la ética y la investigación jurídica.

Cuando estos modelos se implementan bajo principios de transparencia y acompañamiento docente, no sustituyen la enseñanza tradicional, más bien la amplían, dotando así al proceso educativo de dinamismo, creatividad y pertinencia social.

5. Implicaciones éticas de la inteligencia artificial generativa

5.1. El dilema de Ícaro: innovación sin límites o caída ética

Toda tecnología conlleva un doble filo. Así como el fuego que permitió cocinar alimentos y forjar herramientas, también dio origen a la guerra. Si adaptamos la metáfora clásica de Ícaro a la actualidad en el contexto educativo, los estudiantes fascinados por el poder de la IA corren el riesgo de "volar demasiado alto" y perder el sentido del esfuerzo, la reflexión y la autoría intelectual.

En el ámbito universitario, este dilema de manifiesta en el aumento del plagio automatizado, de la dependencia cognitiva y la pérdida de habilidades críticas. Kasneci et al. (2023) advierten que el uso indiscriminado de la IA puede reducir la capacidad de pensamiento analítico al delegar la resolución de problemas complejos a un sistema que solo imita la inteligencia humana.

En la enseñanza del derecho, donde la argumentación lógica y la ponderación ética son esenciales, el uso acrítico de la IA puede erosionar la formación de futuros juristas; es por ello que se debe adoptar un enfoque equilibrado que fomente el uso ético, creativo y formativo de la tecnología sin caer en prohibiciones inflexibles ni en permisividad total.

5.2. Principios de ética académica en el uso de la IA

De acuerdo con la UNESCO y el marco ético de la IA en la educación superior, el uso responsable y consciente de estas herramientas requiere de la observación de cuatro principios:

1. Transparencia: los estudiantes deben reconocer y declarar el uso de IA en sus trabajos académicos.

2. Autenticidad: la IA puede asistir, pero no sustituir, el pensamiento crítico individual

3. Supervisión humana: las decisiones pedagógicas deben seguir dependiendo del juicio docente.

4. Equidad: las instituciones deben garantizar el acceso justo a las tecnologías, evitando brechas digitales.

La aplicación de estos principios implica establecer normas, así como promover una ética digital universitaria en la que el uso de IA se convierta en una oportunidad para fortalecer la integridad académica y no en una amenaza para ella.

5.3. Deontología docente y responsabilidad institucional

El papel del docente en este nuevo escenario es crucial. El profesor ya no es el único depositario del conocimiento, sino un curador de información y guía crítico que orienta el uso ético de las herramientas digitales. De ahí que la formación docente en competencias tecnológicas y éticas sea una prioridad. Pensando en una institución líder, se debe destacar que esta misma tiene el deber de llevar a cabo el desarrollo de políticas que aseguren un equilibrio entre innovación y responsabilidad.

Como advierte Cárdenas (2023), la IA no debe suplantar el criterio humano, sino fortalecerlo mediante la supervisión ética, autorregulación y evaluación crítica de los resultados generados.

En los niveles de conocimiento existe uno denominado de segundo nivel, porque se coloca por encima del nivel de conocimiento científico que está a la cabeza de la clasificación de los niveles de conocimiento; este nivel de conocimiento que está por encima del conocimiento científico, es nada

más y nada menos que el conocimiento filosófico, el cual versa sobre la pertinencia, uso, aplicación del conocimiento científico en aras de considerar que dicho nivel de conocimiento solo es considerado al servicio de la humanidad y para su bienestar; pero si por el contrario ese conocimiento científico será utilizado para dañar al tejido social; luego entonces habrá que ponderar en una escala de valores su utilidad, y la puesta al servicio de la humanidad.

De tal forma que se debería privilegiar al conocimiento filosófico, poniéndolo por encima del conocimiento científico, quien tendrá que evaluar la pertinencia del mismo conocimiento científico.

> "El Nivel Filosófico en palabras de López Durán (2004) se puede fraccionar en dos subniveles principales: uno ético y otro epistemológico. En el primero se analizan las consecuencias personales y sociales de la actividad profesional, a la luz de consideraciones axiológicas. (...) hora se cuestiona la validez axiológica de las normas jurídicas y la actuación de los profesionales del derecho en los diversos ámbitos sociales. La filosofía, en este caso considerá globalmente las consecuencias de lo jurídico, da un paso hacia atrás para considerar al derecho como un todo. (...) El nivel epistemológico se convierte en una especie de consciencia critica de la ciencia jurídica. Somete a procesos de revisión las bases, aparentemente sólidas, de la construcción normativa". (p. 23)

6. Estrategias institucionales para una integración responsable de la IA

6.1. Alfabetización digital ética

Hoy en día las instituciones educativas enfrentan el reto de regular sin censurar. En lugar de prohibir el uso de la IA, se deben establecer marcos normativos flexibles y adaptativos. Chan (2023) propone un modelo de gobernanza universitaria que combine tres dimensiones:

- *Normativa:* definición de criterios claros sobre el uso permitido de IA en tareas, tesis y exámenes.

- *Formativa:* creación de asignaturas o módulos sobre ética de la tecnología.

- *Creativa:* impulso de proyectos interdisciplinarios donde la IA se use como laboratorio de innovación educativa.

Esto podría traducirse en la creación de una comisión de ética digital, encargada de supervisar buenas prácticas y promover la reflexión colectiva sobre el impacto de la IA en la educación jurídica.

7. A MANERA DE CONCLUSIÓN. HACÍA UNA PEDAGOGÍA TRANSDISCIPLINARIA

La IA nos ha obligado a pensar el replantear la estructura del conocimiento. Morin (2001) señala que "el conocimiento pertinente debe situar toda información en su contexto", esto puede interpretarse como que una pedagogía transdisciplinaria permitiría integrar saberes de la información, la comunicación, la sociología y la ética al campo del derecho, dando como resultado, profesionales más completos, críticos y adaptables.

El modelo transdisciplinario propuesto en la UNAM por Juárez Quezada (2024) plantea precisamente esta convergencia: combinar la tecnología, la pedagogía y la ética para construir un modelo de enseñanza multimodal, donde la IA funcione como herramienta de mediación y no de sustitución.

En este contexto, la IA no representa el fin del pensamiento jurídico, sino una oportunidad para renovarlo. La clave está en formar juristas digitales éticos, capaces de usar la tecnología para defender los principios del derecho, la justicia y los derechos humanos digitales.

Pero el trabajo es extenso e involucra a varios agentes como son los académicos, las autoridades institucionales, los alumnos; de tal forma que se piensa en un ideal que sería que estos tres agentes tengan delimitadas las formas de uso de esta herramienta que es la IA, en el proceso de enseñanza aprendizaje, y sobre todo se piensa en que el ideal sea actualizar y renovar constantemente las directrices tanto éticas como las jurídicas que regulen esta actividad. Debido al cambio y transformación de estas tecnologías es necesario estar actualizándose constantemente.

Y finalmente se requiere de una pedagogía transdisciplinaria que englobe lo necesario para el buen manejo de la IA, en la función académica y el proceso enseñanza aprendizaje.

De tal forma que es necesario entender conceptos básicos de ingeniería y su lenguaje, así como el derecho y sus necesidades de regulación en el cambio de las tecnologías, la ciencia de la comunicación y sus nuevas formas de interrelacionarse de los individuos; y derivado de ello la ciencia de la sociología haría su entrada triunfal, ya que es esta la encargada de explicar lo complejo del proceso de socialización; y siendo hoy una prioridad la socialización de los seres humanos, se cumple entonces con la multidisciplinariedad; finalmente la ética como la cereza del pastel haciendo gala en el énfasis de la honestidad académica e intelectual, partiendo de principios axiológicos.

Queda mucho por estudiar y explorar, tanto como se transforma la vida del ser humano en sociedad a raíz de la integración de manera consciente de la IA como herramienta en este caso del proceso de enseñanza-aprendizaje;

tal como queda reflejado en la historia a raíz de una pandemia de COVID en 2020, que nos obligó a cobrar consciencia de nosotros mismos, nuestra fragilidad pero a su vez nuestra resistencia y la capacidad de adaptación a las nuevas formas de interacción, de aprendizaje con las herramientas que contienen IA, y la inclusión de manera consciente de esta en casi todas las labores del hombre y de su vida en sociedad.

Bibliografía

Cárdenas, J. (2023). *Inteligencia artificial, investigación y revisión por pares: escenarios futuros y estrategias de acción*. Revista Española de Sociología, 32(4), 1–15. https://doi.org/10.22325/fes/res.2023.184

Chan, C. K. Y. (2023). *A comprehensive AI policy education framework for university teaching and learning.* International Journal of Educational Technology in Higher Education, 20(1), 1–25. https://doi.org/10.1186/s41239-023-00408-3

Du, H., Liu, G., Niyato, D., Zhang, J., Kang, J., Xiong, Z., Ai, B., & Kim, D. I. (2023). *Generative AI-aided Joint Training-free Secure Semantic Communications via Multi-modal Prompts.* arXiv. http://arxiv.org/abs/2309.02616

Gallardo, C. A. (2002) *Curso de Teorías de la Comunicación*. UNAM.

Gallent, C., Zapata, A., & Ortego, J. L. (2023). *El impacto de la inteligencia artificial generativa en educación superior: una mirada desde la ética y la integridad académica.* RELIEVE, 29(2). https://doi.org/10.30827/relieve.v29i2.29134

Gamow, G. (2007). *Biografía de la física.* Alianza Editorial.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Kutyniok, G., Michaeli, T., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Weller, J., & Kasneci, G. (2023). *ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education.* Learning and Individual Differences, 103(102274). https://doi.org/10.1016/j.lindif.2023.102274

Lara Rosano, F. J. (1992). *La inteligencia artificial y su influencia en las ciencias sociales y las humanidades.* En P. González Casanova e I. Méndez (Eds.), *Matemáticas y Ciencias Sociales* (pp. 305–335). Miguel Ángel Porrúa.

López, D. R. (2019) Esboso de un curso de sociología jurídica.

López D. R. (2008) Lo oculto en la enseñanza del Derecho. México, Porrua.

López D. R. (2025) "Aprendizaje y enseñanza universitaria" en Juárez, Q. G. (coord.) *Didáctica y Metodología de la Enseñanza Superior I* .

López D. R. (2004) Metodología Jurídica, Iure editores, México, p. 23.

Morin, E. (2001). *Los siete saberes necesarios para la educación del futuro.* UNESCO.

Nicolescu, B. (1996). *La transdisciplinariedad: manifiesto.* UNESCO/CIRET.

Simon, H. A. (1983). *Why Should Machines Learn?* En R. S. Michalski, J. G. Carbonell & T. M. Mitchell (Eds.), *Machine Learning: Symbolic Computation* (pp. 25–37). Springer.

Trejo Medina, D. (2024). *Teoría general de la administración. Apuntes para abogados.* Pacholabra Ediciones.

Trejo, M. D. Inteligencia Artificial. Fundamentos para abogados y estudiantes de Derecho. México, Tirant lo blanch, 2025.

UNESCO. (2023). *Marco ético global sobre la inteligencia artificial en la educación.* Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. https://unesdoc.unesco.org

Nuevos desafíos para los profesionales del derecho: ética y responsabilidad en el uso de la inteligencia artificial

Por María Victoria Gisvert (*)

Resumen: El rápido auge de la inteligencia artificial ha generado nuevas oportunidades a nivel global, como, por ejemplo, aumentar la eficiencia laboral mediante la automatización de tareas. Sin embargo, estos rápidos cambios también plantean profundos dilemas éticos, que surgen del potencial que tienen los sistemas basados en IA para reproducir prejuicios y amenazar los derechos humanos, entre otros. La IA plantea cuestiones éticas y de responsabilidad, como el riesgo de sesgos o errores en los sistemas. Quienes utilicen estas herramientas novedosas deben estar preparados para abordar estas cuestiones y garantizar que la IA se utilice de manera responsable y ética. La responsabilidad siempre es del profesional y eso no debe perderse de vista. Además, en el ámbito judicial, donde cada decisión puede causar un perjuicio irreparable, es necesario que los procesos y razonamientos detrás de las conclusiones que arrojan estos sistemas sean trazables y auditables. Esto no solo aumenta la confianza en las decisiones tomadas por ellos, sino que también asegura que estas decisiones puedan ser revisadas y cuestionadas cuando sea pertinente. Omitir lo anterior podría llevar a un escenario donde las decisiones basadas en IA sean incomprensibles e inimpugnables. Ahora bien, frente la falta de regulación al respecto, por un lado, la abogacía organizada delinea una guía de recomendaciones para un uso ético y responsable de la inteligencia artificial, y por otro lado, la justicia parece empezar a tomar medidas, al encontrarse ante supuestos de utilización inadecuada por parte de los profesionales. La clave está entonces en cómo usamos estos sistemas para fortalecer la producción jurídica en los distintos ámbitos de aplicación del derecho, haciéndolo de forma ética y responsable.

(*) Abogada (UNLP), vickygisvert@gmail.com. Especialista en Derecho Procesal Profundizado (Univ. Notarial Argentina), maestranda en Derecho Procesal (UNLP), diplomada en Derecho Procesal Informático (UNLZ). Coordinadora de la Especialización en Litigación y Gestión del Caso (UDE). Profesora de grado y posgrado de Derecho Procesal II (UNLP y UDE). Consejera directiva del Colegio de Abogados de La Plata y de la Facultad de Ciencias Jurídicas y Sociales de la UNLP. Delegada ante la Comisión de Derecho Informático de la FACA.

Palabras clave: Derecho; inteligencia artificial; ética; responsabilidad.

Abstract: The rapid rise of artificial intelligence has created new opportunities globally, such as increasing work efficiency through task automation. However, these rapid changes also pose profound ethical dilemmas, stemming from the potential for AI-based systems to reproduce biases and threaten human rights, among other issues. AI raises ethical and liability issues, such as the risk of bias or errors in systems. Those using these novel tools must be prepared to address these issues and ensure that AI is used responsibly and ethically. Responsibility always rests with the practitioner, and this must not be lost sight of. Furthermore, in the judicial field, where any decision can cause irreparable harm, it is necessary that the processes and reasoning behind the conclusions these systems produce be traceable and auditable. This not only increases confidence in the decisions they make but also ensures that they can be reviewed and challenged when appropriate. Failure to do so could lead to a scenario where AI-based decisions are incomprehensible and beyond appeal. However, in the face of the lack of regulation on this matter, on the one hand, the organized legal profession has outlined a guide of recommendations for the ethical and responsible use of artificial intelligence, and on the other, the justice system appears to be beginning to act, faced with cases of inappropriate use by professionals. The key, then, lies in how we use these systems to strengthen legal production in the various areas of legal application, doing so in an ethical and responsible manner.

Key words: Law; Artificial Intelligence; ethics; responsibility.

1. La automatización en la abogacía

En el contexto actual, los modelos de inteligencia artificial para conversaciones, como *Chat GPT*, pueden ser una herramienta avanzada, aunque no están exentos de proporcionar respuestas y soluciones incorrectas. Además de requerirse una adecuada capacitación para obtener resultados con el uso de la herramienta, se demanda sólidos conocimientos jurídicos para poder detectar y corregir errores u omisiones del sistema.

En este sentido, a pesar del riesgo de automatización de tareas codificables y con bajo nivel de calificación, a medida que se requiere mayor nivel de especialización y pensamiento crítico disminuye el riesgo de reemplazo de la persona humana.

De acuerdo con lo expuesto, se pueden destacar algunas de las implicancias que la utilización de la inteligencia artificial tiene en el ejercicio profesional de la abogacía, cuyo norte parte de lograr una mayor eficiencia y precisión en la gestión a través de la automatización y agilización de tareas que permitirán a los abogados centrarse en funciones más relevantes y es-

tratégicas y que además buscan reducir los errores humanos y optimizar el tiempo dedicado a cada tarea. Entre ellas podemos mencionar:

Análisis de documentos: los abogados podrán analizar documentos legales y extraer información relevante de manera rápida y precisa. Además, se facilitará la generación automática de documentos legales, como contratos y escritos, ahorrando tiempo y esfuerzo en su redacción.

Análisis de datos: la IA puede analizar sentencias judiciales para identificar tendencias o patrones, identificar transacciones fraudulentas en grandes conjuntos de datos y evaluar el riesgo de que ocurra un evento negativo, como un litigio o una quiebra. Desde ya que no todas las tareas que requieren un análisis de datos son susceptibles de ser automatizadas. Es el caso de las tareas que requieren un alto grado de conocimiento legal o que son sensibles al contexto y por ende pueden ser difíciles de automatizar.

Predicción de resultados: la IA podrá ayudar a los abogados a predecir el resultado de un caso basándose en datos históricos y patrones identificados. El incremento de la calidad de la información permitirá tomar mejores y ajustar sus estrategias de acuerdo con las probabilidades de éxito.

Revisión y análisis automático de contratos: la agilización en la revisión y análisis automático de contratos y documentos permitirá a los/as abogados/as centrarse en aspectos más estratégicos.

Creación de alertas inteligentes: la IA puede utilizarse para crear alertas basadas en información pública, como cambios en normativas, leyes o recursos, que pueden ser relevantes para casos concretos.

Chatbots legales y asistentes virtuales: estos programas informáticos son capaces de interactuar con los usuarios en lenguaje natural, lo que significa que pueden responder preguntas y proporcionar asesoramiento legal básico.

Los *chatbots* legales funcionan a través de plataformas de mensajería instantánea como *Facebook*, *Messenger* o *WhatsApp*. El usuario simplemente envía un mensaje al *chatbot* con su pregunta o problema legal y el programa responde automáticamente.

La inteligencia artificial detrás de estos se basa en algoritmos de aprendizaje automático, que les permiten "aprender" a partir de la experiencia. Esto significa que cuanto más se utilicen, más precisos serán sus respuestas.

Para funcionar correctamente, deben estar programados con una base de datos sólida de conocimientos jurídicos. Este conocimiento debe ser actualizado regularmente para asegurar la precisión y relevancia del asesoramiento proporcionado.

Son rápidos, económicos y están disponibles las veinticuatro horas del día, los siete días de la semana. Sin embargo, deben ser considerados como una herramienta complementaria a la asistencia legal tradicional y de ninguna manera un sustituto completo.

Estudio jurídico en la nube: la implementación de los distintos sistemas de presentaciones y notificaciones electrónicas y los paradigmas virtuales, exigen que los profesionales puedan acceder a la información desde cualquier lugar y en cualquier momento.

Con un *software* de gestión para estudio jurídico se pueden desarrollar todas las herramientas de productividad e información en un mismo lugar y con total seguridad. Esto permite la optimización del trabajo y proporciona eficiencia en el desempeño y eficacia en los resultados. Además de contar con toda la información jurídica necesaria para generar las estrategias para cado caso, se puede gestionar la rutina de la oficina con precisión.

Tener un estudio jurídico en la nube hace posible optimizar el trabajo remoto, aumenta la productividad y mejora el servicio ofrecido a los clientes y hace posible la reducción de costos en la empresa. El uso de tecnologías de computación en la nube, como Gmail y Dropbox, sin duda alguna, ya son inherentes a la vida de los profesionales del derecho.

Comparto acá con ustedes un secreto, mi herramienta de gestión favorita es Trello: a través de sus tableros se pueden mantener distintas tareas organizadas, los plazos controlados y los miembros del equipo coordinados. Cada tablero automatiza las tareas y los flujos de trabajo y se puede vincular a otras aplicaciones.

Finalmente, y respecto a los estudios jurídicos en la nube, quisiera resaltar una pequeña pero fundamental acotación que se encuentra vinculada a lo que alojamos en la nube: nada más y nada menos que información y datos de nuestros clientes. Es decir, que, el supuesto consagrado en el artículo 69 de la ley N 5177 que establece que "[e]l estudio profesional es inviolable, en resguardo de la garantía constitucional de la defensa en juicio y del secreto profesional. En caso de allanamiento dispuesto por autoridad competente, deberá darse aviso, bajo pena de nulidad, al Colegio de Abogados del domicilio donde se practique la medida, el cual podrá estar presente durante su realización", es completamente aplicable en resguardo de la información que se almacena en la nube del estudio jurídico, ya que el secreto profesional podría verse comprometido al ingresar en los dispositivos donde se encuentra el acceso a la nube (celulares, computadoras, tablets).

ChatGPT: la aplicación en el ámbito legal fundamentalmente tiene el potencial de agilizar procesos y mejorar la eficiencia, obteniendo beneficios como:

-Análisis rápido y eficiente de grandes volúmenes de información.

-Identificación de patrones y tendencias.

-Optimización de la gestión de documentos legales.

-Asistencia en la redacción de documentos legales.

-Investigación predictiva en base a datos históricos.

-Evaluación de riesgos.

Sin embargo, es fundamental ejercer un control posterior y supervisar el trabajo generado por la plataforma, ya que existen casos documentados de errores graves en el uso de la IA en el campo del derecho.

2. Lineamientos para un uso ético y responsable de la inteligencia artificial

Por su parte, la abogacía organizada, a través de la Comisión de Derecho Informático de la Federación Argentina de Colegios de Abogados, elaboró los "Lineamientos para un uso ético y responsable de la Inteligencia Artificial" con el objetivo de optimizar la práctica profesional y contribuir a la confianza pública.

Estos lineamientos se presentan como una guía de buenas prácticas esencial para la utilización de IA en el ejercicio profesional por parte de la Abogacía resguardando la integridad profesional y la confianza pública en la labor de los profesionales del derecho.

En ese marco, se establece que todo uso de la IA en el ejercicio del derecho debe regirse por los siguientes principios fundamentales:

Principio de debida aptitud tecnológica: el profesional tiene el deber de formarse y de mantenerse actualizado en los avances de la tecnología que utilice. Esto implica comprender los beneficios, riesgos y limitaciones de las herramientas de IA que emplea en su práctica jurídica.

Principio de supervisión humana indelegable: ninguna decisión o presentación judicial puede ser delegada completamente en un modelo de IA. El profesional debe ejercer una supervisión activa y crítica de los resultados producidos por estas herramientas antes de su utilización y presentación ante terceros. La intervención y el juicio humano en la labor jurídica es fundamental e indelegable.

Principio de confidencialidad: la información procesada por modelos de IA debe resguardarse conforme a los estándares de confidencialidad y secreto profesional, según lo establecido por el código de ética de cada juris-

dicción. Siempre que fuera posible, las herramientas empleadas deben configurarse para evitar la utilización de la información introducida por parte de empresas y/o terceros.

Principio de transparencia: el profesional debe poder explicar en términos claros el rol que tuvo la IA en la elaboración de escritos o de estrategias, si así fuera requerido por su cliente o por el tribunal.

Principio de responsabilidad y lealtad procesal reforzada: el profesional debe tener una comprensión global de los sistemas que utilice, sus riesgos (como la posibilidad de generar información falsa, citas jurisprudenciales o doctrinales inexistentes o argumentos falaces) y las medidas de mitigación frente a estos. La lealtad procesal reforzada exige que la información presentada ante la contraparte, un tribunal u organismo público sea veraz y previamente auditada por el profesional.

Asimismo, se recomienda utilizar sistemas de IA para redacción de documentos, búsqueda y análisis de doctrina, síntesis e interacción con documentos complejos o extensos, comparación normativa o de criterios jurisprudenciales, generación de argumentos y realización de tareas administrativas tales como agendamiento de audiencias, gestión de causas, transcripciones, entre otras.

Y no se recomienda la adopción de decisiones jurídicas delegadas en sistemas de IA, no verificación de la política de privacidad y del funcionamiento de la herramienta previo a la utilización, introducción de información sensible y archivos o expedientes sin medidas de anonimización, cita de jurisprudencia o doctrina generada por IA sin verificar su existencia y presentación de escritos jurídicos generados íntegramente con IA sin control humano.

Finalmente, para un uso ético y responsable de los sistemas de IA en el ejercicio profesional de la abogacía se aconseja seleccionar la herramienta más adecuada, teniendo en cuenta los términos y condiciones, y las políticas de privacidad y por otro lado, elegir una modelo de IA generativa creado para tareas jurídicas, ya que estos están entrenados con documentos legales y jurisprudencia, lo que reduce significativamente los riesgos de alucinaciones.

3. LO QUE LA JUSTICIA ANDA DICIENDO…

Frente al uso de la inteligencia artificial como herramienta en la producción jurídica, y sobre todo ante la falta de regulación al respecto, la justicia parece empezar a tomar medidas, al encontrarse ante supuestos de utilización inadecuada por parte de los profesionales intervinientes.

En esta oportunidad me interesa analizar lo resuelto por la Sala I de la Cámara de Apelación en lo Civil y Comercial de Morón en los autos caratulados "Acevedo, Gerardo Gabriel c/ Caceres Mareco, Willian Arsenio y Agrosalta Cooperativa de Seguros Limitada s/ daños y perj. autom. c/ les. o muerte (exc. Estado)".

En el caso mencionado *ut supra*, la parte actora recurrió la sentencia de primera instancia y expresó agravios citando, como es usual, diversos antecedentes jurisprudenciales, incluidas sentencias supuestamente dictadas por el Tribunal del recurso.

En esa línea, los camaristas se encontraron con la cita textual de fallos que no pudieron ser hallados en los libros de registro, por lo que concluyeron que tales citas derivaban, posiblemente, del uso de inteligencia artificial en la confección de sus escritos.

Frente a tal situación, la Cámara resaltó que casi la totalidad del agravio se apoyaba en un caso del propio tribunal que no pudo ser hallado, y destacó que en los tiempos que corren y, a partir del conocido fallo de la justicia norteamericana dictado en "Mata v. Avianca", se ha instalado en los tribunales la cuestión de la evocación de precedentes inexistentes, a partir del (inadecuado) uso de sistemas de inteligencia artificial generativa por parte de los profesionales.

Aunque no fue el único fundamento por el cual los magistrados declararon desierto el recurso, el camarista Gabriel H. Quadri puso especial atención al uso indebido de la IA, y aunque destacó que no comparte la tendencia de imponer sanciones por tal motivo al profesional involucrado, ya que este tipo de conductas no encuadran en ninguna de las sancionables por el art. 45 del Cód. Proc. Civ. y Com. (temeridad o malicia) ni dentro del ámbito del art. 35 inciso 3° del Cód. Proc. Civ. y Com., y porque además, como mucho, se trataría de un error en el uso de tecnologías disruptivas y novedosas, concluyó que en este caso resulta necesario tomar ciertas medidas.

En ese entendimiento, y conforme los deberes que menciona el art. 34, inciso 5 del Cód. Proc. Civ. y Com. especialmente en sus apartados b, d y e) se resolvió "poner en conocimiento del letrado lo sucedido, e indicar que, en el caso de utilizar herramientas de inteligencia artificial generativa en sus presentaciones (porque nada impide que lo haga) es necesario, en todos los casos, revisar, chequear y corroborar que la información incluida en sus planteos responda a fuentes (no solo fallos, sino también normas e incluso citas de doctrina) que existan en la realidad. De este modo, se balancea el aprovechamiento de la tecnología (para potenciar el ejercicio profesional) con su uso ético y responsable".

Además, se dispuso a comunicar la decisión al Colegio de Abogados del Departamento Judicial de Morón, solicitando que proceda a difundir entre sus matriculados, la importancia de verificar las citas empleadas.

De esta forma, se buscó fortalecer una práctica responsable y evitar inconvenientes futuros, promoviendo siempre un uso adecuado y ético de la tecnología en el ámbito jurídico.

CONCLUSIÓN

La inteligencia artificial está transformando el sector legal [1] y brindando numerosas ventajas a los abogados que la implementan en sus procesos y estrategias. A pesar de los desafíos éticos y legales, este avance representa una oportunidad para mejorar la eficiencia, precisión y productividad en la abogacía, así como para adaptarse a un mercado laboral en constante evolución.

La creciente adopción de herramientas como *Chat GPT* por parte de los abogados para la búsqueda de jurisprudencia pone de manifiesto la imperiosa necesidad de que los profesionales del derecho comprendan las limitaciones y riesgos inherentes al uso de estas tecnologías.

La inteligencia artificial generativa, a pesar de su capacidad para procesar y analizar grandes volúmenes de información de manera eficiente, puede generar resultados inexactos o incluso erróneos, comprometiendo así la validez de los documentos legales.

Por lo que, a pesar de los beneficios que la inteligencia artificial puede aportar al agilizar tareas en el ejercicio profesional y proporcionarnos una asistencia valiosa, no debe sustituir el juicio crítico y la supervisión humana, elementos esenciales para asegurar la exactitud y la integridad de las presentaciones judiciales. [2]

Es por ello que resulta crucial que los profesionales del derecho se mantengan debidamente informados, competentes y diligentes, que preserven responsablemente la labor de la abogacía y que utilicen los métodos necesarios para supervisar y validar el uso de la IA de manera adecuada.

En definitiva, es imperativo mantener un equilibrio entre la innovación tecnológica y la supervisión humana para garantizar que las presentaciones

---

(1) Fernández Hernández, Carlos, "La innovación en el sector jurídico va a estar marcada por la Inteligencia Artificial". Diario LA LEY, Nº 70, Sección Legal Management, 17 de marzo de 2023, LA LEY.

(2) Tallarico, Milagros Denise, "Consecuencias jurídicas de las alucinaciones en herramientas de Inteligencia Artificial Generativa: Análisis del caso Roberto Mata vs. Avianca Airlines", Microjuris 16/08/2024, Cita: MJ-DOC-17905-AR||MJD17905.

judiciales sean precisas, imparciales y conformes a los estándares éticos requeridos en el ejercicio profesional.

Bibliografía

Comisión Europea, Libro Blanco sobre la Inteligencia Artificial - Un enfoque europeo orientado a la excelencia y la confianza, Bruselas, 2020.

Chumbita, Sebastián C. De sistemas expertos a usuarios expertos. Bases para un uso responsable de la inteligencia artificial generativa, Diario La Ley 16-11-2023 – Thomson Reuters. Ciudad Autónoma de Buenos Aires, 2023.

Chumbita, Sebastián C. "¿Puede la IA equilibrar la balanza de la Justicia?", Diario PERFIL, 10-2-2024. Ciudad Autónoma de Buenos Aires, 2024.

Corvalán, Juan Gustavo. Inteligencia artificial generativa como ChatGPT: ¿Un nuevo Renacimiento? Diario La Ley 5-6-2023 Cita *online*: TR LALEY AR/DOC/1278/2023.

Corvalán, Juan Gustavo Tratado de Inteligencia Artificial y Derecho – Tomo I, 1a ed. - Ciudad Autónoma de Buenos Aires, La Ley, 2021.

Corvalán, Montenegro, Papini, Puig, Pitrau, Quartesan, Sringhini, Protocolo sobre Gobernanza Algoritmica. Trazabilidad, explicabilidad, interpretabilidad y despliegue, publicado en Tratado de Inteligencia Artificial y Derecho, Thomson Reuters, Ciudad Autónoma de Buenos Aires, 2021.

Faliero, Johanna Caterina. Los desafíos jurídicos del Big Data: Tensiones de derechos entre la parametrización analítica, la toma automatizada de decisiones, el targetting y el perfilamiento. Artículo en Suplemento Especial LegalTech II 2019, Editorial LA LEY, Bs. As., Argentina. Año 2019. P. 71-78.

Gisvert, María Victoria, "Ejercicio Profesional en clave digital", El Derecho – Derecho, Innovación y Desarrollo Sustentable, Número 19 – Junio 2024.

Gisvert, María Victoria, "Sesgos de Género en la Inteligencia Artificial: la lucha feminista en la cuarta revolución industrial", Anuario: Principales actividades de la Facultad de Derecho y Ciencias Sociales de la Universidad del Este. 1ª ed. – Librería Editora Platense, La Plata, 2022.

Salierno, Karina V. - Bielli, Gastón E., IA generativa en la praxis legal: responsabilidad profesional y doble control en el uso de LLMs, LL 29/08/2025, 10, accesible desde este enlace: https://escribaniasalierno.com.ar/2025/08/29/inteligencia-artificial-responsabilidad-profesional/).

Tallarico, Milagros Denise, Consideraciones sobre el uso de inteligencia artificial en la búsqueda de jurisprudencia. Editorial Rubinzal Culzoni, 30/07/2024 Cita: RC D 446/2024

La inteligencia artificial para una estrategia probatoria

Por Marcela Tovar Toledo (*)

Resumen: A menudo, el mundo jurídico ha expresado divergentemente que el uso de la IA suplirá el ejercicio personal y hasta ahora indelegable del derecho. No obstante, oponeros a este ineludible cambio, constituirá en algún momento, no solo la constante desactualización de conceptos, decisiones judiciales de alto impacto, si no que implicará de algún modo, una resistencia injustificada a una realidad, ya impostergable. Esta investigación propone de la IA un aliado, convirtiéndose en un asistente para la correcta construcción de una estrategia probatoria en los procesos en que se pretenda la Declaración de la Unión Marital de Hecho (UMH), la cual a través del correcto entrenamiento revelará los hechos jurídicamente relevantes, los medios probatorios pertinentes para su demostración, y alertará sobre los posibles errores en la construcción de la estrategia litigiosa y el curso del proceso, evitando con ello, a quien ejerce el derecho desde el litigio, valoraciones superfluas, irrelevantes e innecesarias. Esta herramienta se articula con el marco normativo y jurisprudencial vigente en materia probatoria, conforme a los artículos 165, 167, 176 y 179 del Código General del Proceso del ordenamiento jurídico colombiano.

Palabras clave: Unión marital de hecho; valoración probatoria; medios probatorios; inteligencia artificial; litigio estratégico.

Abstract: The legal world has often expressed divergent views, asserting that the use of Artificial Intelligence (AI) will eventually replace the personal and, until now, non-delegable exercise of law. However, opposing this inevitable change will, at some point, not only result in the constant obsolescence of concepts and high-impact judicial decisions, but will also amount to an unjustified resistance to a reality that can no longer be postponed. This research conceives AI as an ally—transforming it into an assistant for the proper construction of evidentiary strategies in proceedings seeking the

(*) Abogada de la Universidad del Cauca, especialista en Derecho de Familia de la Universidad del Cauca, especialista en Derecho Público y magister en Derecho de la Universidad ICESI. Docente investigadora de la Universidad del Cauca, mentora del semillero de investigación Derecho, Arte y Sociedad.

Declaration of a couple's cohabitation. Through adequate training, AI can identify legally relevant facts, determine the appropriate means of proof for their demonstration, and warn of potential errors in the design of the litigation strategy and the course of the proceedings. In doing so, it helps legal practitioners avoid superfluous, irrelevant, and unnecessary assessments. This tool aligns with the current normative and jurisprudential framework governing evidence, in accordance with Articles 165, 167, 176, and 179 of the General Code of Procedure within the Colombian legal system.

Key words: Couple's cohabitation; evaluation of evidence; types of evidence Artificial Intelligence, strategic litigation.

"El arte del proceso no es esencialmente otra cosa que el arte de administrar las pruebas"
Jeremy Bentham

1. Introducción

Los cambios introducen casi siempre temores, algunos con un fundamento claro, otros simplemente abarcan el impulso de rechazar lo nuevo, lo que rompe paradigmas, lo que deconstruye. La inteligencia artificial (IA) y su interacción con el derecho no es la excepción.

La IA para la construcción de una adecuada estrategia probatoria en procesos judiciales que tienen por objeto la demostración de la convivencia permanente y singular, entre una pareja hetero u homo parental en Colombia, requiere necesariamente un análisis inicial que contenga datos relevantes y justifique su implementación de forma gradual, salvaguardando el ejercicio profesional de quienes ejercen el derecho desde la actividad litigiosa con la firme convicción de que para ello, deberá siempre, anteponerse su uso desde la ética.

En Colombia, según los datos proporcionados por la Unidad de Desarrollo y Análisis Estadístico (UDAE) del Consejo Superior de la Judicatura, los procesos que pretenden la declaratoria de la Unión de Marital de Hecho regulados por la ley 54 de 1990 y radicados en los despachos judiciales ascienden en el año 2021 a la suma de 10.431 procesos; 2022, a la suma de 5.895; 2023 a la suma de 8.622 y en el 2024, a la suma de 9.269 procesos.

De estos procesos judiciales, algunos tardan más de dos (2) años en obtener sentencia de primera instancia, consecuencia de varios factores, pero en esta investigación concentraremos nuestra atención en el factor asociado a la irreflexiva labor de quien ejerce el litigio y desestima la valiosa oportunidad de construir una verdadera estrategia probatoria.

Lo anterior, obedece especialmente a la falta de metodología en la estructuración de un caso concreto, relacionada especialmente con la imperante necesidad de revelar ante el juzgador congruencia entre los hechos jurídicamente relevantes y los medios probatorios aportados o solicitados.

2. Derecho probatorio: la unión marital de hecho

A través del análisis de casos reales, sentencias hito y doctrina relevante, se evidencian decisiones desfavorables que no responden a la inexistencia de la UMH, sino a errores en la estrategia probatoria planteadas desde la elaboración de la demanda.

El uso de la IA generativa, a través de modelos avanzados de aprendizaje automático pretende de forma asistencial coadyuvar a la adecuada construcción de una demanda que tenga por objeto probar la convivencia permanente de una pareja durante al menos dos (2) años, por eso, la actuación del profesional del derecho será indelegable. No obstante, la asistencia de la IA le revelará posibles errores en la construcción de hechos impertinentes, facilitando la organización estratégica de las pruebas y su relevancia en el proceso, de tal forma que, incluso alerte sobre la necesidad de omitir elementos probatorios inconducentes o contrarios al interés litigioso.

Los resultados que se obtienen al final de un proceso judicial son en gran medida, producto de las actuaciones de los sujetos procesales (demandante-demandado) quienes guiados por sus apoderados tendrán la ardua tarea de administrar los medios probatorios.

No recae el resultado de forma exclusiva en quien administra justicia, pues este en gran medida es guiado por quien elabora o desestima la oportunidad de construir una verdadera estrategia probatoria.

La Corte Suprema de Justicia en reiteradas oportunidades ha mencionado que la declaración judicial de la UMH es acreditada mediante diversos medios probatorios, resaltando el valor de los testimonios en ausencia de pruebas documentales directas.

No obstante, como bien lo señaló el Tribunal Superior de Medellín (T.S.M, Sala de Familia, sentencia 042 del 27 de febrero de 2024 M.P. Gloria Montoya Echeverri), el deber del juez de procurar la verdad material no puede ser interpretado como una autorización para suplir la carga probatoria que les corresponde a las partes.

Esta concepción ha sido respaldada por la doctrina autorizada, notablemente en el *Manual de Derecho Probatorio* (Quijano, J. P. 1986), como por el texto literal del artículo 167 del Código General del Proceso Colombiano, el cual consagra la regla general de distribución de la carga probatoria. El juez,

si bien es director del proceso, rara vez posee un conocimiento íntimo de los hechos como sí lo tienen las partes.

En los procesos de declaración de UMH, esta situación se agrava por el uso reiterado e irreflexivo de medios probatorios que resultan insuficientes o inconducentes, lo cual conlleva, frecuentemente, a decisiones judiciales que terminan siendo apeladas o incluso sometidas a revisión en sede de casación por errores relacionados con la apreciación o valoración probatoria. Lo anterior, no solo representa un costo para el sistema de justicia en términos de congestión y desgaste institucional, sino que también vulnera los principios de celeridad, economía y eficacia procesal.

Confluyen siempre desde el ejercicio del litigio, la búsqueda de la verdad material que recae en quien administra justicia y el cumplimiento del principio de autorresponsabilidad de las partes en materia probatoria. Esta tensión, cuando no es bien gestionada, deriva en litigios innecesarios, carga procesal excesiva, inseguridad jurídica y, lo más grave, en decisiones que no reflejan con fidelidad la realidad del caso. La propuesta, en consecuencia, busca contribuir al fortalecimiento de la función jurisdiccional mediante el diseño de un instrumento práctico que, sin desnaturalizar el rol de las partes ni del juez, promueva una justicia más eficiente, coherente y confiable.

La presente investigación surge de dos situaciones concretas observadas en la práctica judicial relacionadas con la indebida valoración probatoria: por un lado, se encuentran los procesos que llegan a casación por deficiencias asociadas a la valoración probatoria; y por otro, acciones de tutela interpuestas contra providencias judiciales por la misma causa.

Es entonces necesaria la revisión de diversas fuentes jurisprudenciales y doctrinales relacionadas con procesos que buscan la declaración de un estado civil y sus consecuentes efectos patrimoniales, la cual tiene por objeto identificar los medios probatorios incorporados con mayor frecuencia, cuáles son aceptados o rechazados, y a cuáles se les atribuye mayor fuerza probatoria. Esta revisión permite precisar la conexión entre los hechos de los cuales se pretende su demostración y el medio probatorio propuesto para tal fin.

La epistemología judicial crítica de la Corte Suprema de Justicia plantea una instrucción probatoria de los hechos, que permite un análisis mucho más racional, menos subjetivo o parcializado, aceptando que los jueces y los abogados suelen tener estereotipos que operan como esquemas rígidos, los cuales tienden a monopolizar la manera en que se perciben y valoran ciertos roles sociales, lo que impide una lectura crítica de la prueba, desprovista en algunos casos, de un contexto.

Esta falta de cuestionamiento lleva a que se privilegien aquellas pruebas que se ajustan a las nociones dominantes, mientras que las que provie-

nen de personas o parejas pertenecientes a grupos históricamente excluidos tienden a ser descartadas sin un análisis riguroso. Cuando este tipo de sesgos no es detectado ni corregido, se pone en riesgo la imparcialidad del juicio y se perpetúan decisiones injustas. Y es aquí donde la epistemología judicial crítica se vuelve fundamental no solo en la etapa probatoria, sino en todo el proceso judicial.

Esta iniciativa, relacionada con el diseño y aplicación de un *software* especializado coadyuvará a la creación demandas y el uso responsable de tecnologías emergentes en el derecho procesal colombiano.

En proceso judicial resuelto por la Corte Suprema de Justicia de Colombia, Sala de Casación Civil en sentencia del año 2021 (CSJ Sentencia SC 3462-2021 M.P. Luis Armando Tolosa Villanova; 18 de agosto del 2021), evidencia las complejidades inherentes a la actividad probatoria en procesos dirigidos a obtener la declaración judicial de una unión marital de hecho (UMH), especialmente cuando esta se desarrolla en contextos en los que surgen sesgos estructurales.

El proceso es promovido por JH, quien solicitó la declaración judicial de la existencia de una UMH con el causante CA, aduciendo una convivencia estable, continua e ininterrumpida desde el 4 de diciembre de 2008 hasta el 10 de julio de 2016, fecha en la cual ocurrió el fallecimiento de su compañero.

La singularidad del caso no solo radica en la naturaleza íntima de la relación alegada, sino en el entorno sociocultural que la rodeaba, caracterizado por la necesidad de mantener en la esfera de lo privado una convivencia que, si bien era conocida en el entorno más cercano de la pareja, no se presentaba abiertamente en espacios familiares más amplios debido al temor al rechazo derivado de su orientación sexual.

Durante el trámite, emergieron dos grupos claramente diferenciados de testigos. El primer grupo, conformado por dos (2) testimonios —ambas personas con orientación sexual diversa y allegados a la pareja— quienes aportaron relatos coherentes sobre la convivencia efectiva entre JH y CA en diferentes domicilios, enmarcando la relación dentro de una dinámica afectiva protegida del escrutinio público por razones de bienestar emocional y social. Incluso relataron que CA presentó a JH como su pareja dentro de espacios seguros, resaltando el rol afectivo y de cuidado ejercido por este último hacia la madre del causante.

El segundo grupo, integrado por cinco (5) testimonios, cuestionó la existencia de la UMH. Estos testigos ofrecieron versiones desprovistas de conocimiento directo sobre la intimidad de la pareja, basándose en percepciones externas y en una interpretación patrimonial de la relación, sesgada por prejuicios o visiones heteronormativas de la convivencia. Sus testimonios, en ocasiones contradictorios entre sí, no fueron debidamente con-

frontados con el resto del acervo probatorio, ni valorados con perspectiva de género, ni desde un enfoque integrador del contexto sociocultural de las parejas diversas.

El juez de primera instancia desestimó la solicitud y el Tribunal de segunda instancia confirmó el fallo, acudiendo a una interpretación formalista de los medios de prueba, sin desplegar un examen integral y sistemático de los elementos probatorios, omitiendo aplicar criterios diferenciadores conforme lo exige el artículo 176 del Código General del Proceso. Lo anterior, se tradujo en una apreciación deficiente de la prueba testimonial y en una ausencia de análisis de los elementos contextuales de discriminación estructural que suelen permear las relaciones de pareja no heteronormativas.

La Corte Suprema, en sede de casación, denunció precisamente esta deficiencia argumentativa al señalar que la omisión de una valoración probatoria con perspectiva de género y el desconocimiento del entorno sociocultural de las parejas diversas constituye una forma de discriminación judicial que perpetúa estereotipos y prejuicios contrarios al principio de igualdad sustancial. Así, la Corte destacó la necesidad de adoptar una epistemología judicial crítica, consciente de los sesgos cognitivos que pueden afectar la interpretación de los hechos y la apreciación de las pruebas, especialmente cuando estas emanan de realidades no visibles, clandestinas o marginalizadas por el discurso social dominante.

Este caso ejemplifica de manera paradigmática el problema estructural que subyace en la práctica probatoria de los procesos de declaración de UMH: la ausencia de una metodología probatoria rigurosa, articulada y con enfoque de derechos, que permita conectar adecuadamente los hechos relevantes con los medios probatorios aportados, superando visiones formales y estandarizadas que desatienden la especificidad del contexto. Asimismo, visibiliza la tensión entre el deber del juez de buscar la verdad material y la responsabilidad de las partes en la estructuración de su teoría del caso, en un campo donde la prueba testimonial y el contexto sociocultural deben adquirir un valor reforzado.

La necesidad de diseñar una herramienta metodológica que oriente a los litigantes y jueces en la adecuada selección, incorporación y valoración de los medios de prueba se presenta, entonces, no solo como una medida de fortalecimiento técnico del proceso, sino como una exigencia ética y constitucional para garantizar una justicia inclusiva, sensible al reconocimiento de otros contextos, y comprometida con la realización de derechos en condiciones reales de igualdad.

Los organismos internacionales han incoado por establecer dentro del sistema judicial garantías procesales alrededor del ámbito probatorio, en lo concerniente, encontramos el artículo 8° de la Convención Americana so-

bre Derechos Humanos y el artículo 14 del Pacto Internacional de Derechos Civiles y Políticos.

Ese garantismo en el trámite judicial está enfocado en que las partes en el escenario fáctico o jurídico puedan ser escuchados, así como tener la oportunidad de aportar las pruebas que pretenda hacer valer en el escenario judicial, siendo el juez, el llamado a determinar el valor probatorio de estos.

Es así como, toda decisión judicial deberá estar sustentada en el proverbio del latín *Da mihi factum, dabo tibi ius*, que se traduce "dame los hechos que yo te daré el derecho", con el fin de que los aspectos fácticos puedan encontrar el sustento jurídico apropiado a través de los medios probatorios que están aceptados por la normatividad interna de cada Estado (Bernal, V.M. 2021).

En el marco del derecho procesal colombiano contemporáneo, desde la vigencia del anterior Código de Procedimiento Civil, se adoptó el sistema de la sana crítica para la valoración probatoria. La entrada en vigor del Código General del Proceso (ley 1564 de 2012), reafirma la importancia de fortalecer el principio del libre convencimiento judicial. No obstante, es importante señalar que dicho convencimiento no puede ser arbitrario, sino que debe construirse a partir de juicios de valor sustentados en parámetros racionales y objetivos.

En ese sentido, el artículo 176 del Código General del Proceso establece que "[l]as pruebas *deberán ser apreciadas en conjunto*, de acuerdo *con las reglas de la sana crítica*, sin perjuicio de las solemnidades prescritas en la ley sustancial para la existencia o validez de ciertos actos", reglón seguido, resalta la norma que "[e]l juez expondrá siempre razonadamente el mérito que le asigne a cada prueba". (destacado fuera del texto original).

En consonancia con lo anteriormente indicado, el artículo 178 del precitado estatuto señala que "[e]l juez rechazará, mediante providencia debidamente motivada, las pruebas ilícitas, las notoriamente impertinentes, las inconducentes y las manifiestamente superfluas o inútiles". Esta disposición refleja la voluntad del legislador de someter la valoración probatoria a estándares mínimos de licitud, pertinencia, conducencia y utilidad, de manera que la función jurisdiccional se ejerza con apego a los principios de racionalidad y motivación.

En cuanto a la regulación de la figura jurídica de la Unión Marital de Hecho en el ordenamiento colombiano, resulta preciso señalar que esta fue regulada por la ley 54 de 1990 (Congreso de Colombia, 31 de diciembre de 1990), mediante la cual se conceptualizó dicha figura y se estableció el régimen patrimonial aplicable a los compañeros permanentes. En desarrollo de este marco normativo, la ley 979 de 2005 (Congreso de Colombia, 27 de julio del 2005) instituyó los mecanismos probatorios idóneos para la acreditación de la Unión Marital de Hecho.

Como consideración inicial de relevancia jurisprudencial, resulta procedente hacer referencia a la Sentencia C-257 de 2015 (Corte Constitucional. M.P Gloria Stella Ortiz Delgado; 6, mayo, 2015), proferida por la Honorable Corte Constitucional, en la cual se declaró la exequibilidad de la ley 979 de 2005, bajo el condicionamiento interpretativo según el cual el régimen de protección allí consagrado es igualmente aplicable a las uniones maritales de hecho conformadas por parejas del mismo sexo.

Por otra parte, la H. Corte Suprema de Justicia se ha pronunciado respecto la valoración probatoria en los procesos declarativos de Unión Marital de Hecho, como es el caso de la Sentencia SC 3982-2022 (C.S.J, SC 3892-2022. M.P Luis Alonso Rico Puerta; 13 de diciembre del 2022), donde se casa parcialmente una providencia proferida por la Sala de Familia del Tribunal Superior del Distrito Judicial de Medellín, puesto que, a consideración de la Alta Corte, el Tribunal realizó una inadecuada apreciación de las pruebas, y no efectuó una valoración conjunta de los medios probatorios de acuerdo con las reglas de la sana crítica, como lo ordena el artículo 176 del estatuto adjetivo.

Aunado a lo anterior, es imperioso traer a colocación el último pronunciamiento del Alto Tribunal en la Sentencia SC 1422 de 2025 (C.S.J, SC 1422 de 2025. M.P Martha Patricia Guzman Alvarez; 22 de mayo del 2025), reconociendo un contexto particular y creando con ello, una figura respecto a los efectos patrimoniales de la UMH, la "Sociedad de Hecho Especial entre compañeros permanentes", la cual surge ante la imposibilidad de declarar la sociedad patrimonial cuando uno de los convivientes tiene una sociedad conyugal vigente.

Autores como Jairo Parra Quijano, han decantado que el sistema de la libre valoración probatoria tiene el propósito de evitar que el juez adopte decisiones arbitrarias, debido a que se encuentra en la obligación de utilizar las reglas de la experiencia, lógica, de la historia, y psicología para que en cada proceso administre justicia con más acierto, ya que valorará la prueba de acuerdo con el caso concreto. De igual forma, deberá el administrador de justicia motivar la valoración probatoria, garantizando el derecho a la contradicción y defensa (Quijano, J. P. 1986, p. 77).

A su vez, Hernando Devis Echandía plantea que:

> "[S]in la prueba estaríamos expuestos a la irreparable violación del derecho por los demás, y el Estado no podría ejercer su función jurisdiccional para amparar la armonía social y restablecer el orden jurídico. Gráficamente expresa ese concepto el viejo Adagio: tanto vale no tener un derecho, cuanto no poder probarlo. Es decir, la administración de justicia sería imposible sin la prueba, lo mismo que la prevención de los litigios y de los ilícitos penales; no estaría orden jurídico alguno" (Echandía, H. D., & Velloso, A. A, 1984, p. 13).

La incorporación de tecnologías emergentes en la administración de justicia, y en particular en el desarrollo de la valoración probatoria dentro de los procesos declarativos de UMH, exige una construcción teórica y normativa cuidadosa, fundada sobre principios jurídicos sólidos. En efecto, la propuesta de desarrollar una herramienta basada en inteligencia artificial para organizar, correlacionar y verificar el cumplimiento probatorio de los hechos invocados en demandas de esta naturaleza debe obedecer a un marco normativo integral que garantice su legitimidad, funcionalidad y compatibilidad con los fines constitucionales del proceso judicial.

En este contexto, el principio de legalidad probatoria se erige como uno de los ejes fundamentales del sistema. Conforme al artículo 176 del Código General del Proceso (ley 1564 de 2012), las pruebas deben ser apreciadas de forma conjunta, de acuerdo con las reglas de la sana crítica, lo que implica el deber del juez de realizar un análisis racional, lógico, objetivo y motivado. Esta norma se complementa con el artículo 165 del mismo estatuto, que establece que las pruebas tienen por objeto la demostración de los hechos que fundamentan las pretensiones, excepciones y demás aspectos relevantes del proceso. La herramienta, por tanto, debe operar respetando escrupulosamente los medios de prueba establecidos por la ley (documental, testimonial, inspección judicial, etc.), y su aplicación debe orientarse a verificar que estos sean idóneos, conducentes, pertinentes y útiles frente a los hechos jurídicamente relevantes en cada caso.

El diseño de la herramienta puede fortalecer la labor de quienes administran justicia. No obstante, pretende demostrar principalmente que, el ejercicio del derecho de quienes actúan en representación de un interés particular en coadyuvancia con la IA, a través de una herramienta específica podría materializarse en la posibilidad de evitar omisiones o valoraciones parciales por parte del juez, al asegurar una trazabilidad probatoria clara y una asociación precisa entre cada hecho afirmado y los elementos de prueba que pretenden demostrarlo.

En relación con el principio de seguridad jurídica, la jurisprudencia ha sido clara en señalar que la función del proceso es proporcionar certeza y estabilidad a las relaciones jurídicas. En este sentido, una herramienta que permita correlacionar las pruebas con los hechos jurídicamente relevantes, conforme a la normatividad y jurisprudencia vigente, contribuye a reducir la incertidumbre procesal, promueve la coherencia argumentativa y fortalece la confianza en la administración de justicia.

Asimismo, la propuesta debe garantizar el principio de transparencia y verificabilidad técnica, propio de toda implementación tecnológica en escenarios jurídicos. La OCDE (2022) y organismos como la Unión Europea han subrayado la importancia de que las aplicaciones de inteligencia artificial en el sector público sean explicables, auditables y controladas por huma-

nos. Esta exigencia también se refleja en el principio de libertad probatoria del sistema colombiano, que no permite valoraciones automáticas ni algoritmos ocultos. Por lo tanto, la herramienta estará diseñada bajo parámetros de trazabilidad lógica, permitiendo al usuario conocer el razonamiento detrás de las recomendaciones que emite el sistema, sin interferir con la autonomía judicial ni la libertad de defensa.

Finalmente, resulta imprescindible el respeto por el principio de confidencialidad y protección de datos personales, previsto en la ley 1581 de 2012 (Congreso de Colombia 17 de octubre de 2012) y desarrollado por la Corte Constitucional en sentencias como la T-729 de 2002 (Corte Constitucional M.P Eduardo Montealegre Lynett; 5 de septiembre del 2002). Dado que los procesos que pretenden la declaración de la UMH implican el tratamiento de datos sensibles —relaciones personales, fotografías íntimas, registros asociados a la convivencia, de tal forma que la herramienta deberá incluir medidas de anonimización, almacenamiento cifrado y protocolos de consentimiento informado para el tratamiento de la información, garantizando así los derechos fundamentales de los usuarios.

En este marco, resulta pertinente destacar que la doctrina probatoria, especialmente la de Jairo Parra Quijano, ha desarrollado principios fundamentales que deben orientar cualquier intervención técnica en el ámbito probatorio. Así, el principio de necesidad de la prueba impone que todo hecho jurídico relevante debe ser probado para poder ser valorado por el juez, lo que excluye las decisiones sustentadas en conjeturas, intuiciones o saberes no contrastables dentro del proceso judicial. A su vez, el principio de libre apreciación de la prueba, también formulado por el autor, exige que el juez valore la prueba conforme a criterios de lógica, ciencia y experiencia, sin estar sometido a tarifas legales, pero sí a la sana crítica, lo que impone un razonamiento explícito y motivado (Quijano, J. P. 1986).

Igualmente, el principio del empleo de reglas de la experiencia estrechamente relacionado con la sana crítica implica que el juzgador no se limite a una lógica formal, sino que considere inferencias basadas en el sentido común, los conocimientos generales y la experiencia profesional acumulada. Finalmente, el principio del derecho a la prueba, que Parra Quijano vincula directamente con el debido proceso y el principio de contradicción, garantiza a todas las partes la posibilidad de solicitar, practicar, controvertir y hacer valorar pruebas de manera equitativa, equilibrada y efectiva.

3. Sistema de inteligencia probatoria asistida por IA (SIPRAI)

De esta forma, un grupo integrado por cinco (5) estudiantes del Semillero de Investigación Derecho, Arte y Sociedad de la Universidad del Cauca,

identificó la imperiosa necesidad de crear un *software* especializado denominado *SIPRAI Sistema de Inteligencia Probatoria Asistida por IA).* Herramienta que tendría por objeto apoyar el ejercicio del abogado litigante.

Del análisis realizado, **s**e evidenció que los medios probatorios comúnmente solicitados por las partes (demandante-demandado) son de naturaleza documental y testimonial, los primeros utilizados principalmente para probar hechos relacionados con la convivencia, la existencia de una relación afectiva estable y el manejo conjunto de bienes. En los casos analizados, se presentaron documentos como registros civiles, facturas funerarias, fotografías familiares, escrituras públicas, los cuales tenían como propósito demostrar el vínculo, la cohabitación prolongada y la contribución mutua a una vida en común. Los segundos, pruebas testimoniales que pretendían reforzar estos hechos desde una perspectiva subjetiva, —de acuerdo con las posibilidades que tenía cada testigo de conocer de manera cercana la convivencia de la pareja— pretende dar cuenta del contexto.

En cuanto a la valoración judicial, se percibe que la mayoría de las pruebas solicitadas fueron efectivamente decretadas por el juez, aunque en algunos casos no todas llegaron a practicarse en razón a diferentes causas, como, por ejemplo, la no comparecencia de un testigo. Sin embargo, se destaca que no todos los medios practicados fueron efectivamente valorados en la sentencia: algunos documentos fueron mencionados de forma general sin una referencia directa a su contenido probatorio, lo que podría generar vacíos en la motivación judicial. Pese a ello, se observa una mayor tendencia de los jueces a valorar pruebas documentales que presentan un contenido objetivo y verificable, mientras que las pruebas testimoniales son admitidas con mayor cautela, especialmente si no están debidamente relacionadas con hechos jurídicamente relevantes. Este comportamiento refuerza la necesidad de una herramienta que, apoyada en IA, permita hacer seguimiento sistemático al ciclo probatorio completo y facilite una valoración coherente y crítica de la prueba en los procesos de UMH.

El *Sistema de Inteligencia Probatoria Asistida por IA* (SIPRAI) que aquí se plantea busca ser una respuesta jurídica y asistencial al problema estructural que aqueja a los procesos de declaración de UMH en Colombia, esto es, la desconexión sistemática entre los hechos invocados en la demanda y el acervo probatorio allegado.

En la Sentencia SC-3982 de 2022, la Corte Suprema de Justicia advirtió que la indebida apreciación de la prueba por omisión, descontextualización o valoración fragmentaria constituye una de las causas más frecuentes de errores judiciales en procesos civiles de familia. Ello ha conducido a una multiplicación innecesaria de recursos extraordinarios, tutelas contra providencias judiciales y, en consecuencia, a un desgaste institucional que podría ser evitado mediante un uso estratégico de la tecnología jurídica.

Este *software* especializado de valoración probatoria está dirigido principalmente a abogados litigantes. Su arquitectura funcional se basa en tres módulos principales: (i) ingreso y estructuración de hechos jurídicamente relevantes, (ii) carga y categorización de medios probatorios disponibles y (iii) evaluación preliminar de la correspondencia entre hechos y pruebas con base en la legislación y jurisprudencia aplicables. Esta funcionalidad busca facilitar al abogado una revisión sistemática del acervo probatorio, optimizando la preparación de la demanda y/o alegatos de conclusión.

Desde el punto de vista normativo, la herramienta se fundamenta en el marco previsto por el Código General del Proceso, en particular los artículos 165, 167, 176 y 179, que regulan el objeto, carga, valoración y contradicción de la prueba. En el plano sustancial, incorpora el marco legal contenido en la ley 54 de 1990 y la ley 979 de 2005, así como la reciente doctrina jurisprudencial sobre la figura de la "sociedad de hecho especial". Estos elementos serán incorporados a través de una plataforma de razonamiento legal con técnicas de procesamiento de lenguaje natural (PLN), que permitirá al sistema identificar y clasificar hechos, sugerir medios probatorios admisibles según el caso, y generar informes argumentativos con fundamentos normativos, jurisprudenciales y doctrinales.

El plan de acción contempla tres fases: una fase inicial de diagnóstico y análisis de expedientes, con el propósito de identificar patrones en la presentación y valoración de medios probatorios; una fase intermedia de desarrollo y entrenamiento del sistema, en la que se integrarán modelos semánticos basados en jurisprudencia real complementados con fundamentos normativos específicos para fortalecer la validez de los argumentos que se puedan llegar a elaborar con el uso de la herramienta; y finalmente, una fase de implementación piloto, con pruebas controladas en consultorios jurídicos y simulacros judiciales, orientadas a evaluar el impacto en la preparación de demandas, la calidad de los escritos y la eficiencia del litigio. Este enfoque progresivo permitirá validar su utilidad práctica sin alterar el equilibrio del proceso judicial ni vulnerar garantías.

Además, la herramienta proporcionará a los profesionales del derecho un recurso de análisis crítico que facilitará la elaboración de demandas sólidas, provistas de sustento probatorio y conforme a la normativa/precedente judicial vigente. En esencia, representa un aporte académico y práctico para integrar la IA al servicio del proceso judicial, dotando al abogado de herramientas metodológicas que le permita desempeñar plenamente su rol como defensor del derecho material en los litigios de familia.

La presente investigación permitió constatar la existencia de una problemática estructural en los procesos judiciales de declaración de unión marital de hecho (UMH), relacionada con la débil articulación entre los hechos afirmados en la demanda y los medios de prueba incorporados al

proceso. Esta disociación, identificada a través del análisis empírico de expedientes tramitados ante los Juzgados de Familia del Circuito de Popayán, evidencia un patrón de deficiencia argumentativa que impacta negativamente en la valoración probatoria y compromete principios fundamentales del proceso como la legalidad, la motivación judicial, la igualdad de las partes y el debido proceso.

Adicionalmente, se advirtió cómo la ausencia de un enfoque contextual e inclusivo en la apreciación de la prueba puede reproducir patrones discriminatorios, especialmente en casos en los que convergen factores de invisibilidad social o marginación estructural. La jurisprudencia reciente —particularmente la Sentencia SC 3462-2021 de la Corte Suprema de Justicia— demuestra que la omisión de una valoración integral, sistemática y con perspectiva de derechos constituye una causal frecuente de errores judiciales y de inobservancia del principio de verdad material.

En este contexto, la investigación plantea como propuesta metodológica el diseño de una herramienta de apoyo, basada en inteligencia artificial y orientada a la evaluación técnica de la prueba conforme a los principios del sistema probatorio colombiano. Dicha herramienta, denominada *SIPRAI* (*Sistema de Inteligencia Probatoria Asistida por IA*), se estructura en tres módulos funcionales: (i) Estructuración de hechos jurídicamente relevantes; (ii) clasificación y análisis preliminar del acervo probatorio y su relación con los hechos anteriores; y (iii) diagnóstico normativo y jurisprudencial. Su diseño busca asistir al abogado litigante en la organización técnica y jurídica de su teoría del caso, y al mismo tiempo, ofrecer al operador judicial un modelo lógico para reforzar la motivación fáctica y jurídica de sus decisiones.

SIPRAI no emitirá juicios de valor sobre los hechos y los medios probatorios practicados o aportados para sustentarlos. No obstante, podrá apoyar la identificación del cumplimiento —o la ausencia— de ciertas reglas concretas de valoración, tales como la conducencia, pertinencia o utilidad de los medios de prueba. Este *software* especializado será entrenado con base en información específica de cada etapa del proceso, incluyendo las apreciaciones profesionales que el jurista tenga respecto de la admisibilidad o práctica de cada prueba, conforme a su criterio jurídico y experiencia.

El desarrollo de SIPRAI está sustentado en los artículos 165, 167, 176 y 179 del Código General del Proceso, y responde a una necesidad concreta dentro del contexto de transformación digital del sistema judicial colombiano. Su implementación progresiva inicialmente en entornos académicos permitiría evaluar su impacto en la racionalización del litigio, la mejora de la calidad de la argumentación probatoria y la reducción de cargas innecesarias para la administración de justicia.

4. CONCLUSIONES

En síntesis, se concluye que la valoración probatoria en los procesos de UMH requiere no solo de un reforzamiento teórico, normativo y jurisprudencial, sino también de soluciones asistenciales que promuevan una justicia eficiente, motivada y respetuosa de los principios constitucionales. La herramienta propuesta representa un primer paso hacia ese horizonte, articulando los desarrollos de la inteligencia artificial (IA) con las exigencias de rigor jurídico y garantía procesal propias del Estado social de derecho.

Finalmente, debo realizar un especial reconocimiento a los estudiantes Manuela Rubio Correa, María Isabel Méndez, María Victoria Burbano, Juan David Gironza, María Camila Muñoz y Luisa Fernanda Castaño, integrantes del Semillero de Investigación, Derecho, Arte y Sociedad de la Universidad del Cauca, quienes han trabajado en esta propuesta de investigación y contribuyeron significativamente a lo reflejado en este capítulo.

REFERENCIAS BIBLIOGRÁFICAS

Bernal, Víctor Manuel. *Módulo de formación autodirigida sobre valoración de la prueba en materia* [en línea]. [s.l.]: Escuela Judicial "Rodrigo Lara Bonilla", 2021 [consultado el 7, marzo, 2025]. 179 p. Disponible en Internet: https://escuelajudicial.ramajudicial.gov.co/sites/default/files/biblioteca/21.%20MÓDULO%20DE%20FORMACIÓN%20AUTODIRIGIDA%20SOBRE%20VALORACIÓN%20DE%20LA%20PRUEBA%20EN%20MATERIA%20CIVIL%20Y%20COMERCIAL_compressed.pdf.

Colombia. Congreso de la República. *Ley 54 de 1990*. Por la cual se definen las uniones maritales de hecho y régimen patrimonial entre compañeros permanentes. *Diario Oficial* No. 39.615, 31 de diciembre de 1990.

Colombia. Congreso de la República. *Ley 979 de 2005*. Por medio de la cual se modifica parcialmente la *Ley 54 de 1990* y se establecen unos mecanismos ágiles para demostrar la unión marital de hecho y sus efectos. *Diario Oficial* No. 45.982, 27 de julio de 2005.

Colombia. Congreso de la República. *Ley 1564 de 2012*. Por medio de la cual se expide el *Código General del Proceso*.

Colombia. Congreso de la República. *Ley 1581 de 2012*. Por la cual se dictan disposiciones generales para la protección de datos personales. *Diario Oficial* No. 48.587, 17 de octubre de 2012.

Colombia. Corte Constitucional. *Sentencia T-729/02* (5, septiembre, 2002). M.P. Eduardo Montealegre Lynett, *Expediente T-467467*.

Colombia. Corte Constitucional. *SC 257/15* (6, mayo, 2015). M.P. Gloria Stella Ortiz Delgado, *Expediente D-10462*.

Colombia. Corte Suprema de Justicia, Sala de Casación Civil y Agraria. *SC 3462-2021* (18, agosto, 2021). M.P. Luis Armando Tolosa Villanova, *Expediente N° 25754-31-10-001-2017-00070-01*.

Colombia. Corte Suprema de Justicia, Sala de Casación Civil. *SC 3892-2022* (13, diciembre, 2022). M.P. Luis Alonso Rico Puerta, *Expediente N° 05001311000520190026702*.

Colombia. Corte Suprema de Justicia, Sala de Casación Civil, Agraria y Rural. *SC 1422-2025* (22, mayo, 2025). M.P. Martha Patricia Guzmán Álvarez, *Expediente N° 68001310005202100314O1*.

Echandía, H. D., & Velloso, A. A. (1984). *Compendio de la prueba judicial: Compendio de pruebas judiciales* (Vol. 1). Rubinzal-Culzoni.

OECD/CAF (2022). *Uso estratégico y responsable de la inteligencia artificial en el sector público de América Latina y el Caribe. Estudios de la OCDE sobre Gobernanza Pública*, OECD Publishing, Paris, https://doi.org/10.1787/5b189cb4-es.

Quijano, J. P. (1986). *Manual de derecho probatorio* (Vol. 1). Ediciones librería del profesional.

Justicia e inteligencia artificial: propuesta de armonización procesal a través de una ley modelo

Por Rosa A. Avila Paz de Robledo [(*)] *y Diego Robledo* [(**)] [(***)]

Resumen: Esta investigación explora el vínculo entre inteligencia artificial (IA) y Justicia, con foco en su regulación jurídica y constitucional, los derechos humanos involucrados, y las prácticas judiciales actuales. Analiza cómo profesionales de la magistratura y la abogacía utilizan IA generativa en sus tareas, y qué implicancias jurídicas, procesales y éticas surgen de ello. El estudio examina el marco normativo vigente, los desafíos que plantea la IA en el ámbito judicial, y las tensiones entre automatización, garantías procesales y responsabilidad profesional. Se aplica la dogmática jurídica y el método comparado procesal, así como la construcción de un prototipo ideal como matriz de análisis. También se utiliza la perspectiva sistémica y el enfoque de políticas públicas para analizar esta relación entre IA y Justicia y formular las conclusiones. Como cierre, se propone una ley modelo para regular el uso de IA en procesos judiciales, con el objetivo de armonizar criterios, proteger derechos fundamentales y orientar buenas prácticas en el ecosistema de justicia digital y algorítmico.

Palabras clave: Justicia; inteligencia artificial; ley modelo; armonización procesal.

(*) Doctora en Derecho y Ciencias Sociales, Universidad Nacional de Córdoba, UNC. Doctora Honoris Causa, Universidad Nacional de La Rioja, UNLaR. Docente investigadora, Categoría I Ministerio de Educación de la Nación. Profesora titular y de posgrado, UNC. Prof. titular regular y directora del Instituto de Derecho Procesal, de la Maestría en Derecho Procesal y Doctorado en Ciencias Jurídicas de UNLaR. Miembro del Instituto Iberoamericano de Derecho Procesal, la Asociación Argentina de Derecho Procesal, y la Asociación Internacional de Derecho Procesal. Abogada y mediadora. E-mails: rosa.avila@unc.edu.ar, ravila@unlar.edu.ar y rosaavilapaz@gmail.com

(**) Doctor en Derecho y Ciencias Sociales, Universidad Nacional de Córdoba, UNC. Postdoctorado en UNC y en la Universidad de Bolonia. Docente investigador Categoría III-Ministerio de Educación de Nación. Profesor en UNC y profesor adjunto regular, coordinador académico la Maestría en Derecho Procesal y del Doctorado en Ciencias Jurídicas de la Universidad Nacional de La Rioja. Miembro del Instituto Iberoamericano de Derecho Procesal, Asociación Argentina de Derecho Procesal, y la Asociación Internacional de Derecho Procesal. Abogado y mediador. E-mails: diego.robledo@unc.edu.ar, drobledo@unlar.edu.ar y drobledoavilapaz@gmail.com

(***) Los autores expresan su especial agradecimiento al Prof. Dr. Gonzalo Ana Dobratinich por su invitación para contribuir con el artículo al libro Derecho e inteligencia artificial de la Facultad de Derecho de la Universidad de Buenos Aires, Argentina.

Abstract: This research explores the relationship between artificial intelligence (AI) and Justice, focusing on its legal and constitutional regulation, the human rights involved, and current judicial practices. It analyzes how members of the judiciary and legal professionals use generative AI in their work, and the legal, procedural, and ethical implications that arise from this use. The study examines the existing regulatory framework, the challenges AI poses in the judicial sphere, and the tensions between automation, procedural safeguards, and professional responsibility. It applies legal dogmatics and comparative procedural methods, along with the construction of an ideal prototype as an analytical matrix. A systemic perspective and a public policy approach are also employed to examine the interplay between AI and Justice and to formulate conclusions. As a final contribution, the research proposes a model law to regulate the use of AI in judicial proceedings, aiming to harmonize procedural standards, safeguard fundamental rights, and guide best practices within the algorithmic and digital justice ecosystem.

Key words: Justice; Artificial Intelligence; model law; procedural harmonization.

1. INTRODUCCIÓN

La presente investigación parte de un diálogo que relaciona a la inteligencia artificial (en adelante IA) con el campo [(1)] de la Justicia a partir de sus ejes conceptuales.

Se parte de un concepto dinámico de la inteligencia artificial siguiendo el lineamiento de la UNESCO que lo entiende como "aquellos sistemas con capacidad para procesar datos de forma similar a un comportamiento inteligente y abarca generalmente aspectos de razonamiento, aprendizaje, percepción, predicción, planificación o control" (UNESCO, 2022) (UNESCO, 2025). Complementando este concepto, la Organización de las Naciones Unidas expresa que incluye una gama amplia de tecnologías a los que identifica como "sistemas adaptativos de autoaprendizaje" y de acuerdo a la tecnología lo clasifica en base a "los propósitos (como el reconocimiento facial o de imágenes), las funciones (como la comprensión del lenguaje y la resolución de problemas) o los tipos de agentes (incluidos los robots y los coches autoconducidos)", en su conceptualización incluyen métodos que combinan la visión, el reconocimiento de voz y la robótica con la posibilidad de potenciar o aumentar las capacidades humanas, así como el desarrollo de

(1) Entendido como "el espacio social organizado en y por el cual tiene lugar la transmutación de un conflicto directo entre partes directamente interesadas en un debate jurídicamente reglado entre profesionales que actúan por procuración y que tienen en común su conocimiento y reconocimiento de la regla del juego jurídico, es decir, las leyes escritas y no escritas del campo (Bourdieu, 2001, pág. 164).

"la capacidad de procesamiento informático y el desarrollo de técnicas de manejo de datos" (Organización de las Naciones Unidas, 2025).

A partir de lo expresado, la presente investigación documental es teórica y se encuentra situada en la República Argentina con aportes de tecnopraxis de la dogmática jurídica (Atienza, 2025). A su vez, se aplica para el análisis una perspectiva sistémica y de políticas públicas (Pereira Campos, 2025) y de derecho procesal comparado (Oteiza, 2019).

El método comparado, siguiendo a Cappelletti y Garth, parte de la cooperación interdisciplinaria para analizar el corpus de investigación (1977). Asimismo, se construye una matriz en base a ejes para una ley modelo (Avila Paz de Robledo, R. A. (directora), 2005, pág. 81), pues como enseñaba Taruffo "la construcción de 'prototipos ideales' de los diversos ordenamientos puede establecer un nivel de análisis sobre el cual la comparación es posible y provechosa, evitando perderse en la masa infinita de detalles" (2006).

Un antecedente necesario en esta temática es la investigación llevada a cargo por Jordi Nieva Fenoll que vincula al proceso judicial y a la inteligencia artificial (2018) y de manera más reciente la investigación dirigida por Rosa Avila Paz de Robledo (Avila Paz de Robledo, R. A., 2024).

2. Marco normativo e inteligencia artificial

El tratamiento jurídico de la inteligencia artificial requiere la comprensión integral y sus implicancias en la vida humana y de las instituciones. Uno de los desafíos consiste su acelerada metamorfosis a la que está esencialmente sujeta conforme los avances tecnológicos y la centralidad que ocupa de manera exponencial en tareas de la vida cotidiana de las personas y de las instituciones.

2.1. Modelos regulatorios

Regular o no regular es la cuestión. Luego del equilibrio entre permisos, obligaciones y prohibiciones en el ámbito jurídico surgen otras preguntas ¿regular qué?, ¿regular cuánto?, ¿regular cómo?, ¿cuál camino tomar?

Desde el derecho comparado, se observa que al menos tres modelos regulatorios de la inteligencia artificial. En primer lugar, el *modelo centrado en la dignidad de la persona humana*, y por consiguiente en la privacidad y protección de los derechos humanos, procurando un equilibrio entre innovación y supervisión rigurosa, como es el caso de Europa que ha adoptado en la Unión Europea una Ley de Inteligencia Artificial, en segundo lugar el *modelo enfocado estratégicamente en el control estatal y el progreso tecnológico*, como el caso de China, y en tercer lugar, el *modelo centrado en el*

mercado con un marco regulatorio flexible como el caso de Estados Unidos (Benzal Medina, 2025, pág. 80).

Por otro lado, es posible diferenciar entre modelos de acuerdo al grado de generalidad y especialidad, en tanto es posible contar con leyes generales para la inteligencia artificial o leyes particulares.

En este sentido, se destaca que a nivel Europeo, el 2 de febrero de 2025 entró en vigencia el Reglamento (UE) 2024/1689 del Parlamento y del Consejo de la Unión Europea —en sus capítulos 1 y 2— con el objeto de armonizar las normas vinculadas a la IA. A través de este reglamento se promueve el uso de inteligencia artificial confiable y centrada en las personas, que asegure altos estándares de salud, seguridad y el pleno respeto de los derechos consagrados en la Constitución (Parlamento Europeo y del Consejo de la Unión Europea, 2024).

En el ámbito del Mercado Común del Sur (MERCOSUR) no hay una regulación general como en la Unión Europea, pero si se ha firmado la Declaración de Ministros y Altas Autoridades sobre Derechos Humanos de los Estados Partes del MERCOSUR sobre los principios de derechos humanos en el ámbito de la inteligencia artificial (Mercado Común del Sur, 2023) que expresa que "la inteligencia artificial debe partir de la centralidad humana para ser desarrollada, programada, entrenada, utilizada, monitoreada, evaluada y regulada, y hacerlo con base en los principios de garantía, promoción, protección y defensa de los derechos humanos reconocidos nacional e internacionalmente por los países del MERCOSUR", asimismo incorpora como principios la igualdad y la no discriminación, atendiendo a sesgos de género y étnicos raciales, la privacidad y protección de datos personales, en cuanto a la libertad de expresión expresa que "las plataformas digitales de comunicación e información deben asegurar la integridad de la información en los medios digitales y evitar la programación y utilización de algoritmos de recomendación que violen el derecho a la libertad de expresión" y también incluye los principios de transparencia y explicabilidad algorítmica, control humano de la tecnología, responsabilidad y rendición de cuentas y la colaboración multisectorial para desarrollo responsable de la IA.

La Asamblea General de la Organización de las Naciones Unidas a través de la res. 79/1 aprobada el 22 de septiembre de 2024 incluyó el *Pacto digital global* como integrante la declaración "El Pacto para el futuro" en que pone de relieve que las tecnologías emergentes como la inteligencia artificial "son importantes para facilitar el desarrollo sostenible y están cambiando nuestro mundo de manera radical. Las posibilidades que ofrecen a las personas y el planeta de progresar, tanto en la actualidad como en el futuro, son enormes. Tenemos la determinación de materializar ese potencial y gestionar los riesgos aumentando la cooperación internacional, interactuando con las instancias pertinentes y promoviendo un futuro digital inclusivo, res-

ponsable y sostenible" y en este sentido, decidió establecer un *Panel Científico Internacional Independiente sobre Inteligencia Artificial* integrado por 40 miembros de carácter multidisciplinar y conformado con representación geográfica equilibrada para promover el conocimiento científico e iniciar un *Diálogo Mundial sobre la Gobernanza de la IA* (2024).

Luego por res. 79/325 de fecha 26 de agosto de 2025 en el marco de la Asamblea General de la Organización de Naciones Unidas se decidió que este Panel sea una "plataforma para debatir la cooperación internacional, intercambiar las mejores prácticas y la experiencia adquirida, y facilitar debates abiertos, transparentes e inclusivos sobre la gobernanza de la inteligencia artificial con miras a permitir que la inteligencia artificial contribuya a la consecución de los Objetivos de Desarrollo Sostenible y a cerrar las brechas digitales entre los países y dentro de ellos", y en particular se ocupa entre otros de: a) el desarrollo de sistemas de inteligencia artificial seguros y fiables, b) las implicaciones sociales, económicas, éticas, culturales, lingüísticas y técnicas de la inteligencia artificial, c) la interoperabilidad y compatibilidad de los enfoques de gobernanza de la inteligencia artificial, d) el respeto, la protección y la promoción de los derechos humanos en el ámbito de la inteligencia artificial; e) La transparencia, la rendición de cuentas y una estricta supervisión humana de los sistemas de inteligencia artificial de modo compatible con el derecho internacional y f) el desarrollo de programas informáticos de código abierto, datos abiertos y modelos de inteligencia artificial abiertos (2025).

Por el momento, se observa que, entre las instituciones internacionales y regionales, hay una preocupación por el tema de la regulación de la IA, y se presenta como un debate abierto. También hay una preocupación por los riesgos y por sesgos y estereotipos que pueden afectar el uso confiable y responsable de la IA (Robledo, 2023) (Avila Paz de Robledo, R.A., 2023).

2.2. Hacia un modelo argentino

En busca de un modelo argentino, se observa la Constitución de la Nación (Congreso de la Nación Argentina, 2025). Pues allí se plasma el plan de vida del Estado de Derecho. Se parte de una base en común que consagra entre las declaraciones y tratados a los que la Constitución Nacional les asigna jerarquía constitucional en el artículo 75 inciso 22, que en los artículos 27 de la Declaración Universal de los Derechos Humanos y 13 de la Declaración Americana de Derechos y Deberes del Hombre reconocen además el derecho de acceso a la cultura, a participar del progreso científico y a gozar de sus beneficios. También en este sentido el Pacto Internacional de Derechos Económicos, Sociales y Culturales prescribe, en el art. 15 inciso b. Y con jerarquía superior a las leyes nacionales, se encuentra el Protocolo de San Salvador (Protocolo adicional a la Convención Americana sobre Dere-

chos Humanos en materia de Derechos Económicos, Sociales y Culturales) que también expresa en el art. 14 b este derecho.

También desde su diseño constitucional adopta la forma de gobierno federal, y ello conlleva considerar que existen como sujetos federales: la Nación, las provincias y la Ciudad Autónoma de Buenos Aires. A su vez, el equilibrio de competencias y poderes, propio de vivir el federalismo —en el que se asegura a los estados locales la administración de su justicia y también a la Nación confía la justicia nacional y federal— genera una diversidad de formas jurídicas, de modelos constitucionales, modelos procesales, modos de reglamentación por las cortes supremas y un variopinto de matices de las prácticas judiciales.

Es a partir de esa diversidad que el análisis comparativo ofrece la oportunidad de revalorizar experiencias locales entre las que se registra una tendencia en las constituciones reformadas entre 2020 y 2025 de incluir un capítulo sobre derechos, deberes y garantías digitales.

2.2.1. Constitución de la Provincia de Jujuy

El texto de la Constitución reformada publicado el 21 de junio de 2023 (2023) introduce constitucionalmente a la Inteligencia artificial. Asimismo, reconoce:

i. La democratización del conocimiento y la inclusión digital, promoviendo la conectividad digital universal y efectiva "como un medio para garantizar la igualdad de oportunidades y el ejercicio de los derechos" (art. 72).

ii. Establece el deber estatal de promover "el acceso a las Tecnologías de la Información y las Comunicaciones, en condiciones de igualdad, accesibilidad y asequibilidad para todas las personas" (art. 72).

iii. Reconoce el derecho al acceso a la innovación y la tecnología en todas sus formas respetando derechos, principios y garantías constitucionales (art. 73).

iv. Asimismo, establece que la "innovación y el desarrollo tecnológico deberán ser respetuosos de los derechos humanos y los principios éticos que establezca la ley".

v. Puntualmente en el artículo 76 se regula con el título "Inteligencia artificial o no humana" el reconocimiento del "derecho de toda persona a utilizar sistemas de inteligencia artificial o no humana, basados en métodos computarizados de algoritmos, datos y modelos que imitan el comportamiento humano y automatizan procesos complejos, así como otros futuros desarrollos que surjan en este campo".

vi. También el artículo 76 establece que una ley "sujetará estos sistemas a los principios de legalidad, transparencia, responsabilidad, privacidad y protección de datos, seguridad, no discriminación y rendición de cuentas, garantizando el acceso a la justicia en caso de vulneración de derechos y consagrando la acción de solicitud de revisión humana cuando sea necesario".

vii. También el artículo 76 plantea que, a cargo del Estado, está la "educación y el debate público sobre los desafíos éticos y jurídicos que plantean estos sistemas".

viii. El artículo 76 establece el principio de primacía de los derechos humanos y de las libertades y garantías constitucionales a favor de las personas.

ix. También incluye en cuanto a la protección de datos personales el derecho a conocer y tener acceso a la información que se encuentre en archivos, registros y bases de datos, tanto físicos como digitales, públicos o privados. Asimismo, se debe respetar los "principios de finalidad, consentimiento, calidad, veracidad, seguridad, confidencialidad y autodeterminación informativa" (art. 77).

x. Queda a cargo del Estado la protección de datos personales de la población, así como la seguridad de los datos del Estado, así como la cooperación nacional e internacional en materia de protección de datos personales, para garantizar la seguridad y el respeto de este derecho en el ámbito global.

Este articulado ha de interpretarse de manera coherente al interpretar la garantías judiciales de derecho de defensa y debido proceso previstos en el artículo 29, así como lo regulado en cuanto al Poder Judicial entre el artículo 167 a 172, como la atribución para incorporar tecnologías y herramientas de gestión que coadyuven a la agilización de la tramitación y resolución de las causas, así como los principios procesales entre los cuales se destaca la toma de decisión debidamente fundada, dentro de un plazo razonable, y su ejecución, así como la celeridad y eficacia en la administración del caso y su resolución.

2.2.2. Constitución de la Provincia de La Rioja

La última reforma constitucional fue el 11 de junio de 2024 en la que se incluyeron en tres artículos puntuales los derechos y deberes digitales, así como la regulación de la inteligencia artificial (2024). Así sus notas salientes son:

i. El deber estatal de garantizar el respeto a la dignidad humana y los derechos de la persona en los entornos digitales como condición fundamental para su desarrollo en libertad (art. 38).

ii. El deber estatal de promover la neutralidad tecnológica y la autodeterminación digital mediante deberes específicos. Así como el fomento de la innovación, la investigación y el desarrollo científico y tecnológico para reducir asimetrías (art. 38).

iii. Bajo el título de "Ciudadanía Digital y Gobierno Abierto" se consagra en el artículo 39 que el Estado tiene a su cargo el impulso de la ciudadanía digital "mediante acciones positivas que promuevan el uso de tecnologías de la información de modo ético, seguro y transparente". Esta se integra con "el sistema de gobierno electrónico y democracia digital, que consiste en la administración de los recursos del Estado mediante las Tecnologías de la Información y la Comunicación, y sus principios orientadores basados en la transparencia, rendición de cuentas y participación ciudadana".

iv. A su vez, el artículo 39 establece que se "promoverán métodos alternativos de resolución de conflictos a través de las tecnologías".

v. Bajo el título de "Macrodatos" el artículo 40 expresa que los "datos masivos, inteligencia artificial de datos, datos a gran escala o *big data* a los conjuntos de datos complejos que precisan de aplicaciones informáticas no tradicionales de procesamiento de datos". Y deja a cargo del estado la obligación de sancionar una ley especial en relación al procedimiento de tratamiento de los macrodatos que garantice "la protección de las personas con el tratamiento de los datos personales, la circulación libre y transparente de datos de carácter público, de interés general y humanitario".

Se observa que se integra a los derechos y garantías judiciales previstas constitucionalmente.

2.2.3. Constitución de la Provincia de Santa Fe

La Constitución de la Provincia fue reformada fue publicada en Boletín Oficial el 11 de septiembre de 2025 (Convención Constituyente de la Provincia de Santa Fe, 2025). Como antecedente de la reforma se señala que la ley 14.384, que declara la necesidad de la reforma constitucional, habilitó "la discusión en el sentido de incorporar principios en materia de derechos y deberes digitales, ciudadanía digital y gobierno abierto" (Poder Legislativo de la Provincia de Santa Fe, 2024). Al leer el texto reformado de la Constitución se observan algunas notas salientes:

i. Se parte de la premisa por la cual "el Estado reconoce a la persona humana su eminente dignidad" y expresa la obligación que alcanza a todos los órganos del poder público "a respetarla y protegerla" (art. 9°).

ii. Se incorpora un reconocimiento de equivalencia jurídica en relación a los entornos digitales prescribiendo que "los derechos previstos en esta Constitución son aplicables en entornos digitales" (art. 28).

iii. También expresa que la "Provincia impulsa el desarrollo y uso ético de las tecnologías de manera segura y orientados al bien común, preservando la centralidad y la dignidad de la persona humana" (art. 28).

iv. Se garantiza el "derecho al acceso universal, equitativo, asequible, sin discriminación, continuo y de calidad a la tecnología, a la conectividad y a la infraestructura tecnológica" y para ello establece que se adopten políticas públicas adecuadas para eliminar las brechas digitales (art. 28).

v. Reza que el Estado "promueve la construcción de una ciudadanía digital inclusiva, la alfabetización digital crítica, el respeto a la privacidad, la autonomía mental y la no manipulación de pensamientos, emociones, acciones o decisiones mediante el uso de tecnologías" y afirma que el consentimiento debe ser libre, informado, manifestado de forma expresa, y revocable (art. 28).

vi. Se reconoce que la violencia discursiva en entornos digitales constituye una forma de deterioro del debate público democrático (art. 16).

vii. En cuanto a la protección de datos personales en entornos digitales consagra el derecho que abarca el ejercicio de "control en el uso, gestión, circulación y conservación de su información personal", así como el "derecho a conocer de forma clara, accesible y comprensible los criterios, parámetros y lógicas utilizadas en los sistemas automatizados o algorítmicos de toma de decisiones y a la intervención de una persona humana cuando esa decisión pueda afectar sus derechos, especialmente en los que se implemente inteligencia artificial u otras tecnologías emergentes" (art. 29).

viii. A su vez consagra el derecho a vivir en un entorno físico y digital seguro, libre de violencias y amenazas, lo que se armoniza con el deber estatal de adoptar políticas públicas integrales de ciberseguridad.

ix. En cuanto a los servicios de interés público, a cargo de la Provincia o de terceros que presten estos servicios, establece el deber de adoptar "sistemas algorítmicos transparentes y auditables y promover mecanismos de evaluación de impacto y resguardo frente a sesgos o discriminación" (art. 29). Este deber es puntualmente relevante en cuanto al servicio de administración de justicia y formas extrajudiciales de solución de controversias.

x. En cuanto a la Justicia, se garantiza "el acceso a la justicia y la resolución de conflictos en un plazo razonable" (art. 6°) y se consagra el deber de tribunales y jueces de resolver las causas dentro de los plazos legales (art. 132) y de dar motivación suficiente a las sentencias y autos interlocutorios bajo pena de nulidad (art. 131).

La construcción de un ecosistema de justicia digital y algorítmico es fruto de una decisión institucional y una necesidad de ajuste a los tiempos que corren.

2.2.4. Constitución de la Provincia de Formosa

La reforma constitucional publicada el 8 de septiembre de 2025 de la Constitución formoseña (2025) tiene entre sus notas salientes:

i. El Estado debe garantizar el respeto a la dignidad humana y los derechos de la persona en los entornos digitales (art. 73).

ii. Se consagra la "ciudadanía y participación digital" y que expresa a través de acciones positivas se promoverá el "uso de tecnologías de la información de modo ético, seguro y transparente" (art. 73).

iii. Se consagra el derecho de acceder a los derechos de cuarta generación que surgen en el marco de la sociedad del conocimiento y el desarrollo científico y tecnológico (art. 75).

iv. También para que el derecho a acceder sea en igualdad de oportunidades, pone a cargo del Estado la ejecución de "políticas activas para asegurar conectividad, alfabetización digital y accesibilidad a los integrantes de grupos vulnerables de la comunidad" atendiendo al bien común (art. 75).

v. En cuanto a la garantía de inviolabilidad de papeles privados y correspondencia, se incluye expresamente "las comunicaciones telegráficas, telefónicas, cablegráficas, digitales, electrónicas o de cualquier otra especie" (art. 17).

vi. El artículo 76 expresamente establece que "se impulsará el desarrollo e implementación de inteligencia artificial, datos masivos, nanotecnología, biotecnología, robótica y demás tecnologías actuales o emergentes que contribuyan al progreso y al bienestar de la comunidad".

vii. Se destaca en esta regulación del artículo 76 que el reconocimiento como principios rectores para la regulación provincial a tres puntualmente: la soberanía tecnológica, la inclusión laboral y la apropiación social del conocimiento.

viii. Asimismo, el artículo 76 expresa que la regulación "deberá garantizar el respeto a los derechos humanos, la democracia y el Estado de derecho, así como el control humano significativo, no discriminación algorítmica, trazabilidad, transparencia y evaluación de impacto social, laboral y ambiental".

Se armoniza con los principios y garantías judiciales expresados constitucionalmente.

2.2.5. Balance de derecho comparado

Se observa la tendencia hacia una regulación constitucional de derechos y deberes digitales y en relación a la inteligencia artificial. Asimismo,

se verifica entre las reformas constitucionales de provincias argentinas entre 2020 y 2025 que se da tratamiento jurídico en los textos constitucionales como una realidad que existe y requiere su regulación. A su vez, se observa que las reformas mencionadas siguen el modelo regulatorio centrado en la persona humana, su dignidad y la protección de los derechos humanos. Varios de los textos constitucionales de las reformas encomiendan a una ley especial que contemple el tratamiento jurídico de la IA.

3. Hacia la constitución de un ecosistema de justicia digital y algorítmico

Entre la literatura se utiliza términos tales como proceso electrónico, proceso digital, justicia digital, y otros sui generis. También se observa que la literatura los emplea indistintamente, más allá de las diferencias conceptuales, en ocasiones se debe al desarrollo e implementación tecnológica que comenzó con gradualidad y progresividad y se aceleró a partir del año 2020. En el ámbito nacional, en la *Justicia federal* de acuerdo a las Acordadas 4/2020 y 31/2020 de la Corte Suprema de Justicia de la Nación la totalidad de los procesos son *íntegramente* electrónicos. Se observa que la digitalización del expediente en soporte papel fue una fase necesaria, pero que actualmente, en el ámbito judicial existen expedientes que nacen electrónicamente. A su vez, la firma digital le brinda características y presunciones de autenticidad, autoría e integridad.

En Argentina existe la decisión jurídica y de política judicial de construir un gobierno judicial abierto, transparente y con rendición de cuentas que se encuentra en transformación hacia un ecosistema judicial digital (Avila Paz de Robledo, 2023). Se propicia la constitución de un *ecosistema de justicia digital y algorítmico* como un término que permita integrar la interconexión, la interoperabilidad entre plataformas tecnológicas y la dinámica metamorfosis tecnológica y algorítmico. Ello viene dado en función que en el ámbito judicial se encuentra en ejecución numerosas experiencias de IA y de inteligencia artificial generativa. También se observa que la reglamentación viene dada principalmente por la actividad de las Cortes Supremas de Justicia, así como del Consejo de la Magistratura, tanto en el ámbito nacional como en el ámbito de las justicias locales.

3.1. Implementación de indicadores de gestión judicial

Como expresa Ana Dobratinich, fue en el año 1949 que se publicaba el artículo de Lee Loevigner "Jurimetrics: the next step forward" —inmediatamente después de tres años que se invente la primera computadora digital electrónica ENIAC— formalizando el uso de las computadoras en el espacio del derecho y esos problemas iniciales de la jurimetría fueron reemplazados poco a poco (Ana Dobratinich, 2022, pág. 207). Quizás la complejidad o

quizás el procesamiento en la máquina de algo que aún operaba bajo la lógica de un expediente en soporte papel hicieron de su desarrollo algo inesperado. Las condiciones tecnológicas emergentes actuales con la inteligencia artificial permiten formas de interacción en un espacio virtual que va desde la celebración de contratos hasta la administración de justicia, y ha transformado las prácticas y reglamentaciones, con la digitalización de documentos, la automatización de tareas, y más aún, esta metamorfosis no se detiene.

A partir de 2025, el Consejo de la Magistratura de la Nación desarrolló una innovación en el sistema de gestión judicial (SGJ-LEX100) destinada a magistrados y funcionarios al incorporar indicadores de gestión que permiten recuperar y visualizar información clave para monitorear al sistema de justicia y en forma particular indicadores sobre: i) expedientes en trámite; ii) registro del libro de entrada; iii) movimientos actuales; iv) y sentencias registradas.

Esta información actual y confiable, que se genera de modo automatizado con el uso de herramientas informáticas, juega un rol clave para la toma de decisiones, monitoreo de gestión y oportunidades de mejora. Ahora como observa Harari, "para investigar los algoritmos, las instituciones reguladoras necesitarán no solo analizarlos, sino también traducir sus descubrimientos en relatos que los humanos puedan entender" (Harari, 2024, pág. 384), es decir que esta información además requiere ser comprensible para poder emplearla. Está claro que así el sistema de gestión genera información cuantitativa en base a variables estadísticas orientadas a una gestión más ágil, eficiente y transparente. Ofrece datos actuales que pueden alimentar el *design thinking* del ecosistema de justicia digital y algorítmico a partir de las prácticas cuantificadas. Posibilita generar datos que tornan factible el control de la ciudadanía y el fortalecimiento institucional.

3.2. Política de comunicación judicial clara como un eje que sostiene el ecosistema

La comunicación desempeña un rol protagónico en esta era de la información. Desde luego, en el ámbito de la Corte Suprema de Justicia de la Nación se decidió a través de la res. 2512/2024 crear el Portal de la Justicia Argentina (Corte Suprema de Justicia de la Nación Argentina, 2025) a los fines de optimizar la accesibilidad, la transparencia y la eficiencia del sistema judicial en Argentina.

En tanto política pública judicial de comunicación se desarrolla a partir del postulado de gobierno abierto y acceso a la información pública permitiendo a la ciudadanía y a operadores conocer y tener una visión integral de la Justicia en Argentina. A su vez, articula diversas jurisdicciones con el propósito de generar un acceso equitativo a la información y desarrollar la confianza en la administración de justicia.

Se destaca que uno de los ejes que se comunica es Inteligencia artificial. Puntualmente, comunica noticias como novedades y prácticas y herramientas empleadas en todo el país. De esa manera, al recoger y articular la información de portales de internet oficiales de las justicia federal y justicias locales, permite construir una democracia deliberativa que conozca, comprenda y participe de la construcción del ecosistema de justicia digital y algorítmico.

3.3. Inteligencia artificial como estrategia de lenguaje claro y accesible de la jurisprudencia

A partir de la entrada en vigencia de lo dispuesto por la Acordada 10/2025 de la Corte Suprema de Justicia de la Nación se implementó en la plataforma de jurisprudencia de los tribunales federales y nacionales que se publican en el sitio web de la Corte, la inteligencia artificial genera automáticamente un resumen con lenguaje claro y accesible (2025). Esta medida está dirigida a todos los operadores que acceden a las sentencias desde la página de la CS.

De acuerdo a la Federación Internacional del Lenguaje Claro, un escrito "está en lenguaje claro si su redacción, estructura y diseño son tan transparentes que los lectores a los que se dirige pueden encontrar lo que necesitan, entender lo que encuentran y usar esa información" (Marazzato Sparano, 2021, p. 4). Es decir, que el uso de esta herramienta de lenguaje claro tiene un efecto directo en la capacidad para conocer la información judicial y tomar parte, es decir participar como ciudadanos sin depender de como expresa Böhmer los abogados como "igualadores retóricos" (2008) en tanto traductores del lenguaje técnico al lenguaje de la ciudadanía.

4. Jurisprudencia sobre el uso de la IA generativa por la judicatura y la abogacía

En este apartado se recoge dos casos paradigmáticos de uso de la IA, uno por parte de un juez y otro por parte de un abogado. Siendo dos casos de jurisprudencia locales y de los primeros en la materia, sirven como aprendizaje continuo.

4.1. Abogacía y uso de IA generativa: Caso "Giacomino"

La Sala II de la Cámara de Apelación en lo Civil y Comercial de Rosario a cargo del magistrado Oscar R. Puccinelli, en agosto de 2025, resolvió en el Caso "Giacomino", que el abogado de la parte actora empleó IA para la redacción de un escrito que subió al expediente con "dos citas jurisprudenciales" que favorecían su postura, y luego se retractó de dichas citas porque habían sido generadas por la IA generativa: en la sentencia el tribunal expresa que "[p]asados los autos a resolver, y luego de analizados los argumentos de las partes, se advirtió que ciertas citas jurisprudenciales realizadas

por el letrado de la actora no pudieron ser halladas pese a haberse realizado una búsqueda exhaustiva, razón por la cual se dictó una medida para mejor proveer, solicitando al letrado que identifique la fuente de tales citas. Al contestar el requerimiento, el letrado aclaró que utilizó un sistema de inteligencia artificial y que volcó el resultado encontrado de buena fe, de modo que resulta claro que las citas mencionadas no fueron chequeadas por tal curial" (Giacomino, Cesar Adrián y otros c/ Monserrat, Facundo Damián y otros s/ daños y perjuicios, 2025).

El tribunal consideró que la responsabilidad profesional del abogado se encontraba comprometida no solo ante el tribunal por introducir citas de jurisprudencia generadas por la IA, sino también en cuanto a la relación profesional con su cliente en tanto no había un consentimiento informado válido que habilite al letrado a introducir la información del caso en un *chatbot* de inteligencia artificial generativa. Explica el tribunal en el fallo, que el letrado tenía un deber de cotejar la cita que respaldaba su argumentación. Puesto que los *chatbots* en sus términos y condiciones informan que se trata de un sistema experimental y que sus resultados deben ser cotejados.

El considerando 6 expresa sobre "las citas que el letrado acompañó sin chequear en su pieza recursiva y que alega haberlo hecho de buena fe, por los motivos expresados supra y sin que esto implique una sanción por parte de este tribunal, corresponde hacer un llamado de atención al letrado en cuestión y, además, a fin de prevenir que este tipo de situaciones se generalice en el fuero, anoticiar al Colegio de Abogados de Rosario del contenido de esta resolución a fin de que adopte las medidas que correspondieren, especialmente las enderezadas a poner en conocimiento a los colegiados acerca de la problemática y de los riesgos que implica la utilización de este tipo de herramientas sin chequear los resultados que ofrece en la labor preparación de documentos que se presenten ante los tribunales" Y en su resolución se oficia al Colegio de Abogados de Rosario "a los fines de que tome conocimiento de la problemática suscitada por el uso irreflexivo de los resultados de consultas a *chatbots* con inteligencia artificial generativa" (2025).

De este caso, surge el deber indelegable de verificar las fuentes cuando se utilice IA generativa. A su vez la sentencia advierte que su uso requiere responsabilidad y criterio profesional. Se valor como positivo que no hay una censura o prohibición del uso de la IA generativa, sino más bien un llamado de atención para tomar consciencia, no solo a nivel individual sino también a nivel colegial.

4.2. Judicatura y uso de la IA: Caso "Providencia del Chubut c/P.R.A"

En Chubut, la Cámara en lo Penal de Esquel anuló una sentencia de condena penal dictada por un tribunal de primera instancia en lo penal por

vulnerar el deber fundamentación de la sentencia así como el principio de transparencia pues utiliza IA generativa para la redacción del acto procesal de decisión de condena sin cumplir con "Directivas para el uso ético y responsable de inteligencia artificial generativa en el Poder Judicial de Chubut" elaboradas por el Tribunal Superior de Justicia de la Provincia de Chubut por el Acuerdo Plenario N° 5435 de fecha 1 de abril de 2025 (Providencia del Chubut c/ P.R.A. , 2025).

En el considerando del fallo de Cámara, en el voto de la Magistrada Carina Paola Estefanía, se expresa "al acudir a los argumentos vertidos por el juez en su sentencia —respecto de los cuales la Defensa apoya su queja—, hemos advertido que en el renglón anterior al inicio del Punto IV, titulado 'Resolución Sobre las cuestiones probatorias planteadas por la Defensa' se puede leer: 'Aquí tienes el punto IV reeditado, sin citas y listo para copiar y pegar' (*sic*). Dicha frase da cuenta, objetivamente, que el magistrado ha utilizado un asistente de inteligencia artificial (IA) generativa de aquellos que se utilizan para realizar tareas como escribir, planificar, aprender y explorar ideas y se caracteriza por su capacidad para procesar y comprender información de diversas fuentes, como texto, imágenes y videos, funcionando con un enfoque en el aprendizaje automático y el procesamiento del lenguaje natural. Se desconoce qué asistente utilizó y cuál fue el requerimiento realizado, el que podría inferirse de la respuesta que le ha brindado el asistente artificial que ha sido volcada en la sentencia -antes citada- y tiene relación con la resolución de planteos nulificantes de la Defensa, respecto de las dos pruebas de cargo más relevantes de estos actuados".

Sucede que en abril de 2025 el Tribunal Superior de Justicia de la Provincia de Chubut aprobó las "Directivas para el uso ético y responsable de inteligencia artificial generativa en el Poder Judicial de Chubut" con el propósito de:

i) Fomentar el uso ético y responsable a los fines de mitigar los riesgos inherentes al emplear IA generativa (v.gr. sesgos, alucinaciones, falta de transparencia, dependencia excesiva y sobreconfianza y discriminación).

ii) Garantizar el control humano, garantizando que toda decisión judicial siga estando bajo la responsabilidad de magistrados/as y funcionarios/as competentes. En esta recomendación expresamente dice "[d]e ninguna manera puede delegarse la toma de decisiones en la IAGen".

iii) Proteger datos personales y confidenciales, desplegando herramientas de anonimización cuando sea necesario.

iv) Promover capacitación y actualización continua a los fines de garantizar un uso efectivo y seguro de la IA generativa.

A su vez, valora como antecedente que la Comisión Iberoamericana de Ética Judicial en su 24° dictamen referido al uso ético de la inteligencia ar-

tificial en la labor jurisdiccional en el que se afirma que "[l]a implementación de los sistemas de IA puede mejorar la eficiencia y ser de ayuda en los procesos judiciales. No obstante, debe mantenerse la supervisión o revisión humana para prevenir cualquier funcionamiento indebido o desviado que pueda afectar a la adecuada prestación del servicio judicial" y establece un deber de transparencia: "para un servicio público de justicia en una sociedad democrática no se recomiendan los sistemas de 'cajas negras', es decir aquellos en los que se incluyen datos de los que no se conoce su procesamiento".

En este caso, la Cámara invoca un antecedente jurisprudencial de la Provincia de Río Negro ante la verificación de citas inexistentes en los escritos de las partes un Tribunal de segunda instancia sentenció "corresponde hacer un severo llamado de atención a los abogados en cuestión por cuanto, grave es haber utilizado las herramientas de la IA sin el control responsable, pero más aún lo es el inventar deliberadamente fallos, citas y jurisprudencia en general que resultan inexistentes" (Cámara de Apelaciones en lo Civil, Comercial, Familia y Minería, Gral Roca. Sentencia 199/2025 "MJL c/Peugeot, Citroen Argentina S.A y otra s/ sumarísimo").

Este caso, enseña que el uso de IA generativa por la judicatura requiere el control humano y más aún, la responsabilidad a cargo del operador jurídico que la emplea. Como se expresó el desvío de las directivas de control humano y de transparencia vulneran las garantías de debido proceso digital, así como de la defensa en juicio y del juez natural en tanto la función judicial del juez es indelegable, y el uso de la IA generativa ha de seguir la reglamentación vigente y las normas de ética judicial.

5. Conclusiones

En lo personal se propicia el tratamiento jurídico de la inteligencia artificial a partir a partir de un marco regulatorio centrado en la persona humana, en su dignidad y en la protección de los derechos humanos. Antes de poder hablar de un marco jurídico claro para el ámbito judicial, es preciso que se dé la deliberación en relación a un marco regulatorio de la inteligencia artificial en general.

A su vez, se observa que, en Estados como Argentina, en la que el federalismo permite que las partes y el todo convivan armónicamente respetando la diferencia en la igualdad y generando marcos regulatorios autónomos, se considera necesaria una deliberación que abarque el gobierno multinivel para armonizar marcos regulatorios, garantizando no solo la interoperabilidad tecnológica entre plataformas sino también la jurídica. Más aún, el cómo que es la propuesta a la que se apunta es a través de una ley modelo para el tratamiento general de la IA y una ley modelo especial destinada al tratamiento jurídico procesal, gestión judicial y del ejercicio profesional de la abogacía.

El horizonte del ecosistema judicial digital y algorítmico se encuentra unido intrínsecamente a un proceso de reforma indefinida. Así como en tiempos del mito, Sísifo cada día debía volver a subir la roca por la montaña, cada innovación tecnológica, digital, algorítmica requiere un ajuste que va más allá del horizonte de la norma positiva para incluir la praxis, la dimensión axiológica, cultural y social.

El enfoque de política pública sobre la Justicia permite analizar en forma sistémica desde el diseño hasta su monitoreo. Aplicar la inteligencia artificial para generar información para tomar decisiones en base a un escenario actual es una innovación muy positiva en Argentina. Así como utilizar la IA para generar un lenguaje claro y accesible de la información judicial y como política comunicacional fortalecen los pilares democráticos al potenciar el acceso a la información judicial.

La jurisprudencia novedosa que relevamos sobre el uso de la inteligencia artificial generativa por parte de la judicatura y de la abogacía, revela que el desafío es compartido. Que la responsabilidad es un espejo que abarca a todos los actores del sistema. Y más aún permite una interpretación innovadora de los derechos y garantías procesales a la luz de la IA con enfoque de derechos humanos.

Se propicia la elaboración de una ley modelo de la inteligencia artificial centrada en la persona humana y su dignidad, con la protección de los derechos humanos y bajo el control y supervisión humana, que permita la trazabilidad, explicabilidad y transparencia, y para el bien común. En definitiva, esta ley modelo que se propone es la vía para la armonización procesal de la IA en la Justicia con miras a realizar la humanización del proceso judicial.

Bibliografía

Ana Dobratinich, G. (2022). Inteligencia artificial y justicia: aplicabilidad de la tecnología en las decisiones judiciales en Argentina. *Revista Direitos Culturais*, 17(42), 203-216. doi: http://dx.doi.org/10.20912/rdc.v17i42.761

Asamblea General de la Organización de las Naciones Unidas. (2024). *El Pacto del Futuro*. Naciones Unidas A/RES/79/1 (págs. 1-64). Nueva York: ONU.

Asamblea General de la Organización de las Naciones Unidas. (2025). *Resolución aprobada por la Asamblea General de la Organización de las Naciones Unidas el 26 de agosto de 2025*. Naciones Unidas A/RES/79/325 (págs. 1-4). Nueva York: ONU.

Atienza, M. (2025). La dogmática jurídica como tecno-praxis. En G. B. Manuel Atienza & P. G. Josep Aguiló Regla (Eds.), *La dogmática jurídica como tecno-praxis* (págs. 17-54). Lima, Perú: Palestra.

Avila Paz de Robledo, R. A. (directora). (2005). *Manual de Teoría General del Proceso*. Córdoba: Advocatus.

Avila Paz de Robledo, R. A. (2023). Actos procesales electrónicos en el fuero civil nacional y local. *Anuario del Centro de Investigaciones Jurídicas y Sociales*, XXI, 217-233.

Avila Paz de Robledo, R. A. (2024). Proceso judicial digital e inteligencia artificial: nuevos horizontes. *Anuario del Centro de Investigaciones Jurídicas y Sociales*, XXII, 130-149. Recuperado el 20 de 09 de 2025, de https://revistas.unc.edu.ar/index.php/anuariocijs/article/download/44571/44746

Avila Paz de Robledo, R. A. (2023). Formación de operadores en oralidad y procesos por audiencias atendiendo el acceso a justicia de colectivos vulnerables, para construir "sociedades justas, pacíficas e inclusivas" -ODS 2030-. En A. A. Procesal, *XXXI Congreso Nacional de Derecho Procesal*. San Juan: Asociación Argentina de Derecho Procesal.

Benzal Medina, E. (2025). *El tratamiento jurídico de la inteligencia artificial*. Barcelona: Atelier.

Böhmer, M. (2008). Igualadores y Traductores. La Ética del Abogado en una Democracia Constitucional. En M. Alegre, R. Gargarella & C. Rosenkrantz, *Homenaje a Carlos S. Nino*. Buenos Aires: La Ley, Facultad de Derecho de la Universidad de Buenos Aires.

Bourdieu, P. (2001). La fuerza del derecho. Elementos para una sociología del campo jurídico. En P. Bourdieu, *Poder, derecho y clases sociales*. Bilbao: Desclée de Brouwer.

Cappelletti, M. & Garth, B. (1977). *Access to justice: the newest wave in the worldwide movement to make rights effective*. *Buffalo Law Review*, 27, 181.

Congreso de la Nación Argentina. (2025). *Constitución de la Nación*. Capital Federal, Argentina: Boletín Oficial de la República Argentina.

Convención Constituyente de la Provincia de Santa Fe. (2025). *Constitución de la Provincia invencible de Santa Fe. Reforma 2025*. Convención Constituyente de la Provincia de Santa Fe (págs. 1-76). Santa Fe: Boletín Oficial de la Provincia de Santa Fe. Recuperado el 01 de 10 de 2025, de https://www.santafe.gob.ar/ms/reforma-constitucion/wp-content/uploads/sites/86/2025/09/CONSTITUCION-DE-LA-PROV-DE-SF-2025_VF-WEB.pdf

Corte Suprema de Justicia de la Nación Argentina. (2025). *Acordada 10/2025*. Ciudad de Buenos Aires: CSJN.

Corte Suprema de Justicia de la Nación Argentina. (30 de 09 de 2025). *Justicia Argentina*. Obtenido de Justicia.ar: https://justicia.ar/

Giacomino, C. A. y otros c/ Monserrat, F. D. y otros s/ daños y perjuicios, *100677143l0* 21-11893083-2 (Cámara de Apelación en lo Civil y Comercial de Rosario, Sala II Agosto de 2025).

H. Legislatura de la Provincia de Formosa. (2025). *Constitución de la Provincia de Formosa*. Formosa: Boletín Oficial de la Provincia de Formosa.

H. Legislatura de la Provincia de Jujuy. (2023). *Constitución de la Provincia de Jujuy*. San Salvador: Boletín Oficial de la Provincia de Jujuy.

H. Legislatura de la Provincia de La Rioja. (2024). *Constitución de la Provincia de La Rioja*. La Rioja: Boletín Oficial de la Provincia de La Rioja.

Harari, N. Y. (2024). *Nexus. Una breve historia de las redes de información desde la Edad de Piedra hasta la IA* (J. Ros, Trad.). Buenos Aires: Debate.

Mercado Común del Sur. (2023). *Declaración de Ministros y Altas Autoridades sobre Derechos Humanos de los Estados Partes del MERCOSUR sobre los principios de derechos humanos en el ámbito de la inteligencia artificial*. XLII Reunión de Altas Autoridades sobre Derechos Humanos del MERCOSUR (págs. 1-8). Brasilia: MERCOSUR.

Nieva Fenoll, J. (2018). *Inteligencia artificial y proceso judicial*. Madrid: Marcial Pons Ediciones Jurídicas y Sociales S.A.

Organización de las Naciones Unidas. (01 de 10 de 2025). *Organización de las Naciones Unidas*. Obtenido de Inteligencia artificial: https://www.un.org/es/global-issues/artificial-intelligence

Oteiza, E. (2019). Derecho procesal comparado. Perspectiva contextual desde América Latina. En G. P. Posada, *Justicia y proceso en el Siglo XXI. Desafíos y tareas pendientes* (págs. 653-666). Lima, Perú: Palestra.

Parlamento Europeo y del Consejo de la Unión Europea. (2024). *Reglamento (UE) 2024/1689. Reglamento de la Inteligencia Artificial. Diario Oficial de la Unión Europea*. Recuperado el 30 de 09 de 2025, de https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=OJ:L_202401689

Pereira Campos, S. (2025). *Hacia una perspectiva sistémica y de política pública*. Lima: Palestra.

Poder Legislativo de la Provincia de Santa Fe. (2024). *Ley N° 14.384* (págs. 1-10). Santa Fe: Boletín Oficial de la Provincia de Santa Fe.

Providencia del Chubut c/ P.R.A., Carp. OFIJU NIC No 6209 – Leg. Fiscal N° 59560 (Cámara en lo Penal de la Circunscripción de Esquel 15 de 10 de 2025).

Robledo, D. (2023). Inteligencia artificial en el proceso de mediación y los sesgos algorítmicos de género. *Revista de Derecho Procesal*, (2), 379-408.

Taruffo, M. (2006). El proceso civil de "civil law": Aspectos fundamentales. *Revista Ius et Praxis*, 12(1), 69-94. doi:http://dx.doi.org/10.4067/S0718-00122006000100004

UNESCO. (2022). *Recomendación sobre la ética de la inteligencia artificial adoptada el 23 de noviembre de 2021* (págs. 1-45). París: Organización de las Na-

ciones Unidas para la Educación, la Ciencia y la Cultura. Recuperado el 01 de 10 de 2025, de https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

UNESCO. (01 de 10 de 2025). *Recomendaciones éticas*. Obtenido de Inteligencia artificial: https://www.unesco.org/es/artificial-intelligence/recommendation-ethics

Apuntes sobre la inteligencia artificial aplicada a la administración de Justicia [*]

Por Sonia Gamboa [**], *Andrés Botero* [***] *y Kevin Valdivieso* [****]

Resumen: Este escrito hace reflexiones generales sobre qué es y cómo funciona la inteligencia artificial (IA), así como su utilidad para el campo jurídico (IA-D) y el judicial (IA-J). Además, indica que ante la IA-J hay miedos alimentados por la ignorancia; sin embargo, es una tecnología imparable que ya está en operación, aunque, por el momento, de forma limitada. Resulta que muchos sistemas de IA-J ya están en condiciones de sentenciar casos, pero, hasta el momento, el uso que se les da es fundamentalmente de auxiliadores de los operadores judiciales, cosa que seguramente cambiará a corto plazo dado las ventajas que la IA reporta (en especial, porque incrementa la seguridad y la certeza jurídicas, que tanto se han perdido en las últimas décadas, y porque sería una potente herramienta en la lucha contra la corrupción de alto nivel y la delincuencia organizada), por un lado, y una vez las personas conozcan más de su forma de actuar, por otro. Esto no impide que se ejerza un pensamiento tecnológico crítico que alimente una regulación necesaria de la IA, la IA-D y la IA-J, regulación que, en épocas de globalización, debe ser más mundial que nacional.

Palabras clave: Inteligencia artificial; derecho; justicia; jueces; tecnología.

Abstract: This paper makes general reflections on what Artificial Intelligence (AI) is and how it works, as well as its usefulness for the legal field

---

(*) Resultado del proyecto de investigación 4233, financiado por la Universidad Industrial de Santander (UIS). Este artículo se realizó con base en la metodología de investigación documental-bibliográfica (Botero, 2016), con enfoque hermenéutico crítico. Igualmente este texto parte de uno ya publicado: (Botero et al., 2024).

(**) Doctora en Educación por la Universidad Pedagógica Nacional. Ingeniera de sistemas y magíster en Informática por la UIS. Profesora titular de la Escuela de Ingeniería de Sistemas e Informática y miembro del grupo de investigación Sistemas y Tecnologías de Información –STI– de la UIS. ORCID: http://orcid.org/0000-0002-3563-608X. Correo electrónico: scgamboa@uis.edu.co.

(***) Doctor en Derecho por la Universidad de Buenos Aires (Argentina) y doctor en Derecho por la Universidad de Huelva (España). Abogado y filósofo. Profesor titular de la Escuela de Filosofía y miembro del grupo de investigación Politeia de la UIS. ORCID: http://orcid.org/0000-0002-2609-0265. Correo electrónico: aboterob@uis.edu.co.

(****) Filósofo de la UIS. ORCID: https://orcid.org/my-orcid?orcid=0009-0003-8473-4908. Correo electrónico: kevinarley1967@gmail.com.

(AI-Law) and the judicial field (AI-J). Also, this text indicates that in the face of AI-J there are many fears fed by ignorance. However, it is an unstoppable technology that is already in operation, although, for the moment, in a limited way. It turns out that many AI-J systems are already in a position to sentence cases, but, so far, the use they are given is fundamentally auxiliary of judicial operators, which will surely change in the short term given the advantages that AI It reports (especially, because it increases legal security and certainty, which has been lost so much in recent decades, and since it would be a powerful tool in the fight against high -level corruption and organized crime), on the one hand, and once people know more about their way of acting, on the other. This does not prevent a critical technological thought from being exercised that feeds a necessary regulation of AI, AI-Law and AI-J, regulation that, in times of globalization, must be more worldwide than national.

Keywords: Artificial intelligence; law; justice; judges; technology.

1. INTRODUCCIÓN

La reflexión en torno al avance tecnológico dentro del derecho y a sus efectos en la rama judicial es algo que se ha hecho desde la génesis misma de la tecnología computacional. Por ejemplo, la primera publicación que analizaba el uso y el efecto que pueden tener los ordenadores en el panorama judicial fue tres años luego de la invención de la primera computadora digital electrónica ENIAC [(1)]. También, mucho antes de los avances computacionales, ya se implicaba el derecho con los diferentes avances tecnológicos a través de la historia, por ejemplo, en lo que Pérez (2004) llama paradigmas tecno-económicos [(2)]. Con el paso del tiempo, en cuanto la tecnología computacional se estableció como la herramienta fundamental para cualquier proceso humano, en el campo de lo judicial se concretaron esfuerzos por formalizar procedimientos, primero, en las décadas de los 70 y 80 del siglo XX, aunque no hubo propiamente desarrollos de software significativos en el área, sí se establecieron directrices para el análisis descriptivo de datos y modelos estadísticos de sentencias, que buscaban reducir disparidades

(1) La publicación fue "Jurimetrics: the next step forward" de Loevinger (1949). Ana (2021) dice que este camino se sigue recorriendo, pues con cada avance tecnológico-digital, las preguntas y los problemas previos se reemplazan conforme a las demandas de cada época.

(2) La autora propone cinco revoluciones tecnológicas sucesivas desde 1770 hasta el 2000, con su respectivo detonante: i) la revolución industrial en 1771, con la apertura de la hilandería de algodón de Arkwright en Cromford; ii) la era del vapor y los ferrocarriles en 1829, con la prueba del motor a vapor Rocket para el ferrocarril Liverpool-Manchester; iii) la era del acero, la electricidad y la ingeniería pesada, con la inauguración de la acería Bessemer de Carnegie en Pennsylvania en 1875; iv) la era del petróleo, el automóvil y la producción en masa, con la salida del primer modelo "T" de la planta Ford en Michigan en 1908; y v) la era de la informática y las telecomunicaciones, con el anuncio del microprocesador Intel en California, en 1971 (Pérez, 2004, p. 79). La autora introduce el término "lógica de lo nuevo" para referirse a la adaptación (en el transcurso del tiempo) de cada período.

en las decisiones judiciales. En los 90, empezaron a operar plataformas como Westlaw y LexisNexis, que permitieron la investigación legal, a partir de bases de datos estructuradas, lo que representó una de forma más rápida y eficiente de acceder a los precedentes judiciales y a la normativa vigente. En la década del 2000 se introdujeron los algoritmos de evaluación de riesgo criminal, como COMPAS, que permitían calcular una probabilidad de reincidencia de acusados, como base para tomar decisiones sobre sentencias. Al tiempo que el desarrollo teórico en ciencias cognitivas y computacionales ha potenciado el alcance de la inteligencia artificial (IA), de tal manera que cada vez puede simular más procesos que se consideraban muy complejos o exclusivamente humanos, estos desarrollos incipientes de *software* para la administración de justicia se han potenciado con cada nuevo avance de la IA.

Con base en lo anterior, este texto se divide en dos secciones. La primera, a partir de las preguntas orientadoras: ¿qué es una IA? y ¿qué capacidades tienen las IA hoy en día?; y la segunda, a partir de las preguntas ¿qué desarrollos de IA se usan hoy en el mundo jurídico?, ¿cómo ha afectado o podrá afectar la IA al mundo jurídico, en general, y el judicial, en particular?, y ¿qué alternativas de mitigación de riesgos y de aprovechamiento máximo de la IA pueden aplicarse? Finaliza el escrito con las respectivas conclusiones y bibliografía.

2. La IA. Nociones básicas.

Como se sabe, la IA está cada vez más presente no solo en las vidas cotidianas, sino en sectores de alto impacto como el médico, el militar, el aeroespacial, el político y el jurídico, entre otros. Los dispositivos móviles constantemente ofrecen información sobre rutas óptimas, estado del tiempo, noticias y lanzamientos de las preferencias de sus propietarios; además, permiten estar en contacto permanente con todo tipo de asuntos de interés personal o laboral, al tiempo que almacenan los movimientos, las preferencias, los hábitos y hasta los estados de ánimo de sus usuarios, para ofrecer información que pudiera ser de interés y se ajuste a algún modelo comercial o político que haya pagado por esos datos. Para ello se vale de tecnologías como sistemas de IA generativa, algoritmos de procesamiento de señales, sistemas predictivos, etcétera, todos ellos construidos sobre modelos de IA cada vez más acertados, eficientes, intuitivos y amigables. Se está frente al auge de fenómenos que tres décadas atrás eran parte de la ciencia ficción, como la popularización de vehículos autónomos de Tesla y la normalización de establecer vínculos afectivos con sistemas de inteligencia artificial conversacional, por mencionar algunos casos. Pero ¿qué se entiende por IA?

Russell y Norvig (2004) definen la IA como "el estudio de los agentes que reciben percepciones del entorno y llevan a cabo acciones" (p. XIX). Percepción se refiere a la información de determinado tipo que la IA puede

recibir del entorno, mediante algún mecanismo o interfaz diseñado para este fin. Así mismo, la IA, de manera autónoma, lleva a cabo acciones o da respuestas al entorno como resultado de procesar la información recibida, para cumplir con el objetivo para el que fue diseñada; a lo anterior se le conoce como función del agente o del dispositivo (Gamboa, 2020, p. 118). Este agente se considera inteligente por la capacidad de responder o adaptarse autónomamente a las circunstancias que encuentra en su entorno, lo cual, si bien no es una capacidad exclusivamente humana, sí ha resultado de interés para la computación automatizar, principalmente, procesos inteligentes humanos. Vista como dispositivo (como un desarrollo concreto y dispuesto en el mundo), la IA cumple el objetivo para el cual fue diseñada, mediante algoritmos, que son líneas de comandos que ejecuta para producir información o acciones, llamadas *outputs*, a partir de la información que recibe del entorno, o *inputs*. En palabras de Maranhão et al. (2021), un "algoritmo es cualquier procedimiento de ordenador bien definido que posee algún valor agregado en la calidad de sus entradas (inputs), generando otros valores en la salida (outputs), de manera que puede ser considerada una herramienta para solucionar problemas" (p. 145).

También se ha definido la IA como una imitación de los procesos de la inteligencia humana (Alarcón et al., 2019); como una rama de la computación dedicada a crear máquinas que piensan y aprenden (Andreu, 2021); o como el desarrollo de un producto atribuible a procesos mentales artificiales (Cáceres, 2023) [3]. Cerdio (2021) atribuye tres características a los sistemas de IA:

1) La capacidad de mejorar su desempeño para la obtención de cierto objetivo mediante la adquisición, procesamiento y aplicación de datos, conocimientos o habilidades.

2) Que son sistemas diseñados para desempeñarse (i) de manera autónoma o (ii) mediante la asistencia de cierta intervención humana o (iii) mediante la modificación de ciertas reglas simbólicas o modelos; en todos los casos (i-iii), como la principal herramienta para alcanzar cierto objetivo.

3) La capacidad para interactuar adrede e influenciar su entorno (p. 70).

Por su parte, el adjetivo "artificial", según Simon (2006), denota un dispositivo "producido por el hombre" (p. 4) para imitar *lo natural*. En una mirada filosófica, aquí se considera preciso delimitar, con Gottfredson [4], el

(3) Russell y Norvig (2004), de igual forma, reconocen dos enfoques en las definiciones de IA: centrado en la conducta y centrado en los procesos mentales, a partir de los cuales proponen cuatro categorías: sistemas que piensan como humanos, sistemas que piensan racionalmente, sistemas que actúan como humanos y sistemas que actúan racionalmente (p. 2).

(4) "Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (Gottfredson 1997a, p. 13). "(…) highly general information-processing capacity

concepto de *inteligencia* a una capacidad mental de comprensión y adaptación, la misma que las IA parecen portar hoy en día. Cabe señalar que, como lo relatan Gallagher y Zahavi (2008, pp. 1-5), casi en simultaneidad con el origen de la computación, emergen las llamadas ciencias cognitivas como el referente teórico que impulsaría el nacimiento de la IA, y que se vale, entre otras disciplinas, de la tradición de la filosofía de la mente, que ha llegado a juzgar, más recientemente y desde diferentes enfoques, al cerebro a la luz de las capacidades computacionales del momento.

En cuanto la IA busca imitar procesos mentales humanos, cabe preguntar si pudiera llegar a destronar al ser humano del pedestal cognoscente. Aunque no se pueda saber a mediano plazo, la filosofía ofrece argumentos que dejan ver que aún no se está tan cerca de ese evento. Campos (2021), por ejemplo, argumenta en favor de la prevalencia de los humanos, pues la inteligencia y la intuición de las personas resultan del *sentido* que le dan a sus hábitos y experiencias, capacidad que las máquinas no poseen. Campos (2021) expone, incluso, la no existencia de alma o espíritu en los sistemas inteligentes, como garantía de que la IA no pondrá en riesgo la condición prevalente de la especie humana. Sin embargo, cada vez que una IA logra imitar tareas o procesos que se creían exclusivamente humanos, la respuesta a la pregunta por lo esencial o lo exclusivamente humano debe ser reelaborada. Si se creía, por ejemplo, que la producción de piezas artísticas y creativas era exclusivamente humana, la reciente exhibición de obras pictóricas y musicales, entre otras, creadas por IA ha mostrado que no. Empero, más que destronar a la humanidad, estas tecnologías la han liberado de tareas y oficios que exceden en fuerza, tiempo y capacidad de ejecución, lo que no impide que el uso de IA pueda resultar destructivo para la humanidad, como podría serlo en asuntos militares o con el uso indebido de información privada que capturan las IA [(5)], por mencionar dos casos que pueden alimentar el miedo a estas tecnologías.

Como ya se mencionó, los algoritmos contienen los comandos a ejecutar para lograr el objetivo para el cual fue diseñada la IA. Por ejemplo, para crear un sistema que juegue ajedrez, primero debe implementarse los algoritmos de los movimientos básicos de las fichas, luego los de jugadas compuestas por varios movimientos y su relación con el estado actual del

---

that facilitates reasoning, problem solving, decision making, and other higher order thinking skills" (Gottfredson, 1997b, p. 81).

(5) En el caso del reconocimiento facial, por ejemplo, en favor de la seguridad, se aduce que su "propósito principal es identificar a conocidos criminales o personas con una orden de arresto" (Domingo, 2021, p. 23), pero se crítica la vulneración de los derechos humanos que esto pueda generar. Sanabria et al. (2022) advierten de varios riesgos de la IA en cuestiones tales como: derecho a la intimidad, protección de datos personales, libre actuar del ser humano, libertad de expresión, derecho de la libertad de reunión y asociación, derecho a la igualdad, creación de perfiles étnicos, garantías personales, garantía de igualdad procesal, garantía de valoración razonable de la prueba, aplicación del principio de transparencia y publicidad de la prueba, principio de no autoincriminación y sistema penal que mire el hecho y no la peligrosidad de la persona.

oponente, etcétera. No obstante, si una persona juega ajedrez contra una máquina y la vence, puede pensar que las personas son más inteligentes que ella, a fin de cuenta es una creación humana, pero si esta máquina aprende mejores movimientos mientras juega y genera en su base de datos todas las jugadas posibles para deducir la mejor en cada momento, a un punto de volverse invencible, ¿podría decirse que la IA es más inteligente que las personas, al menos como jugadoras de ajedrez?

Esto lleva al paradigma epistémico de las IA de que aprenden por sí mismas, es decir, que están en capacidad de establecer relaciones entre los datos que reciben y almacenan, para cumplir mejor su objetivo; pero no lo hacen desde cero. Podría pensarse que Deep Blue, la computadora que derrotó al campeón mundial Kasparov en 1997, aprendió a jugar ajedrez por sí sola, pero realmente no contaba con esa capacidad, sino con la de calcular 200 millones de posiciones por segundo, a partir de una base de conocimiento y algoritmos de decisión que operaban para lograr su objetivo: ganar partidas de ajedrez, en condiciones de torneo estándar.

Hoy en día las IA operan con mecanismos *machine learning* (cuya más reciente generación es el *deep learning*, que supera las IA deterministas basadas en representación del conocimiento en reglas lógicas y simbólicas), que funcionan mediante la ejecución de modelos estadísticos y grandes volúmenes de datos que les permiten elegir, con bastante precisión, un curso de acción apropiado para los objetivos determinados para ellas, con la particularidad de que entre más datos capture la IA puede reconocer nuevos patrones que le permiten ajustar los parámetros del modelo para un mejor desempeño en el cumplimiento de tales objetivos. Es a esta capacidad a la que se refiere el aprendizaje de máquina o *machine learning* (Manrique, 2020).

En este aprendizaje [(6)], dependiendo del conjunto de datos con los que opera, es posible que el modelo se ajuste a ciertos sesgos no deseados. Es el caso de las investigaciones en biotecnología que buscan que una IA pueda diagnosticar ciertas enfermedades, pero si los *datasets* con los que se entrena el modelo corresponden a datos de otras regiones geográficas, es posible que el modelo sujete el diagnóstico a características ajenas a los pacientes de otra región. Estos sesgos, si bien no desvían las decisiones de la IA de los objetivos inicialmente planteados y pueden ser corregidos en versiones posteriores, hacen necesaria la supervisión constante de sus resultados, especialmente en sus primeras fases de aplicación. A este aspecto se suma la imposibilidad de que la IA se valga de una cierta *dación de sentido* propia que, en el caso del experto humano, contribuye a que la decisión tenga una

(6) Que puede ser supervisado (opera con un conjunto de datos etiquetados previamente), no supervisado (no cuenta con datos etiquetados previamente, pues su objetivo es reconocer patrones en el conjunto de datos) o por refuerzo (opera por prueba y error, y se le indica cuáles decisiones son acertadas).

más alta probabilidad de acierto, como la que ofrece, por ejemplo y hasta el momento, el médico especialista en el diagnóstico de una enfermedad en la que él tiene mucha experiencia, pues esta dación de sentido se asocia, en la construcción de sistemas expertos, a una comprensión intuitiva de la situación que responde a un conocimiento tácito resultado de la experiencia del experto, pero que no puede expresarse fácilmente, por lo menos no en este momento, en reglas formales o sistemas simbólicos, como lo explican Dreyfus y Dreyfus (1986) en su célebre conferencia "From Socrates to Expert Systems".

Con todo ello, puede afirmarse que, hasta ahora, la IA no está en capacidad de actuar con total autonomía, aunque en algunos casos (que se irán incrementando con el tiempo), sí ofrecen resultados más cercanos a lo deseado que los que lograría un ser humano sin esta intervención, por lo que, en cuanto esta experticia pueda representarse en reglas formales o sistemas simbólicos, podría esperarse que las aplicaciones IA lleguen a operar con mayor independencia del hombre, mas no necesariamente implica que se deba eliminar la continua supervisión humana sobre sus resultados.

En este panorama, ¿qué pasa, entonces, si un sistema inteligente se equivoca? ¿Cómo afrontar situaciones éticas, morales o jurídicas creadas por equivocaciones de la IA? ¿Cómo deben juzgarse? [7] ¿Cómo regular el desarrollo de las IA de cara a la preservación de los humanos y de los ecosistemas? Y ¿cómo se mitigan los riesgos desde el diseño y el desarrollo? Estas y otras preguntas, que anticipan el impacto ético de lo desconocido o lo no desarrollado aun, son tema recurrente en los textos sobre IA. Por ejemplo, Martino (2021) aborda una ética digital; Churnin (2011) analiza los retos éticos y jurídicos de la IA, y Otero (2023) se refiera a las implicaciones de la IA en los derechos fundamentales, por mencionar tres obras.

Según Martino (2021), la posibilidad de que las IA lleguen a operar con total autonomía para cumplir sus objetivos, que aprendan a resolver otros asuntos y que lleguen a comunicarse entre ellas, propicia el miedo evolutivo a lo desconocido y a lo novedoso [8]. No obstante, el miedo es una respuesta

---

(7) En el ámbito jurídico es objeto de debate si las IA pueden cometer ilícitos, si se les puede imputar delitos y deben responder por ello ante un juez (humano o artificial) o sobre quién recae la responsabilidad. Las consideraciones éticas al respecto hacen énfasis, por una parte, en la observación permanente sobre posibles sesgos que operen en la toma de decisiones, sobre la transparencia que "el algoritmo" debe ofrecer sobre la forma en que toma decisiones, entre otros aspectos de diseño; por otra parte, se ha buscado establecer modelos de responsabilidad que aluden a las empresas desarrolladoras, a diseñadores de software, al usuario, etc. Vale recordar que, por ejemplo, a los textos generados por las IA no se les puede registrar derechos de autor, por lo que tampoco se les puede atribuir plagio o responsabilidad alguna sobre lo que expresa el texto. Sobre estas y otras responsabilidades (jurídicas y morales) de la IA, leer a Churnin (2011) y Cerdio (2021).

(8) Miedo alimentado, especialmente, por la ciencia ficción contemporánea. Recuérdese las leyes de Asimov: "1. Un robot no puede hacer daño a un ser humano ni, por inacción, permitir que un ser humano sufra daño. 2. Un robot debe obedecer las órdenes de los seres humanos, excepto si esas órdenes entran en conflicto con la primera ley. 3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o la segunda leyes" (Narváez, 2023,

evolutiva exitosa hasta cierta medida, pues salvaguarda la especie ante lo extraño, que no tiene que ser necesariamente bueno; pero si el miedo termina por anquilosar la exploración-innovación, la especie se estanca. En consecuencia, es necesario un equilibrio entre miedo-exploración, para producir un beneficio general, incluso en estos temas.

3. La IA en el derecho (IA-D) y en lo judicial (IA-J)

En el campo jurídico, la IA es objeto de estudio desde diferentes enfoques. Por una parte, preocupan a quién se le asigna la responsabilidad de los actos de la IA-D y la IA-J. Por otra parte, la IA genera desconfianza y miedo en abogados y jueces, por las posibilidades de desplazamiento laboral, pues este campo parecer ser uno de los próximos espacios a ser colonizados por ella. Estos y otros riesgos han sido tomados en consideración cuando se evalúa el impacto del uso de IA en el derecho y en lo judicial. Así, se encuentran lecturas a favor [(9)], en contra e, incluso, lecturas escépticas. Por ejemplo, Contreras et al. (2021) se preguntan por el impacto que tendrá el desarrollo de la IA en la forma en que se ejerce la abogacía. Un aspecto positivo es que la IA ayuda a los operadores jurídicos a filtrar información relevante, haciendo el proceso más eficaz y en menos tiempo (Sánchez, 2019). Por su parte, Ramírez (2022) expone:

> "La IA tiene múltiples implicaciones en el mundo jurídico entre las que se destacan las siguientes: a) Proporciona una gran cantidad de conocimiento derivado del análisis de los datos y produce nuevo razonamiento que puede contribuir a la interpretación y análisis de los fenómenos. b) Conjuga y contrasta los datos cuantitativos y cualitativos para resolver problemas y emitir conclusiones que pueden verificarse por los individuos. c) Procesa datos que permiten predecir tendencias o posibilidades de respuesta ante fenómenos. d) Plantea buenas prácticas para adelantar investigaciones y posibilidades de cruce de información relevante en el ámbito criminal. Entre otros" (p. 144).

Las posturas negativas auguraban, hasta hace pocos años, que sería imposible que las IA pudiesen fungir como jueces emitiendo sentencias, es decir, resolviendo conflictos a partir de normas jurídicas. Siempre se aludía a problemas técnicos y asuntos humanistas-morales como límites infranqueables de una IA-J. Dentro de ese paradigma se hizo famoso el argumento de que las máquinas nunca podrán juzgar pues: carecen de emociones necesarias para hacer lo justo y no pueden argumentar, discernir sobre la

---

p. 70). La ciencia ficción plantea la posibilidad de que estas leyes sean violadas, emergiendo escenarios apocalípticos que aumentan la desconfianza sobre las IA.

(9) Ana (2021) dice: "quienes están a favor argumentan que no se trata de un reemplazo, sino de un auxilio desde el cual la tecnología, mediante su compilado de métodos, medios y capacidad resolutiva a gran escala, ayudaría a la toma de decisiones" (p. 6).

norma vigente, valorar la prueba, ni actuar por sí mismos en tanto solo hacen aquello para lo que fueron programados –y de hacerlo se le estaría entregando la rama judicial a los diseñadores de IA lo que atentaría contra su independencia–, etcétera. Otras posturas señalan la insuficiencia sintáctica del lenguaje de la IA en relación con el derecho [(10)]. Al respecto, Villegas (2023) expresa:

> "[e]xiste una barrera natural de la inteligencia artificial en el mundo jurídico y es el simple hecho de que las computadoras pueden procesar con números, pero no tan fácilmente con palabras. Más aun, el mundo jurídico se encuentra repleto de conceptos abiertos a múltiples interpretaciones, así como de construcciones legislativas que emplean una muy compleja estructura sintáctica, lo que dificulta aún más su procesamiento por la inteligencia artificial. Esta barrera natural ha impedido que esta tecnología impacte el mundo jurídico como lo ha hecho respecto de otras industrias" (p. 91).

Sin embargo, la IA-J está incrementándose cada día. ¿Por qué el interés a favor en la IA-J? Hay varios motivos: 1) ofrece una administración de justicia económica (entre otras cosas, porque reduce al máximo los costos de transacción que tiene la justicia humana en la actualidad) y altamente veloz (entre más rápido se den las decisiones judiciales, mejor para todos); 2) está siempre actualizada (tiene sistemas de captación de información en tiempo real, no solo en asuntos jurídico-normativos, sino también probatorios); 3) redacta y escribe mucho mejor que buena parte de los abogados; 4) cuenta con una gran capacidad argumentativa (por lo que tiene un alto poder convincente); 5) puede analizar de forma simultánea muchos criterios (tanto normativos como factuales), en tiempo real, con capacidad de ponderar y decidir sobre ellos; 6) ofrece más seguridad al usuario pues tiene altos niveles de neutralidad e imparcialidad, dentro de los límites mencionados con anterioridad como la supervisión humana, lo que aumenta su legitimidad y sirve de freno a la corrupción en la rama judicial; 7) evita exponer a los jueces humanos en casos difíciles, como serían los de corrupción de las altas esferas del Estado o en delitos contra organizaciones criminales transnacionales (¿cómo amenazar, extorsionar o sobornar a una máquina?); 8) ofrece respuestas más seguras y ciertas pues elimina muchos sesgos cognitivos, emotivos e ideológicos del juez humano (aumentados, por ejemplo, por teorías como el neoconstitucionalismo que hace muy difícil juicios de predictibilidad de la decisión [(11)]),

---

(10) Martino (2021) se refiere a que el método de oratoria en el razonamiento legal no es ajeno a las IA: "Algo que hay que decir, y que a menudo olvidamos, es que el razonamiento legal y retórico y no lógico tiene que ver con el deseo de convicción. El razonamiento legal es más retórico que lógico. Sería conveniente que los abogados y jueces practicaran una vieja asignatura de las universidades medievales: la oratoria. Hay un programa que se ocupa de esto: el Proyecto Debate, cuya base de conocimiento consiste en unos 10.000 millones de frases, tomadas de periódicos y revistas" (p. 24).

(11) Como se sabe, desde hace varias décadas, la cultura jurídica occidental ha hecho uso de dispositivos normativos (como los principios jurídicos y las reglas cada vez más casuísticas) y académicos (como el resurgimiento del iusnaturalismo y la entronización del neoconstitucionalismo)

lo que volverá más predecible el derecho; 9) dará cierta uniformidad a las decisiones judiciales, a diferencia de lo que sucede cuando son muchos jueces, cada uno con sesgos diferentes, los que resuelven casos similares, todo lo cual aumenta la certeza de las reglas jurídicas; 10) resuelve cualquier problema que se le somete (a diferencia de la tendencia a concentrar la justicia humana estatal en los asuntos de mayor cuantía, dejando las causas menores a sistemas con una alta tasa de ineficacia o ambigüedad); y 11) aumenta la legitimidad de la rama judicial y del Estado, pues los casos operativos de IA-J han obtenido altos niveles de satisfacción por parte de los usuarios; entre otros.

Ahora hay que preguntarse qué funciones trajo, tiene y podrá tener en un futuro el desarrollo de la IA en el campo judicial (IA-J). Resulta relevante considerar lo correspondiente a la ejecución/aplicación del derecho. Por ejemplo, si la IA puede emitir juicios con valor jurídico, redactar memoriales o sentencias, conseguir y valorar pruebas en favor o en contra de lo alegado por un actor procesal, considerar atenuantes o agravantes, considerar decisiones de casos similares anteriores, etcétera. Cualquier toma de partido en torno a la IA-J dependerá de si las máquinas podrán resolver mejor que los jueces humanos conflictos jurídicos mediante normas de carácter particular como lo son las sentencias.

Sobre la incursión que actualmente tiene la IA en el campo jurídico, el *Montreal AI Ethics Institute* (Faulkner, 2025) ofrece un análisis del rol de la IA en el sistema de justicia, principalmente, en Canadá, pero teniendo en cuenta las aplicaciones actuales de la IA a nivel global. Este informe hace referencia a cinco aplicaciones de la IA en el sistema de justicia, desde aquellas que cumplen funciones más operativas hasta las que constituyen un soporte para la toma de decisiones sobre sentencias y sobre elaboración de casos. Estas son:

1. Gestión de casos: IA que ejecuta tareas administrativas automáticamente y optimiza flujos de trabajo de los juzgados. Este tipo de IA-J puede programar audiencias, priorizarlas según la urgencia o la complejidad, seguir el progreso de casos y gestionar documentos legales, entre otras tareas.

---

que amplifican las respuestas jurídicas ante cualquier conflicto jurídico (Botero, 2006). En pos de la justicia (según lo que entienda por ella el juez o el académico) se han sacrificado certeza y seguridad normativas, generando ambigüedad en el sistema. Obviamente, siempre hay márgenes de ambigüedad en el derecho, en especial porque este es lenguaje y todo lenguaje tiene una zona de penumbra, sumado a que algunos grados de ambigüedad son útiles pues disminuye la obsolescencia de la norma escrita y le permite al juez resolver de mejor forma casos muy excepcionales. Pero lo que ha sucedido en los últimos tiempos es que la ambigüedad, la inseguridad y la incerteza jurídicas aumentan mucho más allá de lo deseable, lo cual genera efectos sociales muy peligrosos, en especial que el ciudadano no tiene certeza de qué debería hacer o no hacer antes de la conducta. Hace algunos años, en una entrevista hecha Guibourg, mencionada por Botero (2021), aquel dijo que la ambigüedad que producen las nuevas posturas iusnaturalistas y neoconstitucionalistas generará una contrarrevolución que llevará al derecho a su otro extremo: la búsqueda de una certeza y seguridad máximas, que también es peligrosa. La IA sería una respuesta tecnológica a esa búsqueda de una certeza y seguridad *útiles*, pero ya depende de la *regulación* que no se vuelva una reacción extrema, que la misma IA permita cierta ambigüedad que sería útil al sistema.

2. Investigación jurídica: IA que busca y recupera documentos jurídicos de forma rápida y precisa, los analiza y ofrece información relevante y recomendaciones como resultado. Este tipo de IA identifica patrones en procesos judiciales similares ya resueltos. También, esta IA-J evalúa y detecta inconsistencias en argumentos legales.

3. Evaluación de riesgos y predicción de reincidencia: IA que analiza grandes *datasets* sobre antecedentes penales, factores demográficos, patrones de comportamiento, etcétera, así como el historial (crediticio, delictivo, de salud, entre otros) de las partes procesales, para calcular, por ejemplo, un riesgo de reincidencia, lo que es muy relevante para el juzgador.

4. Recomendación de sentencias: IA que analiza precedentes legales, la normativa vigente, las pruebas aportadas y evaluaciones de riesgo de los juzgados, entre otros criterios, para sugerir a los jueces cómo sentenciar un caso concreto.

5. Análisis predictivo de resultados judiciales: IA que analiza patrones de miles de casos pasados para estimar la probabilidad de varios resultados legales, lo que se constituye en una herramienta que ayuda a los abogados a preparar mejores estrategias para sus casos, y a los jueces y a los legisladores a tomar decisiones basadas en una gran cantidad de datos sobre alternativas de resolución de conflictos que eviten juicios innecesarios y a asignar de manera más eficiente los recursos para los juicios.

Estas cinco funciones tienen en común que representan para el sistema ventajas como procesar grandes volúmenes de datos en poco tiempo con el fin de ofrecer análisis, patrones, predicciones que permiten una asignación más eficiente de recursos a los procesos, dando una mayor certeza sobre los posibles resultados en los juicios, eliminando juicios innecesarios y, especialmente, mitigando los efectos de sesgos cognitivos, emotivos e ideológicos y prejuicios de los juzgadores; pero, paradójicamente, en todos los casos se expresa la preocupación de que, en cuanto se desplaza a los humanos de sus roles, sí se mitigan los sesgos humanos, pero se imponen los sesgos inherentes a la programación inicial y a los datos con los cuales se han entrenado los algoritmos. Entonces, si no se garantiza transparencia en la operación de la IA y un seguimiento constante a sus resultados, es posible que, por una parte, en cuanto los humanos vayan confiando cada vez más en los algoritmos, vayan delegando sus funciones a la IA sin comprender del todo cómo ella opera o toma decisiones; y, por otra, que no se pueda salvaguardar los derechos de los acusados y garantizar la equidad en los procesos judiciales.

Este informe propone estrategias para mitigar estos efectos negativos que se prevén con el uso de la IA en procesos judiciales. Estas estrategias consisten en concebir marcos que ofrezcan desde pautas de diseño hasta regulaciones regionales e internacionales. En términos de diseño de las IA,

se recurre al concepto de inteligencia artificial explicable (XAI), que exige documentar cómo la IA toma las decisiones, de manera que el modelo sea comprensible y genere confianza para los humanos implicados: tanto usuarios directos como indirectos. Para mitigar posibles sesgos, se propone usar *datasets* representativos de diversos grupos poblacionales e implementar técnicas de aprendizaje automático conscientes de la equidad para detectar patrones discriminatorios y reponderar los datos de entrenamiento. Por tanto, es necesario implementar mecanismos de revisión y supervisión en las primeras fases, mediante los cuales los jueces humanos evalúen críticamente los resultados de la IA antes de emitir un fallo final. Se reitera la necesidad de poner en marcha las regulaciones en curso sobre diseño y uso de la IA.

Sin embargo, este informe del *Montreal AI Ethics Institute* corresponde a un desarrollo temprano de la IA-J, pues considera que esta solo sirve para auxiliar al juez humano (*Judicial Decision Support System*), cosa que puede cambiar prontamente si es que se le permite a la IA-J dirigir y sentenciar procesos judiciales por su propia cuenta, dado que ya se cuenta con algunas IA-J que lo pueden hacer, pero que no se les permite desde las instancias oficiales.

En el panorama colombiano, tanto la Fiscalía General de la Nación, con PRISMA, como la Corte Constitucional colombiana, con PretorIA, manejan aplicaciones de *software* desarrolladas con *machine learning*. Sobre PretorIA, Ramírez (2022) comenta:

> "En Colombia, los tribunales nacionales empiezan a resolver asuntos utilizando la inteligencia artificial con sistemas como pretorIA. Un sistema predictivo de detección inteligente de sentencias e información, su nombre se inspira en el pretor, auxiliar de la justicia romana, capaz de agrupar, analizar y clasificar información de las más de 2500 sentencias diarias que recibe la Corte Constitucional. La inteligencia artificial está reduciendo la demanda de abogados, al menos en lo que se refiere a ciertas tareas, que hoy por medio de la automatización se pueden hacer con mayor rapidez y precisión" (p. 150).

PRISMA, por su parte, tiene la función de predecir, a partir de su base de datos, si un sujeto puede llegar a reincidir en una actividad delictiva (Fernández, 2021). La base de datos almacena perfiles de individuos que ya están inscritos en el sistema penal acusatorio, de forma tal que el algoritmo de PRISMA ofrece como *output* una predicción porcentual de la probabilidad de reincidencia de un imputado, con base en "las características del individuo, del último evento criminal y de los antecedentes criminales de cada individuo" (p. 18), en comparación con la ocurrencia de haber reincidido o no de otros individuos con características similares. Para ello se vale de modelos predictivos computacionales, cuya validez ha sido probada con otros fenómenos de comportamiento similar. De esta manera, cuando un juez debe

determinar si otorga o no medida de aseguramiento a un imputado o qué tipo de sanción debe asignarle a un condenado, esta herramienta compara y complementa la información del proceso con lo disponible en la base de datos, auxiliando al operador judicial en su decisión de qué medida otorga, dependiendo del riesgo de volver a delinquir [(12)]. En palabras de Fernández (2021) la "inteligencia artificial es usada por la Fiscalía de Colombia para optimizar procesos, encaminar mejor los recursos y evitar cometer errores al momento de solicitar una detención preventiva" (p. 200), especialmente teniendo en cuenta que la reincidencia es uno de los mayores problemas que enfrenta la política criminal del Estado y un factor determinante para saber qué medida o sanción aplicar.

Según la Fiscalía, con PRISMA se busca:

1. Desde un punto de vista de eficiencia, reducir la reincidencia criminal y hacer un uso más razonable de los cupos carcelarios escasos, y,

2. Desde un punto de vista de justicia, hacer un uso más proporcional de las medidas de aseguramiento intramural, intentando que estas se concentren en aquellos individuos con altos niveles objetivos de riesgo de reincidencia criminal (Fiscalía General de la Nación, s.f.).

Fernández (2021) resalta, también, que sistemas como PRISMA deben contar con la descripción de cómo llegó a los resultados que ofrece, puesto que, aunque esta IA no pretende ser la última palabra en el proceso, sí es una ayuda importante al operador para tomar una decisión según la normativa vigente; sin embargo, advierte sobre PRISMA que:

> "Las personas tienden más a interpretar que el sistema informático es más objetivo y por tanto a creer en él. Sin embargo, si se declara una detención preventiva basada únicamente en el resultado de un programa de inteligencia artificial, quien en últimas tomó la decisión no fue el juez, fue PRISMA, en su calidad de acusador. Si bien formalmente el output de PRISMA no es una decisión judicial, ¿materialmente lo será?" (p. 206).

Por tanto, está claro que este sistema experto, operativo en la actualidad, funciona más como un apoyo para la decisión final que como un decisor, pues incluso cuando un operador humano acepta la sugerencia del sistema, sigue estando bajo su responsabilidad la decisión. Se está, entonces, ante un sistema que busca apoyar e, incluso, incentivar la decisión del operador judicial humano, por tanto, es un *Judicial Decision Support System*. Así, debe considerarse que la puesta en marcha de estos sistemas de apoyo no representa, como tal, un riesgo al trabajo de jueces y fiscales humanos, puesto que aquellos suplen tareas que, por el tiempo y la capacidad de pro-

(12) Para más información: https://www.fiscalia.gov.co/colombia/wp-content/uploads/Perfil-de-riesgo-de-reincidencia-para-solicitudes-de-medida-de-aseguramiento.pdf.

cesamiento de la máquina, implican mejoras considerables en el sistema judicial, mientras que al operador humano no le sería posible ofrecer el mismo *output* en el mismo tiempo o con la misma precisión [(13)].

En México se ha puesto en marcha el sistema de *machine learning* denominado Expertius I con el objetivo de dar soporte a la toma de las decisiones en el dominio del juicio especial de alimentos (Cáceres, 2023, p. 69). Este sistema cuenta con un respaldo tanto teórico como epistemológico, según la explicación de Martínez (2012):

> "El sistema no consiste en un arsenal de argumentos que deban seguir los jueces, sino que proporciona una propuesta de solución considerando criterios estandarizados en la comunidad, con lo cual posibilita que el juez rechace los argumentos siempre que justifique la razón por la cual los rechaza y que esta justificación revierta al interior del sistema. Los creadores de este sistema, así como los programadores han evitado que sea la sola teoría la que oriente al juzgado para resolver" (p. 843).

La segunda versión de Expertius, mediante su mecanismo *deep learning*, simula procesos cognitivos de los jueces con especial énfasis en las razones morales implícitas en los diferentes casos. Cáceres (2023) afirma que esta versión "no requiere del modelado del razonamiento de los expertos ya que se basa en sofisticados modelos matemáticos capaces de procesar enormes cantidades de información contenidas en bases de datos y recurren a técnicas de minería de datos" (p. 81).

Otras aplicaciones de IA-J que se vienen usando en el mundo para facilitar los trabajos de recolección y análisis de datos, están: el sistema VINCENT, que relaciona "los precedentes más relevantes conectados con una sentencia objeto de consulta" (Cáceres, 2023, p. 86); PROMETEA que se usa en Argentina como herramienta para preparar automáticamente dictámenes judiciales (Contreras et al., 2021) y el robot VICTOR que identifica la repercusión general en un recurso extraordinario ante el Supremo Tribunal Federal brasileño, que tiene funciones de tribunal constitucional, con el fin de aumentar la eficiencia en la tramitación y la resolución final de los procesos que llegan a dicho tribunal, lo que conlleva a una mejora en su *accountability* (rendición de cuentas ante la ciudadanía, especialmente en lo que respecta a la transparencia de sus decisiones y la regulación pública

---

(13) Por demás, en países como Canadá y España operan los sistemas de software Spousal Abuse Risk Assessment (SARA) y VIOGÉN, respectivamente. Dichos sistemas tienen una función similar a la de PRISMA: predecir la reincidencia delictiva; sin embargo, la diferencia de aquellos sistemas *machine learning* con el aplicado en Colombia es que operan solo frente a la violencia basada en género. Roa et al. (2023) exponen que "SARA utiliza veinte (20) factores de riesgo relacionados con la violencia contra la pareja con el fin de predecir la probabilidad de que el agresor pueda reincidir en la conducta violenta después de realizada dicha valoración" (p. 367). En España, VIOGÉN recoge datos con el objetivo de prevenir y facilitar la identificación de posibles futuros agresores (https://www.interior.gob.es/opencms/ca/servicios-al-ciudadano/violencia-contra-la-mujer/sistema-viogen/).

de dicho tribunal) [14]. Por su parte, Lex Machina, de uso, principalmente, en Estados Unidos, es una herramienta de analítica jurídica (*legal analytics*) que aplica IA para convertir datos de casos y sentencias en métricas cuantitativas, para analizar cómo se comportan jueces, abogados, tribunales o tipos de litigios.

Sin embargo, todo cambiará con los llamados jueces robot; por ejemplo, China, en el 2004, había desarrollado versiones para resolver cuestiones penales y atender al público. Granero (2021) cuenta que, a pesar de todas sus potencialidades, estos sistemas se usan solo para analizar crímenes y estandarizar las condenas, al generar automáticamente borradores de sentencias que los jueces pueden aceptar o rechazar según su criterio. Granero considera que la operación del sistema ha sido tan exitosa que cada vez resulta más raro que se rechace una sentencia sugerida por esta IA-J, asunto que ha disminuido significativamente las diferencias cuantitativas entre las condenas en casos similares a lo largo del país; ha mermado los tiempos procesales, así como la mora y la congestión judicial [15]; y ha mejorado los sistema de supervisión tanto de los jueces humanos como de la propia IA-J, ya que cuando un juez rechaza continuamente sentencias proyectadas por esta, los superiores revisan la situación. Granero agrega: "por lo que se refiere a los juzgados chinos, actualmente es bastante frecuente ser atendido por un asistente robot que resuelve preguntas y lo hace en un lenguaje que el ciudadano común pueda entender" (p. 50).

Otro caso actual son los jueces robot de Estonia, que se diferencian de otros sistemas IA-J por contar con la autoridad para tomar decisiones por su propia cuenta. En palabras de Martino (2021): "Estonia no es el primer país que unifica la IA y la ley, aunque podría ser el primero en darle la autoridad para toma de decisiones" (p. 29). Sin embargo, Ana (2021) agrega que, si bien estos jueces robot puedan emitir una decisión, ella podrá ser apelada ante un juez humano. Por tanto, "quienes están a favor argumentan que no se trata de un reemplazo, sino de un auxilio desde el cual la tecnología, mediante su compilado de métodos, medios y capacidad resolutiva a gran escala, ayudaría a la toma de decisiones" (p. 5).

Lo que se busca en el país báltico con la implementación de estos jueces robot es acelerar los procesos, mitigando así la mora y la congestión judicial que pueden suceder por muchos motivos, tanto internos como externos al

---

(14) Este robot es analizado críticamente por Kirtschig y Olsen (2023), quienes cuestionan su *Accountability,* que implica la regulación, el acompañamiento y el control que requiere dicho sistema de IA-J. Esto recuerda el adagio: ¿y quién ronda al alcalde?

(15) Sobre la diferencia entre mora ("el atraso endémico de la justicia, entendido como la demora en la definición de la controversia jurídica sometida a consideración del juez, respecto del tiempo normativo establecido para la decisión del mismo") y congestión ("entendida como el volumen de la demanda superior a la capacidad razonable de respuesta de la Rama Judicial") judiciales, así como sus posibles causas y efectos, véase a Londoño (2008, p. 398).

despacho judicial, así como mejorar los índices de acceso a la administración de justicia, pero una justicia independiente de los poderes políticos y económicos, para aumentar así los niveles de aceptación de la rama judicial por parte de los ciudadanos, lo que, a su vez, desencadena muchos otros *procesos virtuosos*, como la legitimidad del Estado, el *accountability* de la rama y los jueces, la eficacia del derecho, el fortalecimiento de la economía al ofrecer reglas claras a los inversores, etcétera.

Ramírez (2022) dice, con respecto a su aplicación, que:

> "[S]e orienta a juicios menores con el propósito de acelerar casos atrasados de los jueces y secretarios judiciales. Su aplicación funcionará mediante dos partes que cargarán sus documentos e información relevante del caso en una plataforma, donde la IA tomará una decisión que podrá ser apelada por juez humano" (p. 150).

Esto muestra que los sistemas inteligentes no responden a todas las esferas judiciales, sino, más bien, a un grupo selecto de ejercicios jurídicos. En ese sentido, las funciones decisorias finales siguen siendo, hasta ahora, una tarea humana; sin embargo, cada vez más, los sistemas IA vienen acotando terreno. En palabras de Ana (2021):

> "La organización y recuperación de información, la celeridad en el manejo de los datos, el control de los trámites, las facilidades para la producción y corrección de los documentos parecen terreno indiscutido y vistos con gratitud por parte de los actores jurídicos. Pero la mera idea de que esos instrumentos informáticos intervengan en decisiones judiciales parece ser inadmisible" (p. 5).

Aunque a la fecha, la IA-J opera más como *Judicial Decision Support System*, dados los avances vertiginosos de la IA es esperable que no se quede en ese rol y llegue a tomar decisiones, lo que aumenta el temor y la desconfianza de los que se han habituado a las prácticas, buenas o malas, de la justicia operada por humanos.

Por su parte, estos sistemas de IA-J no pasan desapercibidos ante la iusfilosofía. Por ejemplo, si se relaciona la arquitectura de estos sistemas inteligentes con debates en la filosofía del derecho, aquella podría ser objeto de la misma objeción que se hace a los realistas estadounidenses cuando se les expone que los jueces deben decidir con base en los méritos del caso, individualmente considerado, y no tanto conforme a la probabilidad de como decidiría el resto de la población judicial. ¿Cómo resolvería el sistema esta dicotomía entre (a) el juicio de predictibilidad a partir de un juez-estadístico (que busca estandarizar las decisiones) y (b) las circunstancias del caso concreto a partir de un juez-concreto (que busca hacer excepciones justificadas)? En principio se trataría de dos modelos predictivos diferentes: uno basado en las normas y los datos sobre las decisiones tomadas por una determinada población judicial (sobre la sentencia y las circunstancias aso-

ciadas a ella, las condiciones espacio-temporales, etcétera) y el otro basado en los méritos concretos del caso concreto, solo que de antemano se tendría que establecer qué se tomará por un mérito del caso, cuáles se consideran relevantes para la decisión y cómo deben afectan la decisión final del juez.

Sin embargo, cabe la pregunta: ¿es el sistema quien resuelve la dicotomía? En las primeras aplicaciones de IA, los de representación del conocimiento simbólica y basada en reglas, desde la etapa de diseño del software se resuelve si por uno de los enfoques o por los dos, y cuáles son los criterios para resolver cada vez con uno y otro modelo; en las versiones más recientes, basadas en *machine* y *deep learning*, bien podría ser la propia máquina quien resuelva la dicotomía. Pero sea uno u otro, es claro que la IA-D en general y la IA-J en especial darán lugar a una reforma significativa no solo del sistema normativo y judicial, sino también de sus sistemas de lectura y análisis, como lo es la iusfilosofía, asunto que no es objeto del presente escrito.

Por otra parte, como se sabe, la implementación de la IA-J ha generado (y seguirá generando) miedos, particularmente dos: pérdida del control de la rama judicial y de sus decisiones por parte de sistemas (políticos y económicos) hegemónicos; y transformación del oficio de jurista, pues tendrá que acercarse cada vez más a las nuevas tecnologías para realizar competentemente su labor, así como al hecho de que con la IA cada vez habrá menos necesidad de abogados (que ha sido una carrera dominante, cuantitativamente, entre las elecciones profesionales en Occidente [16]). ¿Qué político, si tiene alguna injerencia en la justicia, no temerá este cambio de paradigma? ¿Cuántos casos no se han resuelto desde criterios netamente judiciales para ser resueltos desde los intereses políticos y económicos de turno? ¿No tendrían miedos estas personas privilegiadas por el sistema hegemónico de ser juzgados por una IA-J sobre la cual no tendrían control? Además, ¿qué jurista no tendrá miedo ante este cambio tan significativo en su entorno laboral si resulta que la máquina hace mejor estas tareas y con mejor aceptabilidad por parte de la ciudadanía? Si dichas IA-J automatizan el trabajo judicial humano, entonces, el empleo dentro de la rama judicial puede disminuir significativamente y afectaría especialmente a los que no se adapten, o no sepan cómo hacerlo, a los cambios que se avecinan [17].

Frente a estos miedos, debe tenerse en cuenta que la implicación de los sistemas inteligentes en el derecho (y por tanto en lo judicial) no es un asunto nuevo, pues desde la génesis de los avances tecnológicos y digitales se ha preguntado por la utilidad y la regulación de estos sistemas. Frente a la

(16) Para Colombia, consultar García y Ceballos (2019).

(17) Puede que esta sea la visión más fatalista del asunto y, por tanto, la que más genera problemas. Sin embargo, puede que, más bien, se trate de una transformación del oficio de algunos actores de la rama. Por ejemplo, el diseño de estos sistemas parte de la consideración de la experticia humana, luego la máquina aprende a resolver autónomamente.

utilidad ya se han dado varios elementos de juicio en este escrito, y frente a la regulación, es una asignatura pendiente a nivel nacional [18] e internacional [19], regulación que se ha dificultado por diferentes motivos, como los intereses (económicos, políticos, de razón de Estado, etcétera) encontrados, las dificultades (técnicas, políticas e incluso de técnica legislativa) de regular de forma global (pues hoy día una regulación meramente nacional no sería suficiente) algo que cambia tan radicalmente en poco tiempo, y la preocupación del ser humano con lo desconocido (que lleva a muchos a satanizar la IA-J). Incluso, la utilidad de la IA exige una regulación inmediata y fuerte, como forma de contrarrestar el miedo de muchos a ser reemplazados o, peor aún, superados. Pero se reitera que dicha regulación, por más necesaria que parezca, no se dará fácilmente en el plano que debería darse: el mundial, de un lado, y que se requeriría nuevas técnicas regulativas, pues la velocidad de los cambios de la IA es superior a la de las formas tradicionales de reforma de las normas regulatorias, de otro.

En síntesis, hoy en día se está aplicando la IA-J como un sistema experto asesor y auxiliar del juez-humano, por lo que los cambios no parecen tan radicales, lo que favorecería su aceptación, pero si se permite a las máquinas fungir como jueces y decidir directamente casos, los cambios en el sistema jurídico (incluyendo su subsistema de lectura y análisis, esto es, la academia) y en el judicial serían comparables con la invención de la rueda o la domesticación del fuego. Buena parte del miedo, e incluso del rechazo a la IA-J surge de quienes se han habituado a cierta forma de operar la justicia o de los que se han beneficiado de la forma en que ha obrado. Pero, cabe preguntarse si el miedo de estas personas podrá frenar la IA-J. La respuesta es un no, máxime si la oposición a la IA se dará desde el desconocimiento que alimenta argumentos infundados sobre lo que pueden o no hacer estos sistemas inteligentes, que no son ni serán perfectos, pero hasta el momento permiten vislumbrar buenos resultados sistémicos probados ya con los sistemas actualmente en operación (utilidad), aunque signifique poner en

(18) Al momento de la escritura de este texto, se cuenta en Colombia con un marco ético realizado por el Estado (Guío et al., 2021) y el proyecto de ley 059 del Senado de la República para establecer una política pública para el desarrollo, uso e implementación de la IA, presentado por el senador Juan Diego Echavarría el 1º de agosto de 2023. El proyecto ha sido retirado o archivado en tres ocasiones previas por la controversia que genera entre defensores y detractores de la IA en Colombia; sin embargo, ante la globalización actual, una regulación nacional debe estar articulada con las normas internacionales del tema, que aún están por verse.

(19) Entre los que hay que mencionar, los *Principios de la OCDE sobre la Inteligencia Artificial* del 21 de mayo de 2019 (Disponible en: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449), al que ha sido adherido Colombia (ver la noticia: https://www.mintic.gov.co/portal/inicio/Sala-dePrensa/Noticias/100683:Colombia-se-adhiere-a-acuerdo-sobre-Inteligencia-Artificial-ante-los-paises-de-la-OCDEJ). Además, algunos proyectos de ley para la regulación sobre la IA avanzan en otros espacios como la Unión Europea (a la fecha, falta la aprobación del Parlamento Europeo), Argentina (proyecto de ley que modificaría la ley nacional 25.467), Costa Rica y Panamá (anteproyecto de ley No. 014). Para más información, véase: https://forogpp.com/inteligencia-artificial/regulacion-sobre-ia/ y https://www.europarl.europa.eu/news/es/headlines/society/20230601STO93804/ley-de-ia-de-la-ue-primera-normativa-sobre-inteligencia-artificial

cuestión algunos privilegios adquiridos, lo que supone una necesaria regulación más global que nacional. El mejor escenario sería que los que sienten que serán desplazados por el cambio se informen objetiva y críticamente de lo que ha sido y puede llegar a ser la IA-J para que puedan adaptarse al cambio, de un lado, y para que puedan sugerir una buena regulación cuando sea el momento, de otro.

4. A modo de conclusión

Hace poco se creía que no sería posible sistemas de IA que pudiesen fungir como jueces dirigiendo el proceso y emitiendo sentencias, es decir, resolviendo conflictos a partir de normas jurídicas y acorde con las pruebas, pero existen ya sistemas de IA-J, en su mayoría en fase experimental, o de los que no se sabe mayor cosa por ser secreto industrial o por temas de seguridad nacional. Los que ya están operativos en el mundo, y cuyo funcionamiento es de acceso abierto, solo actúan, hasta el momento, como auxiliares del juez, para ahorrarle trabajo y mejorar sus fuentes de información. Claramente, la IA-J está maniatada actualmente, por lo que aún no resuelven casos directamente (salvo el caso de los jueces robot de China y Estonia), especialmente por los miedos que ha implicado su implementación definitiva. No obstante, no pasará mucho tiempo para que la IA-J asuma algo para lo que está habilitada: dirigir procesos y sentenciar casos.

Los sistemas de IA-J que ya están operando, especialmente para resolver problemas judiciales al interior de (o con) multinacionales o para casos de mínima cuantía, han dado resultados asombrosos con altos niveles de aceptabilidad por parte de los usuarios. A esto se suma que hoy día no hay mayores problemas tecnológicos para que la IA-J pueda dirigir un proceso y sentenciar casos directamente.

Por ejemplo, estos sistemas están en condiciones de argumentar adecuadamente e, incluso, a partir de sistemas expertos de valoración (con capacidad de ponderar con variables tanto jurídicas como factuales), todo según su programación inicial –pero sin quedarse en ella–, en el caso de sistemas de representación simbólica y basados en reglas, o según el propio aprendizaje, en sistemas fundados en *machine learning*. Entre estos sistemas de IA con capacidad argumentativa, el más robusto hasta el momento es Cato, desarrollado en la Universidad de Pittsburgh por Kevin Ashley para el mundo anglosajón. Con respecto a la ponderación probatoria, hay algunos prototipos basados en el uso de Bayes ingenuo [20] y Expertius II, por mencionar solo dos casos.

---

(20) "Bayesiano ingenuo es un algoritmo muy utilizado para resolver problemas de clasificación. El modelo se denomina *naïve* porque trata todas las variables de predicción propuestas como independientes unas de otras. El bayesiano ingenuo es un algoritmo rápido y escalable que calcula las

Además, los sistemas IA desarrollados a partir de *machine learning* y *deep learning*, que en muchos casos resultan más recomendables que los determinísticos basados en reglas, ofrecen resultados fiables y con un alto estándar de seguridad, pues una vez entran a operar mejoran su desempeño en cada transacción, ya que aprenden de su experiencia; es decir, en cada transacción almacenan un nuevo conjunto de datos, cuyas relaciones con el resultado ofrecido le permite a la IA reconfigurar los criterios de decisión para las siguientes transacciones, todo ello sin intervención directa de humanos, aunque es muy recomendable una constante supervisión que, a su vez, para garantizar la mayor objetividad posible, también debe ser auditada. En el panorama de la IA-J, aunque en la actualidad estos sistemas no cuentan con mayor autonomía, se espera que en el futuro sí logren, incluso, tomar decisiones judiciales no supervisadas, al menos en algunos campos donde exista un gran riesgo para los jueces humanos o falta de objetividad e imparcialidad para el juzgado [(21)].

Otro reto de la IA-J es que debe partir de una teoría jurídica computacional, incluyendo los sistemas *machine learning* basados en minería de datos, lo que implica que se deba cuidar la calidad de los datos que, en grandes volúmenes se le ofrecen al algoritmo para lograr su reconocimiento, su procesamiento y su consecuente *output*, pues si el sistema almacena datos errados, inválidos o ilegales como fuente de información, lo que se va a obtener como *output* estará viciado. Esto supone el reto de qué pasaría si se dejara a la propia IA buscar sus fuentes de información, ¿se mejoraría la calidad de los *inputs* y con ello el resultado o, por el contrario, se podría empeorar?

Otro desafío que enfrenta la IA-J es la creencia de que para los sistemas *machine learning* no se puede predecir con exactitud lo que harán cuando estén razonando, como si con aprendizaje se hiciera referencia a que formaran una suerte de autodeterminación, cuando realmente el aprendizaje de máquina se trata de la capacidad de reconocimiento de nuevos patrones o coincidencias entre conjuntos de datos que pueden llevar al algoritmo a tener un desempeño cada vez más cercano al ideal, es decir, a cumplir mejor la función para la que fue diseñada (que no cambia por sí sola). Prejuicios de este tipo en contra de la IA-J van dejando de tener soporte alguno con el conocimiento de las formas en las que operan estos sistemas.

---

probabilidades condicionales para las combinaciones de atributos y el atributo de objetivo. A partir de los datos de entrenamiento se establece una probabilidad independiente. Esta probabilidad proporciona la verosimilitud de cada clase objetivo, una vez dada la instancia de cada categoría de valor a partir de cada variable de entrada". Tomado de: https://www.ibm.com/docs/es/spss-modeler/saas?topic=mining-oracle-naive-bayes (consultado 11-09-2023).

(21) Por temas de espacio, no se puede ahondar en las maneras como la IA-J puede tomar decisiones judiciales por su cuenta, por lo que se remite a Luna et al. (2022). Lo que sí se puede decir es que el derecho, como todas las áreas del saber, debe afrontar los desafíos y las actualizaciones que trae consigo la tecnología digital.

En últimas, se está frente a un camino sin retorno; el actual modo de pensar y de forma de vida es parte, para bien o mal, de los avances tecnológicos. Rechazar lo desconocido no es solución, como tampoco aceptarlo sin más; es necesario desarrollar un pensamiento tecnológico crítico, que se incorpore y se adapte al avance de la IA-J. Este pensamiento implica reconocer que la vida humana, en todas sus facetas, transcurre mediada por la tecnología y que resultaría problemático prescindir de ella una vez ha sido internada en la cultura. Implica, también, ser conscientes de lo expansivo que ha sido el estallido digital en todas las áreas, y que esto puede generar temor a supuestas catástrofes apocalípticas, pero este temor no impedirá que se acerquen cada vez más los sistemas inteligentes y las actividades humanas, las máquinas y las personas, la inteligencia natural y la artificial. Es necesario mitigar estos miedos con información y con regulación (no solo jurídica, pues el derecho no es el único sistema normativo) más global que nacional, pero una regulación que establezca lineamientos para que prevalezca la preservación del ser humano y de su entorno, la protección de sus derechos, la búsqueda del bien común, entre otros principios que, de alguna manera, delimitan el alcance de la IA en lo humano, lo jurídico y lo judicial.

Sin embargo, el problema no es el miedo en sí mismo, sino que este esté asociado con la ignorancia alrededor de la IA, lo que podría generar actos como los que la historia ha mostrado que suceden cuando algunas personas, llenas de temor y prejuicios, actúan en contra de lo que les produce tales cosas. En este sentido, quien juzgue la IA-J sin haberse informado previamente de qué es, de cómo funciona, de sus aplicaciones útiles, etcétera, puede terminar por satanizarla y hacerle el juego a ciertos actores interesados en que la IA-J no opere, en especial en situaciones donde sería muy útil: juzgar casos de corrupción y de delincuencia organizada, ya que una máquina-juez no puede ser coaccionada ni corrompida.

¿Estos miedos y esta ignorancia podrán frenar la llegada plena de la IA-J? Como ya se dijo, aquí se considera que no, a menos que la justicia humana se legitime tanto y en tan poco tiempo ante los ciudadanos, ofreciendo aquello que la ciudadanía exige de sus jueces (que sean motor de convivencia real), que la misma sociedad decline lo que hasta ahora parece ser inevitable.

Referencias

Alarcón, A., Villalba, J. C., y Franco, J. F. (2019). La inteligencia artificial y su impacto en la enseñanza y el ejercicio del derecho. *Prolegómenos*, *22*(44), 7–10. https://doi.org/10.18359/prole.4353.

Ana Dobratinich, G. (2021). Derecho y tecnología. Diálogos in-calculables. En Ana Dobratinich, G. (dir.). *Derecho y nuevas tecnologías* (pp. 1-13). La Ley.

Andreu, N. (2021). El problema de la consciencia en la inteligencia artificial. Una aproximación desde el derecho. En Ana Dobratinich, G. (dir.). *Derecho y nuevas tecnologías* (pp. 225-239). La Ley.

Botero, A. (2006). Ensayo sobre la crisis de la razón jurídica: Formalismo versus principios y/o valores. *Opinión jurídica*, *5*(9), 33-61.

Botero, A. (2016). Sobre el uso de la bibliografía en la investigación jurídica. *Revista Pensamiento Jurídico*, (43), 475-504. https://revistas.unal.edu.co/index.php/peju/article/view/60774

Botero, A. (2021). Derecho, sistema y realidad: ¿el positivismo se hace agua en nuestras manos? *Problema. Anuario de Filosofía y Teoría del Derecho*, (15), 263-306. https://doi.org/10.22201/iij.24487937e.2021.15.16124

Botero, A., Gamboa, S., y Valdivieso, K. (2024). Reflexiones sobre la inteligencia artificial aplicada a la administración de justicia. *Revista Jurídica. Mario Alario D'Filippo*, *16*(33), 160-183. https://doi.org/10.32997/2256-2796-vol.16-num.33-2024-4887

Cáceres Nieto, E. (2023). La inteligencia artificial aplicada al derecho como una nueva rama de la teoría jurídica. *Anales de la Cátedra Francisco Suárez*, (57), 63-89. https://doi.org/10.30827/acfs.v57i.26281

Campos Madrigal, J. (2021). Reflexiones filosóficas sobre la inteligencia artificial, moralidad y consecuencias jurídicas de la IA. En Ana Dobratinich, G. (dir.). *Derecho y nuevas tecnologías*, (pp. 193-208). La Ley.

Cerdio Herrán, J. (2021). Con culpa o sin ella: responsabilidad para sistemas inteligentes. En Ana Dobratinich, G. (dir.). *Derecho y nuevas tecnologías*, (pp. 69-102). La Ley.

Churnin, S. (2011). *Inteligencia artificial retos éticos y jurídicos, y la influencia de los derechos humanos*. Universidad Complutense de Madrid.

Contreras, P., Azuaje, M., Bedecarratz, F., Bozzo, S., Díaz, J. P.y Finol, D. (2021). Enseñanza y aprendizaje de la inteligencia artificial y derecho en Chile: sobre el *minor* en Inteligencia Artificial y Derecho de la Universidad Autónoma de Chile. *Revista pedagogía universitaria y didáctica del derecho*, *8*(2), 281-302. https://doi.org/10.5354/0719-5885.2021.64456

Corte Constitucional (2020). PRETORIA, un ejemplo de incorporación de tecnologías de punta en el sector justicia. Recuperado de: https://www.corteconstitucional.gov.co/noticia.php?PRETORIA,-un-ejemplo-de-incorporaci%C3%B3n-de-tecnolog%C3%ADas-de-punta-en-el-sector-justicia-8970.

Domingo, C. (2021). Utilización del sistema de reconocimiento facial para preservar la seguridad ciudadana. *El Criminalista Digital*, *II época* (9), 20-37. https://revistaseug.ugr.es/index.php/cridi/article/view/20899

Dreyfus, H. L., & Dreyfus, S. E. (1986). *From Socrates to Expert Systems: The Limits of Calculative Rationality.* In C. Mitcham & A. Huning (Eds.), *Philosophy and Technology II: Information Technology and Computers in Theory and Practice* (pp. 111-130). Dordrecht, Netherlands: Springer. https://doi.org/10.1007/978-94-009-4512-8_9

Faulkner, A. (2025). From Case Law to Code: Evaluating AI's Role in the Justice System. Montreal AI Ethics Institute. [https://montrealethics.ai/from-case-law-to-code-evaluating-ais-role-in-the-justice-system/#:~:text=Legal%20Research]

Fernández, M. (2021). El proceso y el precrimen: una lectura literaria del uso de la inteligencia artificial en la detención preventiva. *UNA Revista de Derecho, 6*(1), 190-224. https://doi.org/10.57784/1992/54860

Fiscalía General de la Nación (s.f). Herramienta PRiSMA: Perfil de Riesgo de Reincidencia para la Solicitud de Medidas de Aseguramiento [en línea]. Recuperado de: https://www.fiscalia.gov.co/colombia/wp-content/uploads/Perfil-de-riesgo-de-reincidencia-para-solicitudes-de-medida-de-aseguramiento.pdf.

Gallagher, S. y Zahavi, D. (2008). *The Phenomenological Mind. An Introduction to philosophy of mind and cognitive science.* Routledge.

Gamboa, S. (2020). *Mente y tecnicidad*. Aula de Humanidades.

García Villegas, M. y Ceballos Bedoya, M. A. (2019). *La profesión jurídica en Colombia: Falta de reglas y exceso de mercado.* Dejusticia.

Gottfredson, L. (1997a). Mainstream Science on Intelligence: An Editorial With 52 Signatories, History, and Bibliography. *Intelligence*, *24*(1), 13-23.

Gottfredson, L. (1997b). Why g Matters: The complexity of Everyday Life. *Intelligence*, *24*(1), 79-132.

Granero, H. (2021). Sistemas inteligentes en la justicia. Retos, oportunidades y desafío. En Ana Dobratinich, G. (dir.). *Derecho y nuevas tecnologías* (pp. 43-68). La Ley.

Guío, A., Tamayo, E. y Gómez, P. (2021). *Marco ético para la inteligencia artificial en Colombia*. Consejería Presidencial de Asuntos Económicos y Transformación Digital.

Kirtschig, G. y Olsen, A. (2023). O robô supremo: inteligência artificial, empoderamento e controle. *Seqüência*, *44*(94), 1-24. https://doi.org/10.5007/2177-7055.2023.e79766

Loevinger, L. (1949). Jurimetrics. The Next Step Forward. *Minnesota Law Review. Journal of the State Bar Association, 33*(5), 455-493.

Londoño, M. (2008). La congestión y la mora judicial: el juez, ¿su único responsable? *Revista Facultad de Derecho y Ciencias Políticas*, *38*(109), 385-419.

Luna, F., Perona, R. y Carillo, Y. (2022). Impacto y límites de la inteligencia artificial en la práctica jurídica. *Via Inveniendi Et Iudicandi*, *17*(2), 234-244. https://doi.org/10.15332/19090528.8773.

Manrique, E. (2020). Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo. *Revista Ibérica de Sistemas e Tecnologias de Informação*, (E28), 586-599.

Maranhão, J., Abrusio, J., Almada, M. y López, N. (2021). La inteligencia artificial y el derecho: dos perspectivas. En Ana Dobratinich, G. (dir.). *Derecho y nuevas tecnologías* (pp. 143-166). La Ley.

Martínez, G. (2012). La inteligencia artificial y su aplicación al campo del derecho. *Alegatos*, (82), 827-843.

Martino, A. (2021). ¿Quién teme a los jueces robot? En Ana Dobratinich, G. (dir.). *Derecho y nuevas tecnologías* (pp. 15-41). La Ley.

Narváez, J. (2023). Blade Runner: La inteligencia artificial consciente y el derecho futuro (qué nos dice la ciencia ficción), *Abogacía*, (33), 68-70.

Otero, M. (2023) ¿Puede la inteligencia artificial sustituir a la mente humana? Implicaciones de la IA en los derechos fundamentales y en la ética. *Anales de la Cátedra Francisco Suárez*, (57), 39-61. https://doi.org/10.30827/acfs.v57i.24710

Pérez, C. (2004). *Revoluciones tecnológicas y capital financiero. La dinámica de las grandes burbujas financieras y las épocas de bonanza*. Siglo XXI editores.

Ramírez, P. (2022). Inteligencia artificial y sus aportes al derecho sancionador. En Consejo nacional de Disciplina Judicial (Coord.). *Derecho jurisdiccional disciplinario, desde un enfoque ético, deontológico y preventivo* (pp.141-163). Consejo Superior de la Judicatura.

Roa, M., Sanabria, J. & Dinas, K. (2023). Herramientas de predicción de violencia basada en género y feminicidio mediante la Inteligencia Artificial. *Revista Jurídica Mario Alario D'Filippo*, *15*(30), 360-390. https://doi.org/10.32997/2256-2796-vol.15-num.30-2023-4254

Russell, S., y Norvig., P. (2004). *Inteligencia artificial. Un enfoque moderno* (2da ed.). Pearson Prentice Hall.

Sanabria, J., Roa, M. del P., y Lee Pérez, O. (2022). Tecnología de reconocimiento facial y sus riesgos en los derechos humanos. *Revista Criminalidad*, *64*(3), 61-78. https://doi.org/10.47741/17943108.366

Sánchez, A. (2019) Neuro-evolucionismo y Deep machine learning: nuevos desafíos para el derecho. *Journal of Ethics and Legal Technologies*, *1*(1), 115-136.

Simon, H. (2006). *Las ciencias de lo artificial*. Comares.

Villegas, M. (2023). Entrevista a Paul Handal. Los efectos de la inteligencia artificial en el ejercicio jurídico. *Iuris Dictio*, (31), 89-95. https://doi.org/10.18272/iu.i31.2960

Sección III

Normas, leyes, regulaciones

Contratación de inteligencia artificial por parte del Estado: análisis legal y regulatorio en Argentina [*]

Por Manuela Adrogué [**] y Milagros Ibarzábal [***]

Resumen: El artículo analiza la contratación de sistemas de inteligencia artificial por parte del Estado argentino, destacando tanto sus oportunidades como sus desafíos jurídicos, técnicos y éticos. Se señala que la IA permite mejorar la eficiencia, la rapidez y la calidad de la gestión pública, pero plantea riesgos específicos, como sesgos algorítmicos, vulneración de derechos fundamentales y problemas de protección de datos. Se describen los marcos normativos nacionales e internacionales relevantes, incluyendo la Guía de la AAIP sobre Transparencia y Protección de Datos en IA Responsable, la Recomendación de la UNESCO sobre la Ética de la IA y el Plan Nacional de Inteligencia Artificial. Además, se examinan experiencias comparadas en la Unión Europea, Canadá, Reino Unido y Estados Unidos, destacando instrumentos como el AI Act, el *Pre-Commercial Procurement* (PCP) y las *Innovation Partnerships*. Finalmente, se presentan buenas prácticas para una contratación pública responsable, incluyendo la evaluación de riesgos, la trazabilidad, la auditoría, la formación de equipos interdisciplinarios y la implementación de *sandboxes* contractuales, con el objetivo de equilibrar innovación tecnológica, seguridad jurídica y respeto a los derechos fundamentales.

(*) Agradecemos a Segundo Barda por la asistencia prestada durante la elaboración de este trabajo.

(**) Manuela Adrogué es socia en el departamento de Tecnología de la Información, Privacidad & Inteligencia Artificial en Marval O'Farrell Mairal, donde se especializa en privacidad, protección de *software* y tecnología. Su trabajo incluye la realización transacciones orientadas a la preservación de activos intangibles, con un enfoque particular en derechos de propiedad intelectual y la protección de datos personales.

(***) Milagros Ibarzábal es abogada de la Universidad Austral, graduada con honores. Máster en Derecho Administrativo de la Universidad Austral y Máster en Leyes de New York University. Es socia en el departamento de Derecho Público & Administrativo de Marval O'Farrell Mairal. Se especializa en contrataciones y transacciones con entidades públicas, disputas contra el Estado tanto en sede administrativa como judicial a nivel nacional y en varias jurisdicciones provinciales, cuestiones regulatorias y *Compliance*, ética pública y políticas de integridad.

Palabras clave: Contratación pública; inteligencia artificial; sector público; argentina; derechos fundamentales; transparencia; protección de datos; ética.

Abstract: This article examines the procurement of artificial intelligence (AI) systems by the Argentine State, highlighting both opportunities and legal, technical, and ethical challenges. AI can enhance efficiency, speed, and quality in public administration, but it also raises specific risks, such as algorithmic bias, violations of fundamental rights, and data protection concerns. The study outlines relevant national and international regulatory frameworks, including the AAIP Guide on Transparency and Data Protection in Responsible AI, the UNESCO Recommendation on the Ethics of AI, and the National AI Plan. Comparative experiences in the European Union, Canada, the United Kingdom, and the United States are also discussed, highlighting instruments like the AI Act, Pre-Commercial Procurement (PCP), and Innovation Partnerships. Finally, best practices for responsible public procurement are presented, including risk assessment, traceability, auditing, interdisciplinary teams, and the implementation of contractual sandboxes, aiming to balance technological innovation, legal certainty, and the protection of fundamental rights.

Keywords: Public procurement; Artificial Intelligence; public sector; Argentina; fundamental rights; transparency; data protection; ethics.

1. INTRODUCCIÓN

Si el proceso de digitalización convirtió al Estado en una ventanilla electrónica, la inteligencia artificial está transformando dicha ventanilla en la manera en que el Estado decide y se relaciona con la ciudadanía. ¿Qué significa que un algoritmo atienda consultas, prediga necesidades o detecte problemas urbanos en tiempo real? Lo que en la esfera privada ya se percibe como cotidiano comienza a trasladarse al ámbito estatal, donde la inteligencia artificial no solo agiliza trámites o mejora diagnósticos [(1)], sino que redefine el estándar mismo de lo que entendemos por servicio estatal eficiente.

El uso de la inteligencia artificial promete en el ámbito del sector público mejorar la eficiencia y la velocidad en la adopción de decisiones, ahorrar costos y lograr mejores resultados en la gestión pública.

Por el momento, pareciera que la atención ciudadana es el ámbito en el cual estas soluciones se presentan como más frecuentes. Por ejemplo, en el

(1) Solano, Mauro; Estévez, Elsa; y Peralta, Amalia, *Guía para el uso de IA en el sector público en Argentina*, CIPPEC, Buenos Aires, 2025. Disponible en: https://www.cippec.org/wp-content/uploads/2025/05/Guia-para-el-uso-de-IA-en-el-sector-publico-en-Argentina_2025.pdf.

uso de *chatbots* como "Boti" [2] de la Ciudad de Buenos Aires que incorpora inteligencia artificial generativa basada en grandes modelos de lenguaje [3], o "Tina" la asistente virtual del Estado nacional [4]. Existen también otras aplicaciones más disruptivas, como la de la Ciudad de Mendoza para identificar microbasurales a cielo abierto [5], o la iniciativa del Gobierno de Salta para utilizar un sistema de inteligencia artificial destinado a predecir embarazos adolescentes [6] mediante el análisis de datos recolectados por agentes territoriales [7]. A la vez, algunos organismos nacionales ya han comenzado a explorar su uso de manera más activa [8], aunque no existe al presente un marco normativo específico que regule su adquisición y utilización.

Este escenario plantea interrogantes centrales sobre la incorporación de la inteligencia artificial en la gestión pública: ¿debiera regularse su contratación? y, en caso afirmativo, ¿cómo se regula su contratación?, ¿qué particularidades presentan estos procesos frente a la adquisición de bienes o servicios tradicionales?, ¿qué riesgos jurídicos y éticos emergen?

Abordar estas cuestiones requiere, en primer lugar, precisar el alcance conceptual de la inteligencia artificial, para luego examinar las particularidades de la contratación de la inteligencia artificial por el sector público, los marcos normativos aplicables, las experiencias en el derecho comparado y las buenas prácticas que pueden orientar una implementación estatal responsable.

---

(2) Según surge de la propia web de CABA, los contenidos disponibles en Boti son múltiples y están en constante actualización: a diario se incorporan temáticas relacionadas a la Ciudad para acercar a las personas toda la información que necesitan para hacer gestiones, resolver trámites, sacar turnos, pedir asistencia o conocer las propuestas culturales y turísticas disponibles en Buenos Aires. [ver: https://buenosaires.gob.ar/innovacionytransformaciondigital/boti]

(3) Solano, Mauro; Estévez, Elsa; y Peralta, Amalia, *Guía para el uso de IA en el sector público en Argentina*, CIPPEC, Buenos Aires, 2025. Disponible en: https://www.cippec.org/wp-content/uploads/2025/05/Guia-para-el-uso-de-IA-en-el-sector-publico-en-Argentina_2025.pdf.

(4) "Se lanza Tina, la asistente virtual del Estado nacional" [ver: https://www.argentina.gob.ar/noticias/se-lanza-tina-la-asistente-virtual-del-estado-nacional?utm_source=chatgpt.com]

(5) Mediante un convenio con la Fundación Bunge y Born, se creó el proyecto "Inteligencia artificial aplicada al monitoreo de microbasurales en la Ciudad de Mendoza". Utiliza un algoritmo basado en redes neuronales entrenadas con imágenes aéreas o satelitales de alta resolución, con el propósito de detectar microbasurales de manera periódica y a bajo costo. [Ver: https://www.fundacionbyb.org/post/con-ia-detectan-microbasurales-para-prevenir-problemas-ambientales?utm_source=chatgpt.com].

(6) "El Gobierno de Salta y Microsoft aplicarán inteligencia artificial en la prevención de situaciones sociales". Secretaría de Comunicación. Salta Gobierno. [Ver: https://www.salta.gob.ar/prensa/noticias/el-gobierno-de-salta-y-microsoft-aplicaran-inteligencia-artificial-en-la-prevencion-de-situaciones-sociales-52860]

(7) El sistema nunca llegó a implementarse operativamente: la presión social y mediática frente a sus riesgos de estigmatización y opacidad metodológica llevó a que el proyecto quedara en fase experimental y no se tradujera en políticas públicas concretas.

(8) En el sitio oficial COMPR.AR. en la opción "búsqueda de procesos" al buscar bajo el "Nombre descriptivo del proceso" con las palabras "inteligencia artificial" aparecen solo dos licitaciones que ya fueron adjudicadas: 318-0003-LPR24 (contratación de una solución informática integral basada en inteligencia artificial que fue realizada mediante licitación privada por la Autoridad de Cuenca Matanza Riachuelo) y 55-0027-CDI23 (consultoría para el estudio de la aplicación de inteligencia artificial a procesos del ENARGAS que fue realizado mediante contratación directa por el Ente Regulador del Gas).

2. Qué es la inteligencia artificial

Antes de abordar los desafíos que plantea la contratación de sistemas de inteligencia artificial en el ámbito estatal, resulta necesario delimitar conceptualmente qué debe entenderse por este concepto. La inexistencia de una definición jurídica única y consolidada explica que tanto la doctrina como los organismos internacionales recurran a aproximaciones diversas. Contar con una base conceptual clara no solo permitirá distinguir la inteligencia artificial de otras formas de automatización, sino también identificar los beneficios y riesgos específicos que su utilización conlleva en la esfera pública.

La Guía para entidades públicas y privadas en materia de Transparencia y Protección de Datos Personales para una Inteligencia Artificial responsable, elaborada por la Agencia de Acceso a la Información Pública (AAIP), aporta una distinción particularmente clara. Señala que los sistemas de inteligencia artificial se inscriben dentro de los llamados Sistemas de Decisión Automatizada (SDA), aunque con rasgos diferenciados: mientras que un SDA tradicional opera sobre la base de un conjunto predefinido de reglas, rígidas y diseñadas para tareas repetitivas, la inteligencia artificial incorpora capacidades más complejas, entre ellas el reconocimiento de patrones, la comprensión del lenguaje natural y el aprendizaje adaptativo. En este sentido, la inteligencia artificial utiliza técnicas como el *machine learning* y las redes neuronales, lo que le permite aprender de los datos, mejorar su desempeño con el tiempo y adaptarse a nuevas situaciones [(9)].

Esta caracterización coincide en buena medida con la propuesta de la Comisión Europea, que en su White Paper de 2020 definió a la inteligencia artificial como un conjunto de tecnologías que, mediante software, pueden simular procesos de inteligencia humana tales como el razonamiento, la percepción, la predicción y la toma de decisiones autónomas [(10)]. A su vez, la Organización para la Cooperación y el Desarrollo Económicos (OCDE) ha enfatizado que la inteligencia artificial debe entenderse como un "sistema basado en máquinas que, para un conjunto determinado de objetivos, puede influir en ambientes reales o virtuales produciendo resultados como predicciones, recomendaciones o decisiones" [(11)].

De este modo, la inteligencia artificial se distingue de otras formas de automatización por su capacidad de aprender y evolucionar, reduciendo la

---

(9) Agencia de Acceso a la Información Pública, Guía para entidades públicas y privadas en materia de Transparencia y Protección de Datos Personales para una Inteligencia Artificial responsable, Buenos Aires, 2022.

(10) Comisión Europea, White Paper sobre la Inteligencia Artificial. Un enfoque europeo orientado a la excelencia y la confianza, COM(2020) 65 final, Bruselas, 2020.

(11) Organización para la Cooperación y el Desarrollo Económicos (OCDE), *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, París, 2019.

rigidez de los sistemas tradicionales (por ejemplo, programas con reglas fijas, algoritmos deterministas o procesos mecánicos que solo ejecutan instrucciones predefinidas) y generando, al mismo tiempo, mayores beneficios potenciales y mayores riesgos jurídicos.

3. Particularidades de la contratación de la inteligencia artificial por parte del Sector Público: riesgos y desafíos

Como recién explicamos, a diferencia del *software* tradicional, los sistemas de inteligencia artificial se caracterizan por su capacidad de aprender, adaptarse y modificar su comportamiento a partir de los datos que procesan. Mediante el análisis de grandes volúmenes de información, identifican patrones, extraen tendencias y formulan decisiones basadas en esos insumos [(12)]. Para su correcto funcionamiento, requieren datos de calidad y en gran escala, así como una infraestructura tecnológica robusta (por ejemplo, conectividad de alta velocidad y centros de datos adecuados [(13)]).

Estas particularidades generan riesgos y desafíos específicos que deben ser anticipados en la contratación pública de soluciones de inteligencia artificial, tanto en la etapa previa de planificación como en la redacción de los pliegos licitatorios.

3.1. Sesgos algorítmicos y discriminación

Uno de los riesgos es la posibilidad de que los sistemas de inteligencia artificial reproduzcan o amplifiquen sesgos existentes en los datos, lo que puede conducir a decisiones discriminatorias o arbitrarias. Ello plantea la necesidad de exigir mecanismos de transparencia y explicabilidad en la medida necesaria para verificar el cumplimiento contractual y proteger derechos, que permitan comprender, cuando sea técnicamente posible, cómo se arribó a un determinado resultado [(14)].

3.2. Impacto en derechos fundamentales

El uso de la inteligencia artificial en la Administración pública puede afectar de manera directa el ejercicio de derechos fundamentales, como el

---

(12) Solano, Mauro; Estévez, Elsa; y Peralta, Amalia, *Guía para el uso de IA en el sector público en Argentina*, CIPPEC, Buenos Aires, 2025, p. 4. Disponible en: https://www.cippec.org/wp-content/uploads/2025/05/Guia-para-el-uso-de-IA-en-el-sector-publico-en-Argentina_2025.pdf.

(13) Solano, Mauro; Estévez, Elsa; y Peralta, Amalia, *Guía para el uso de IA en el sector público en Argentina*, CIPPEC, Buenos Aires, 2025, p. 4. Disponible en: https://www.cippec.org/wp-content/uploads/2025/05/Guia-para-el-uso-de-IA-en-el-sector-publico-en-Argentina_2025.pdf.

(14) Solano, Mauro; Estévez, Elsa; y Peralta, Amalia, *Guía para el uso de IA en el sector público en Argentina*, CIPPEC, Buenos Aires, 2025, p. 4. Disponible en: https://www.cippec.org/wp-content/uploads/2025/05/Guia-para-el-uso-de-IA-en-el-sector-publico-en-Argentina_2025.pdf.

debido proceso, la igualdad ante la ley o el acceso a la información. Por ello, la contratación debería incorporar garantías suficientes de respeto y protección de estos derechos, reforzando estándares de legitimidad y control.

3.3. Protección de datos y seguridad de la información

Otro aspecto crítico es la protección de los datos personales y la seguridad de la información. Los pliegos deberían establecer estándares mínimos de seguridad incluyendo la posibilidad de revisiones técnicas, limitadas al alcance necesario para comprobar el cumplimiento de los estándares pactados, sin acceso irrestricto a información sensible o secretos comerciales.

3.4. Interoperabilidad y prevención de dependencia tecnológica

El fomento de estándares abiertos e interoperabilidad es deseable para proteger el interés público y facilitar la continuidad operativa. No obstante, estas exigencias deben guardarse en términos de razonabilidad técnica y económica, evitando que deriven en (i) la divulgación de secretos comerciales, *know-how* o información sensible del proveedor; o (ii) inversiones desproporcionadas que trasladen al contratista costos no previstos.

En consecuencia, la interoperabilidad debería circunscribirse a interfaces y formatos de intercambio necesarios para el uso previsto y la eventual migración ordenada, sin requerir acceso al código fuente, a los pesos de modelos (*model weights*) ni a documentación confidencial, salvo pacto específico y con salvaguardas de confidencialidad y compensación adecuadas.

En conjunto, estos riesgos y desafíos muestran que la contratación de inteligencia artificial por parte del sector público exige pliegos más sofisticados y regulaciones específicas, capaces de equilibrar innovación tecnológica con seguridad jurídica, tutela de derechos y sostenibilidad institucional, a la vez que aseguren a los proveedores de sistemas tecnológicos la certeza y estabilidad necesarias para sostener inversiones y generar confianza legítima en el marco contractual.

En este contexto, resulta relevante analizar qué marco legal existe al presente en Argentina en esta materia.

4. MARCO NORMATIVO VIGENTE

4. 1. Régimen general sobre contrataciones públicas

La contratación de sistemas de inteligencia artificial por parte del sector público argentino carece aún de un marco normativo específico. En conse-

cuencia, su adquisición se encuentra actualmente regida por las disposiciones del régimen general de contrataciones públicas.

Dado el carácter federal del país, cada jurisdicción posee su propio régimen de contratación. No obstante, la mayoría de los ordenamientos locales tienden a alinearse, en cuanto a principios y estructura, con el marco aplicable a la Administración Pública Nacional, que será el objeto de este análisis.

El régimen nacional se encuentra principalmente regulado por: (i) el decreto delegado 1023/2001, que aprueba el Régimen de Contrataciones de la Administración Nacional (decreto 1023); (ii) su dec. reg. 1030/2016 (decreto 1030); y (iii) la disposición 63/2016, que aprueba el Pliego Único de Bases y Condiciones Generales (el Régimen de Contrataciones Públicas). Este conjunto normativo se complementa, en cada procedimiento concreto, con los pliegos de bases y condiciones particulares, los pliegos técnicos y el contrato o la orden de compra correspondiente.

Como regla general, el Régimen de Contrataciones Públicas se aplica a todos los contratos celebrados por la Administración Nacional [(15)], que se presumen de derecho público, salvo supuestos excepcionales en los que, atendiendo a su objeto, puedan quedar sujetos al derecho privado [(16)]. En su ámbito se incluyen, en particular, los contratos de servicios celebrados por las jurisdicciones y entidades comprendidas en la Administración Nacional [(17)].

Los procedimientos para contratar con la Administración Nacional incluyen: (i) la licitación pública, (ii) la licitación privada, (iii) la contratación directa y (iv) la subasta pública [(18)]. Los criterios para determinar qué procedimiento específico se aplicará en el contrato en cuestión dependerán del importe, la finalidad o el tipo de contrato.

---

(15) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 2° refiere a que el régimen será de aplicación obligatoria a todas las jurisdicciones y entidades comprendidas en el inciso a) del artículo 8º de la ley 24.156 y sus modificaciones. Ello incluye a la Administración Nacional, conformada por la Administración Central y los Organismos Descentralizados, comprendiendo en estos últimos a las Instituciones de Seguridad Social.

(16) Dec. 1023/2001, Régimen de Contrataciones de la Administración Nacional, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 1° establece: "Toda contratación de la Administración Nacional se presumirá de índole administrativa, salvo que de ella o de sus antecedentes surja que está sometida a un régimen jurídico de derecho privado."

(17) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 4. El art. 5 prevé los contratos excluidos (por ejemplo, empleo público, compras por caja chica, los que se celebren con estados extranjeros, con entidades de derecho público internacional, con instituciones multilaterales de crédito, los que se financien total o parcialmente con recursos provenientes de esos organismos, entre otros):

(18) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 24.

Una característica central del régimen es que la administración cuenta con prerrogativas especiales de derecho público, justificadas en el interés general, que no suelen estar presentes en los contratos entre particulares [19]. Entre ellas se destacan: (i) la facultad de redactar e interpretar el contrato, resolver dudas y modificarlo por razones de interés público [20]; (ii) declarar la caducidad, rescisión o cancelación del contrato y determinar sus efectos [21]; (iii) rescindir unilateralmente por razones de interés público, incluso sin reconocer lucro cesante [22]; (iv) aumentar o disminuir el objeto hasta un 20% en las condiciones pactadas, y hasta un 30% con aceptación del contratista (con un límite máximo del 35%) [23]; (v) controlar e inspeccionar la ejecución [24]; (vi) imponer sanciones por incumplimiento [25]; (vii) ejecutar directamente el objeto contractual en caso de incumplimiento, utilizando activos y recursos del contratista [26]; y (viii) prorrogar contratos de suministro o servicios periódicos en condiciones similares, con posibilidad de ajustar precios según el mercado [27].

Los contratistas, por su parte, deben continuar cumpliendo sus obligaciones aun frente a decisiones unilaterales de la administración, salvo supuestos de fuerza mayor o imposibilidad razonable derivada de actos de la propia autoridad [28].

Sin embargo, el ejercicio de estas prerrogativas debe ser razonable y no afectar al equilibrio económico-financiero del contrato, en cuyo caso el contratista podrá reclamar la recomposición del contrato [29].

---

(19) Para un análisis más profundo, nos remitimos al trabajo de Guglielminetti, Ana P., "Prerrogativas y garantías contractuales a través del prisma del financiamiento privado", Revista Derecho Administrativo, Número 106, 2016, cita online: AP/DOC/882/2016.

(20) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 12 inc. a).

(21) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 12 inc. a).

(22) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 12 inc. b), segunda parte.

(23) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 12 inc. b), primera parte.

(24) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 12 inc. c).

(25) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 12 inc. d).

(26) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 12 inc. e).

(27) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 12 inc. g).

(28) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 13 inc. c). Dec. 1030/2016, *Reglamento del Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 16 de septiembre de 2016, arts. 25 inc. a), 101 Anexo.

(29) Dec. 1023/2001, *Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 13 de agosto de 2001, art. 13 inc. a).

4.2. Regulación de la inteligencia artificial para el sector público

Pese a la inexistencia de un régimen de contratación sobre inteligencia artificial en el sector público, existen marcos normativos tanto internacionales como nacionales que buscan ofrecer respuestas a los desafíos que supone su utilización en ese ámbito.

La incorporación de la inteligencia artificial en la gestión pública abre un campo de oportunidades, pero también de incertidumbres jurídicas. Como explicamos, la ausencia de reglas claras sobre transparencia, explicabilidad o protección de datos personales genera tensiones entre la innovación tecnológica y la vigencia de los derechos fundamentales. Este escenario plantea la necesidad de examinar los marcos normativos.

A nivel internacional, destacan los Principios de la OCDE sobre Inteligencia Artificial centrados en el ser humano, la equidad, la seguridad y la transparencia y la Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial, que subraya la protección de derechos humanos, la diversidad cultural y la sostenibilidad. Ambos constituyen pilares del marco ético global que orienta las políticas nacionales.

En Argentina, este compromiso se refleja en la *Guía sobre Transparencia y Protección de Datos en Inteligencia Artificial Responsable de la AAIP*, la *Disposición 2/2023 sobre IA Fiable* de la Subsecretaria de Tecnologías de la Información (con principios de explicabilidad, equidad, responsabilidad y supervisión humana) y los lineamientos de la Red Iberoamericana. A ellos se suman normas complementarias (como las resoluciones sobre seguridad de datos, explicabilidad en decisiones automatizadas, fortalecimiento institucional, la creación de un espacio de diálogo interministerial y el Plan Nacional de IA) que conforman un marco integral para orientar el desarrollo y la implementación de esta tecnología.

Naturalmente, la aplicación de estas normas sobre inteligencia artificial al sector público debe ser realizada de manera armoniosa con el resto del ordenamiento jurídico, para evitar contradicciones normativas.

4.2.1. Marco internacional

Para comprender el diseño de los marcos regulatorios argentinos en materia de inteligencia artificial, es importante identificar y entender las fuentes internacionales que los inspiran. En particular, las guías nacionales más relevantes se inspiraron en dos instrumentos de referencia global: por un lado, la Recomendación del Consejo de la OCDE sobre Inteligencia Artificial (2019), y por otro, la Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial (2021).

Ambos documentos establecen un conjunto de principios rectores que orientan a los Estados hacia un uso responsable de la inteligencia artificial, aunque no refieren específicamente a cómo los Estados deben contratarla.

a. OCDE - Recomendación del Consejo sobre Inteligencia Artificial

El marco de principios elaborado por la OCDE en materia de inteligencia artificial se apoya en una serie de reconocimientos orientativos: que la confianza debería considerarse un vector esencial de la transformación digital; que la fiabilidad de los sistemas de inteligencia artificial podría constituir un factor decisivo para su adopción social, aun cuando las aplicaciones futuras y sus implicancias resulten difíciles de anticipar; y que sería necesario un debate público amplio e informado para aprovechar sus beneficios potenciales, limitando al mismo tiempo los riesgos asociados.

La OCDE, en su recomendación sobre inteligencia artificial, sostiene que los Estados miembros y no miembros adherentes [30] deberían promover y aplicar principios de administración responsable orientados a una inteligencia artificial fiable. Estos principios, concebidos como complementarios e interdependientes, buscan servir de guía para todos los actores involucrados en el ciclo de vida de los sistemas, de acuerdo con sus funciones y capacidades.

En primer lugar, se destaca el principio de crecimiento inclusivo, desarrollo sostenible y bienestar que busca alentar a las partes interesadas a implementar prácticas que promuevan beneficios tangibles para las personas y el planeta. Entre sus objetivos se encuentra el fortalecimiento de las capacidades humanas, el estímulo de la creatividad, la inclusión de poblaciones históricamente marginadas, la reducción de desigualdades sociales, económicas y de género, así como la protección de los entornos naturales.

Un segundo eje se refiere al respeto del Estado de derecho, los derechos humanos y los valores democráticos, lo que comprende la equidad y la privacidad. Se propone que los actores de la inteligencia artificial actúen en consonancia con normas jurídicas y principios democráticos, garantizando la no discriminación, la igualdad, la libertad, la dignidad y la autonomía de las personas, además de proteger la diversidad, la justicia social y los derechos laborales. Este principio implica también enfrentar fenómenos como la desinformación y la información engañosa, procurando mecanismos de supervisión humana y correctivos frente a usos indebidos, ya sean intencionales o no.

---

(30) Argentina no es miembro pleno, pero en 2019 adhirió a los *Principios sobre Inteligencia Artificial* aprobados por dicho organismo, convirtiéndose en un "no miembro adherente". Esta categoría implica la aceptación y compromiso de implementar los estándares fijados en el instrumento específico, sin gozar de los derechos de participación ni voto propios de los Estados miembros. Argentina se convirtió en candidato a la adhesión a la OCDE en 2022.

El principio de transparencia y explicabilidad apunta a que los sistemas de inteligencia artificial se desarrollen bajo parámetros de divulgación responsable. Esto supone proporcionar información significativa y comprensible, tanto respecto de las capacidades y limitaciones de los sistemas como de las interacciones que generan. Asimismo, cuando sea viable, debe informarse sobre los datos, factores y procesos que explican las decisiones o resultados, de modo que las personas afectadas puedan comprenderlos y, eventualmente, cuestionarlos.

La solidez, seguridad y protección constituyen otro de los pilares. Los sistemas de inteligencia artificial deberían diseñarse para ser robustos y seguros a lo largo de todo su ciclo de vida, evitando riesgos excesivos en condiciones normales o adversas de uso. Ello exige prever mecanismos de corrección, invalidación o desmantelamiento seguro ante conductas indeseadas, así como reforzar la integridad de la información sin menoscabar la libertad de expresión.

Finalmente, el principio de responsabilidad establece que los actores de la inteligencia artificial deben asumir el deber de garantizar el correcto funcionamiento de los sistemas y el respeto de los principios enunciados. Esto requiere trazabilidad de datos, procesos y decisiones, a fin de permitir un control efectivo de resultados y la implementación de respuestas adecuadas al contexto. También demanda la adopción de un enfoque sistemático de gestión de riesgos en todas las fases del ciclo de vida, incluyendo los vinculados con sesgos, derechos humanos, privacidad, seguridad, derechos laborales y propiedad intelectual. La cooperación entre distintos actores se presenta como un elemento clave para cumplir con este estándar de responsabilidad compartida.

Como corolario, la OCDE recomienda que los Estados adherentes incorporen estos principios en sus políticas nacionales y en su cooperación internacional, prestando especial atención al rol de las pequeñas y medianas empresas. Ello implica impulsar la inversión pública y privada en investigación y desarrollo de inteligencia artificial, fomentar un ecosistema digital inclusivo y seguro, promover entornos normativos flexibles e interoperables, preparar a las sociedades para la transformación del mercado laboral y garantizar una transición justa para los trabajadores, así como reforzar la cooperación internacional orientada a generar estándares técnicos compartidos y mecanismos de evaluación comparables.

b. UNESCO – Recomendación sobre la Ética de la Inteligencia Artificial

La Recomendación sobre la Ética de la Inteligencia Artificial constituye un instrumento normativo de alcance universal que busca orientar el desarrollo y aplicación de los sistemas de inteligencia artificial al servicio de la

humanidad, las sociedades y el medio ambiente, al tiempo que procura prevenir los posibles daños derivados de su uso. A diferencia de otros marcos éticos ya existentes, esta iniciativa no se limita a la enunciación de valores y principios, sino que avanza en la formulación de lineamientos concretos de política pública, con especial atención a la inclusión social, la igualdad de género y la protección de los ecosistemas.

Su finalidad radica en proporcionar a los Estados una base común para la elaboración de normas y políticas en consonancia con el derecho internacional, al mismo tiempo que orienta la actuación de individuos, comunidades, instituciones y empresas privadas hacia la incorporación de la ética en todas las fases del ciclo de vida de los sistemas de inteligencia artificial. En este sentido, la Recomendación enfatiza la necesidad de proteger y promover los derechos humanos, la dignidad y la igualdad, salvaguardando también los intereses de las generaciones futuras y preservando la diversidad cultural y medioambiental.

Asimismo, la recomendación se proyecta como un marco normativo que pretende conjugar principios universales con mecanismos prácticos de implementación, favoreciendo así una gobernanza global de la inteligencia artificial que sea respetuosa de los derechos fundamentales y del entorno natural.

La recomendación se dirige a los Estados Miembros [31] en su doble condición: por un lado, como actores en el desarrollo, implementación y uso de sistemas de inteligencia artificial; por otro, como autoridades encargadas de diseñar y aplicar marcos jurídicos y regulatorios que acompañen todas las fases del ciclo de vida de dichos sistemas. Asimismo, la recomendación pone de relieve la función de los Estados en la generación de condiciones regulatorias que fomenten la responsabilidad empresarial en el ecosistema de la inteligencia artificial. Argentina se adhirió formalmente a esta recomendación, comprometiéndose con los principios de transparencia, equidad, dignidad humana y supervisión humana del desarrollo de sistemas de inteligencia artificial, a través de la disposición 2/2023.

De manera complementaria, la recomendación establece directrices éticas para todos los actores de la cadena de valor de la inteligencia artificial. Estas pautas fijan criterios comunes de evaluación del impacto ético a lo largo de todo el ciclo de vida de los sistemas, favoreciendo un marco de gobernanza fundado en la legalidad, la transparencia y la rendición de cuentas.

---

(31) Argentina es Estado Miembro de la UNESCO desde 1948, integrándose al grupo latinoamericano y caribeño del organismo. Esto le otorga plena participación en sus instrumentos, políticas y convenciones.

La Recomendación se estructura en torno a los siguientes principios: (i) proporcionalidad e inocuidad; (ii) seguridad y protección; (iii) equidad y no discriminación; (iv) sostenibilidad; (v) derecho a la intimidad y protección de datos; (vi) supervisión y decisión humanas; (vii) transparencia y explicabilidad; (viii) responsabilidad y rendición de cuentas; (ix) sensibilización y educación; y (x) gobernanza y colaboración multiactor. Cada uno de estos principios se desarrolla con mayor profundidad en los apartados que siguen.

i. *Proporcionalidad e inocuidad:* En primer lugar, se afirma que las tecnologías de inteligencia artificial no aseguran por sí mismas la prosperidad humana ni ambiental. Todo proceso del ciclo de vida debe limitarse a lo estrictamente necesario para fines legítimos y adecuarse al contexto. Ante riesgos de daño a personas, derechos humanos, comunidades o ecosistemas, corresponde aplicar evaluaciones de riesgo y adoptar medidas preventivas idóneas.

La decisión de utilizar inteligencia artificial y el método seleccionado, entonces, deben:

a) Ser adecuados y proporcionales a un objetivo legítimo;

b) respetar los valores fundamentales y no implicar violaciones de derechos humanos; y,

c) fundarse en evidencia científica robusta y pertinencia contextual.

Las decisiones con impacto irreversible o de vida o muerte exigen decisión final humana. Se excluyen usos de inteligencia artificial para calificación social y vigilancia masiva.

Seguridad y protección: En segundo lugar, se recomienda la prevención, mitigación y, cuando sea posible, eliminación los daños no deseados y las vulnerabilidades a ataques durante todo el ciclo de vida. Ello requeriría de reglas sostenibles de acceso a datos, respetuosos de la privacidad, y el uso de datos de calidad para entrenamiento y validación de modelos.

Equidad y no discriminación: En tercer lugar, los Estados deben reducir brechas digitales y de capacidades, asegurar acceso significativo y ofrecer recursos efectivos frente a sesgos y decisiones discriminatorias. A nivel internacional, los países tecnológicamente más avanzados tienen un deber de solidaridad que favorezca un orden informativo, cultural, educativo y socioeconómico más equitativo.

Sostenibilidad: Se recomienda una evaluación continua de impactos humanos, sociales, culturales, económicos y ambientales, alineada con los

Objetivos de Desarrollo Sostenible [32], reconociendo la naturaleza dinámica de estas metas.

Derecho a la intimidad y protección de datos: En quinto lugar, se sostiene que la privacidad, en tanto presupuesto de la dignidad, la autonomía y la capacidad de agencia, debe ser respetada, protegida y promovida a lo largo de todo el ciclo de vida de los sistemas de inteligencia artificial.

Los marcos de protección de datos y los mecanismos conexos deberían de inspirarse en los principios y normas internacionales [33] aplicables a la recopilación, uso y divulgación de datos personales, así como al ejercicio efectivo de los derechos de los titulares, asegurando siempre un objetivo legítimo y una base jurídica válida para su tratamiento, incluido el consentimiento informado. Asimismo, los sistemas algorítmicos deben someterse a evaluaciones de impacto en la privacidad que contemplen también consideraciones sociales y éticas, incorporando innovaciones propias del enfoque de "privacidad desde el diseño". En este sentido, los actores de la inteligencia artificial asumen la responsabilidad de concebir y aplicar sistemas que garanticen la protección de la información personal en todas las fases de su desarrollo y utilización.

Supervisión y decisión humanas: En sexto lugar, se destaca la importancia de poder atribuir responsabilidad ética y jurídica a personas físicas o jurídicas. La supervisión debe ser individual y también pública.

Así, sostiene que la eficiencia no justifica abdicar del control: los sistemas pueden asistir, pero no desplazan la responsabilidad última ni la rendición de cuentas humanas.

i. Transparencia y explicabilidad: Este principio hace a las condiciones previas para proteger derechos y garantizar regímenes de responsabilidad eficaces. Su alcance debería ajustarse al contexto y equilibrarse con privacidad y seguridad. En particular, se busca que las personas estén informadas cuando decisiones se basan en inteligencia artificial, teniendo derecho a explicaciones comprensibles, y acceder a revisión humana con capacidad de rectificación. En casos graves con efectos sobre derechos, la transparen-

---

(32) Los Objetivos de Desarrollo Sostenible constituyen el núcleo de la *Agenda 2030 para el Desarrollo Sostenible*, adoptada por la Asamblea General de las Naciones Unidas en septiembre de 2015. Se trata de 17 metas globales interrelacionadas destinadas a erradicar la pobreza, proteger el planeta y garantizar el bienestar humano, que abarcan desde la educación y la salud hasta la igualdad de género, la justicia y la acción climática. [Ver: *Objetivos de Desarrollo Sostenible* en https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/].

(33) Destacamos que no existe un tratado internacional universalmente aceptado en materia de protección de datos personales. El mayor esfuerzo de alcance internacional se ha materializado en el Convenio 108 del Consejo de Europa (Convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal, 1981) y en su versión modernizada, el Convenio 108+ (2018), que han servido como referentes globales, aunque no cuentan con adhesión universal. Argentina, sin embargo, es parte y ha ratificado ambos Convenios.

cia podría exigir mayor apertura técnica (p. ej., sobre datos o garantías de equidad).

ii. Responsabilidad y rendición de cuentas: este principio, busca que tanto el Estado como los distintos actores de la inteligencia artificial deben respetar derechos humanos y proteger el ambiente, asumiendo responsabilidades éticas y jurídicas a lo largo del ciclo de vida, también respecto de actores bajo su jurisdicción o control efectivo. En particular, se sugiere la implementación de mecanismos de supervisión, evaluación de impacto, auditoría y diligencia debida con protección a denunciantes, así como auditabilidad y trazabilidad técnicas e institucionales.

iii. Sensibilización y educación: se promueve la alfabetización en inteligencia artificial, datos y ética mediante educación abierta, participación cívica y capacitación multiactor (gobiernos, organismos internacionales, sociedad civil, academia, medios, líderes comunitarios y sector privado), respetando diversidad lingüística y cultural. La formación debe centrarse en impactos sobre los derechos y el ambiente.

iv. Gobernanza y colaboración multifactorial: finalmente, el uso de datos debe respetar la soberanía nacional conforme al derecho internacional. La gobernanza de la inteligencia artificial exige participación de gobiernos, comunidad técnica, sociedad civil, academia, medios, educadores, responsables de políticas, sector privado y órganos de derechos humanos y de igualdad, con estándares abiertos e interoperabilidad.

En conclusión, aunque carece de carácter vinculante, la recomendación constituye un consenso ético que integra derechos humanos, sostenibilidad y principios democráticos, y resulta particularmente significativa para Estados como la Argentina, que han manifestado su compromiso de orientar sus políticas públicas conforme a dichos estándares. Su interpretación debe ser integral, reconociendo que los valores y principios enunciados son complementarios e interdependientes. El desafío central, no obstante, reside en su efectiva traducción al plano normativo y regulatorio interno, evitando que permanezcan en el ámbito meramente declarativo.

Así, tanto la recomendación de la UNESCO como los Principios de la OCDE constituyen referentes internacionales orientados a promover una inteligencia artificial fiable, inclusiva y respetuosa de los derechos humanos. Aunque difieren en su alcance, la UNESCO propone un marco ético universal de carácter declarativo, mientras que la OCDE privilegia lineamientos de gobernanza y políticas públicas con vocación de implementación práctica, ambos coinciden en subrayar valores comunes como la transparencia, la responsabilidad, la supervisión humana, la equidad y la sostenibilidad. Estas coincidencias delinean un consenso internacional incipiente que, si

bien carece aún de fuerza vinculante, ofrece un marco de referencia para los Estados en el diseño de sus políticas y marcos regulatorios.

c. Ámbito Iberoamericano

En el año 2019, la Red Iberoamericana de Protección de Datos (RIPD) aprobó dos instrumentos de especial relevancia para la regulación de la inteligencia artificial en la región: las *Recomendaciones Generales para el Tratamiento de Datos en Inteligencia Artificial* y las *Orientaciones Específicas para el Cumplimiento de los Principios y Derechos que Rigen la Protección de los Datos Personales en los Proyectos de Inteligencia Artificial*. [34] Ambos documentos deben entenderse como partes de un mismo marco normativo: las recomendaciones fijan principios y pautas de carácter general, mientras que las orientaciones tienen la función de reglamentarlas, ofreciendo ejemplos y prácticas concretas para su implementación.

Las recomendaciones generales constituyen un marco amplio y transversal cuyo propósito es orientar a los Estados Iberoamericanos en la incorporación de salvaguardas adecuadas frente a los riesgos que plantea la inteligencia artificial. Allí se subraya la importancia de principios como la transparencia, la proporcionalidad, la responsabilidad y el respeto a los derechos fundamentales. Entre las recomendaciones se destaca la necesidad de: cumplir con las normas locales de protección de datos; efectuar estudios de impacto en privacidad en los casos de alto riesgo; incorporar la privacidad, la ética y la seguridad desde el diseño y por defecto; materializar el principio de responsabilidad proactiva; y establecer esquemas adecuados de gobernanza en las organizaciones que desarrollan productos de inteligencia artificial. A ello se suman obligaciones tales como garantizar la calidad de los datos, utilizar técnicas de anonimización cuando sea posible, respetar los derechos de los titulares y adoptar medidas de confianza y transparencia en el tratamiento de datos personales.

Ahora bien, estos principios no quedarían completos si no existiera un desarrollo operativo que los traduzca en obligaciones concretas. En este punto entran en juego las orientaciones específicas, que se presentan como la continuación natural de las recomendaciones generales. Su finalidad es bajar a tierra los conceptos allí enunciados y brindar a los responsables y encargados del tratamiento criterios prácticos para hacer efectivos los fines de licitud, transparencia, proporcionalidad y responsabilidad. Se trata, por tanto, de un documento que transforma los principios en exigencias técni-

(34) Las recomendaciones de la Red Iberoamericana de Protección de Datos, si bien carecen de fuerza jurídica vinculante, constituyen un instrumento de *soft law* de particular relevancia para Argentina. En su calidad de miembro activo de la Red, el país interviene en su elaboración y, a través de la Agencia de Acceso a la Información Pública, las emplea habitualmente como parámetro interpretativo y guía de orientación en la práctica regulatoria.

cas y jurídicas, vinculando el plano abstracto de los estándares con el plano concreto de los proyectos de inteligencia artificial.

Así, las orientaciones desarrollan de manera detallada cada uno de los principios rectores que implican las recomendaciones generales. El principio de legitimación exige que los datos se traten con consentimiento válido o sobre una base legítima, estableciendo mecanismos claros y accesibles para obtenerlo y documentarlo, con especial atención a datos sensibles o de menores. El principio de licitud impone que el tratamiento respete el derecho interno, el derecho internacional y los derechos fundamentales, lo que en el ámbito público significa actuar dentro de las facultades legales y, en el ámbito privado, incorporar las obligaciones en contratos y acuerdos. El principio de lealtad requiere privilegiar los intereses del titular, evitando medios engañosos o discriminaciones arbitrarias, e impone medidas de monitoreo constante para mitigar sesgos. El principio de transparencia obliga a informar al titular sobre las características del tratamiento y, en inteligencia artificial, se traduce en comunicar el uso de procesos automatizados, explicar de forma comprensible la lógica de los algoritmos y ofrecer mecanismos para mantener el control sobre los datos. El principio de finalidad prohíbe usos incompatibles con los fines originales e impone definir y comunicar claramente las finalidades desde etapas tempranas de desarrollo. El principio de proporcionalidad restringe la recolección de datos a lo estrictamente necesario y promueve el uso de técnicas como la pseudonimización o los datos sintéticos. El principio de calidad exige mantener los datos exactos y actualizados, revisando periódicamente los algoritmos y aplicando técnicas de borrado seguro. El principio de responsabilidad obliga a demostrar cumplimiento, mediante estándares, auditorías, capacitación, documentación y mecanismos de explicabilidad. El principio de seguridad demanda medidas técnicas, físicas y administrativas robustas, incluyendo la notificación de incidentes, la resiliencia a ataques y la trazabilidad de los sistemas. Por último, el principio de confidencialidad impone que solo personas y procesos autorizados puedan tratar los datos, lo que en inteligencia artificial se traduce en controles contractuales y medidas de anonimización o pseudonimización.

Pero el valor de las orientaciones no se limita a los principios. También regulan las obligaciones de los encargados del tratamiento. Se establece que los encargados deben actuar siempre bajo instrucciones del responsable, sin poder de decisión propio sobre alcance o finalidad, y que esta relación debe formalizarse mediante contratos que definan objeto, duración, categorías de datos y obligaciones claras en materia de seguridad, confidencialidad, notificación de incidentes y devolución o supresión de datos. En el caso de proyectos de inteligencia artificial, se subraya la importancia de la supervisión constante de proveedores externos y la inclusión de instrucciones con-

tractuales específicas, dado que la frontera entre responsable y encargado puede volverse difusa.

Finalmente, las orientaciones dedican un apartado específico a los derechos de los titulares, desarrollando mecanismos para garantizar su ejercicio efectivo en el contexto de la inteligencia artificial. Se trata de los derechos de acceso, rectificación, cancelación, oposición, portabilidad, limitación del tratamiento y a no ser objeto de decisiones automatizadas. En términos prácticos, se exige que los sistemas permitan a los titulares acceder a información clara sobre la lógica básica de los algoritmos y sus efectos, impugnar decisiones automatizadas, solicitar intervención humana y administrar sus preferencias de privacidad mediante herramientas accesibles. También se promueve la creación de paneles de control o mecanismos similares que permitan a las personas actualizar, corregir o suprimir sus datos de manera sencilla.

Las *Recomendaciones Generales* y las *Orientaciones Específicas* conforman un andamiaje normativo y práctico indivisible: las primeras fijan los principios rectores de la protección de datos en la inteligencia artificial y las segundas los traducen en obligaciones concretas para responsables y encargados. Su interdependencia asegura que los Estados y, en particular, las entidades públicas que contratan sistemas de inteligencia artificial, cuenten con un marco que combina estándares internacionales de protección de derechos fundamentales con prácticas operativas adaptadas a los desafíos específicos de la era digital.

En este marco, los lineamientos de la RIPD adquieren una relevancia particular para las entidades públicas, que en su calidad de responsables del tratamiento están obligadas a acreditar el cumplimiento de los principios de legitimación, licitud, lealtad y transparencia. Sin embargo, su impacto se extiende también a las entidades privadas que proveen soluciones de inteligencia artificial al sector estatal, puesto que los contratos deben prever cláusulas precisas en materia de seguridad, confidencialidad, gobernanza de datos y notificación de incidentes. Las orientaciones remarcan que, si un encargado se aparta de las instrucciones recibidas y decide por sí mismo sobre los fines o medios del tratamiento, pasará a ser considerado responsable, asumiendo plenamente las obligaciones correspondientes.

4.2.2. Marco normativo argentino

En Argentina se ha asumido, a través del Plan Nacional de Inteligencia Artificial, el compromiso de incorporar los principios y recomendaciones internacionales en materia de inteligencia artificial, con el propósito de afrontar los problemas y desafíos que plantea su integración en procesos y sistemas que pueden generar consecuencias tanto directas como indirectas. Estos desafíos se manifiestan a lo largo de todo el ciclo de vida de los

sistemas de inteligencia artificial, acompañando a la tecnología desde su diseño inicial hasta su operación y mantenimiento, y requieren de un abordaje que respete "derechos humanos, valores democráticos, normativas establecidas, principios éticos..." (35).

En este marco, se establece que el Estado argentino debe "impulsar a través de la adopción de inteligencia artificial un Estado ágil, eficiente y moderno" (36). Para ello, las entidades públicas asumen el compromiso de garantizar que la implementación de estas tecnologías se realice bajo criterios de transparencia, trazabilidad y fundamentación, no solo en la provisión de servicios públicos sino también en la gestión interna de la administración. De este modo, las obligaciones estatales incluyen desde la normalización de datos y la elaboración de pliegos específicos para la adquisición de soluciones de inteligencia artificial, hasta la conformación de equipos técnicos especializados y la captación de funcionarios, todo ello orientado a asegurar que la adopción de la inteligencia artificial en el sector público se traduzca en beneficios concretos para la ciudadanía.

El Estado argentino, a través del cuerpo normativo (en específico, la *Guía sobre Transparencia y Protección* y las *Recomendaciones para una inteligencia artificial fiable*) ha identificado la necesidad de anticipar distintos riesgos, como, por ejemplo:

1. Sesgos y discriminación: reproducción de prejuicios raciales, de género o sociales en ámbitos sensibles como el reconocimiento facial, la vigilancia laboral o las decisiones judiciales. Pueden presentarse distintos tipos de sesgos (37), entre ellos: (a) sesgo de percepción (38), (b) sesgo técnico (39), (c) sesgo de modelado (40) y (d) sesgo de activación (41).

2. Deficiente calidad de los datos: utilización de información desactualizada, errónea u obtenida mediante prácticas cuestionables, lo que compromete la legitimidad del sistema y afecta la privacidad de las personas.

---

(35) Ministerio de Educación, Cultura, Ciencia y Tecnología de la Nación, *Plan Nacional de Inteligencia Artificial*, Buenos Aires, 2019, p. 12.

(36) Ministerio de Educación, Cultura, Ciencia y Tecnología de la Nación, *Plan Nacional de Inteligencia Artificial*, Buenos Aires, 2019, p. 21.

(37) Agencia de Acceso a la Información Pública (AAIP), Guía para entidades públicas y privadas en materia de Transparencia y Protección de Datos Personales para una Inteligencia Artificial responsable, Buenos Aires, 2024, p. 18.

(38) Cuando los datos recopilados no representan adecuadamente a todas las poblaciones, el sistema de inteligencia artificial puede funcionar de forma desigual, favoreciendo o perjudicando a ciertos grupos.

(39) La tecnología en sí misma puede generar sesgos o errores, especialmente cuando los algoritmos funcionan mejor con ciertas variables o características y no con otras.

(40) Cuando un modelo es diseñado manualmente por expertos, puede presentar limitaciones porque no siempre considera todos los factores del entorno, ya sea de forma intencional o por desconocimiento.

(41) El sesgo puede generarse cuando los resultados de un sistema de inteligencia artificial son utilizados de forma parcial o injusta en el entorno donde se aplican.

3. Vulneración de la privacidad: recopilación y análisis masivo de datos personales sin base legal válida.

4. Riesgos de seguridad: posibilidad de ciberataques que manipulen modelos, expongan datos sensibles o comprometan la integridad del sistema.

5. Reidentificación de datos anónimos: riesgo de volver a identificar a personas a partir de datos supuestamente anonimizados, con impacto en su privacidad.

6. Uso indebido de datos personales: empleo de la información para fines distintos a los informados al titular, incluyendo su venta o explotación comercial sin consentimiento.

7. Suplantación de identidad mediante deepfakes: manipulación audiovisual para reemplazar identidades, dañar reputaciones o difundir desinformación.

8. Vigilancia sin consentimiento: monitoreo masivo de empleados o ciudadanos sin autorización, con grave afectación de la privacidad y los derechos fundamentales.

9. Opacidad y falta de transparencia: dificultades para explicar cómo los algoritmos toman decisiones, lo que debilita la confianza pública y aumenta los riesgos de abuso.

a. Guías y recomendaciones argentinas sobre inteligencia artificial

Frente a estos desafíos, adquieren particular relevancia dos instrumentos locales: la Guía de la AAIP sobre Transparencia y Protección de Datos Personales en Inteligencia Artificial y las Recomendaciones para una Inteligencia Artificial Fiable. Ambos buscan traducir los principios internacionales al contexto argentino y dotar a las entidades públicas de pautas que fortalezcan la legitimidad en el uso de la inteligencia artificial. No obstante, su carácter es meramente orientativo, ya que no constituyen normas obligatorias.

Ambos documentos procuran proyectar en el ámbito argentino los principios internacionales previamente examinados, adecuándolos a las particularidades de nuestro entramado normativo e institucional. Si bien carecen de fuerza vinculante en sentido estricto, se erigen como referentes, en tanto ofrecen parámetros concretos para encauzar la incorporación y el uso de tecnologías de inteligencia artificial en el sector público, dotando de legitimidad y previsibilidad a dichas prácticas.

- Disposición 2/2023 de la Jefatura de Gabinete de Ministros Subsecretaría de Tecnologías de la Información: Recomendaciones para una Inteligencia Artificial Fiable

El documento *Recomendaciones para una Inteligencia Artificial Fiable* constituye un marco integral destinado a orientar la implementación ética y responsable de proyectos de inteligencia artificial, particularmente en el sector público argentino. Su finalidad principal es asegurar que el despliegue de esta tecnología se realice con un enfoque centrado en el ciudadano y basado en derechos, de modo que la inteligencia artificial sirva al bien común mientras se mitigan los riesgos asociados a su utilización [(42)].

En primer lugar, se aclara la naturaleza de la inteligencia artificial, diferenciando entre la inteligencia artificial estrecha (diseñada para ejecutar tareas específicas) y la teórica inteligencia artificial general (equiparable a la inteligencia humana). Se resalta que la inteligencia artificial, como artificio tecnológico, no posee atributos humanos tales como consciencia, intención, autodeterminación o libertad. Por ello, resulta esencial comprender que la responsabilidad jurídica y ética última recae siempre en los actores humanos responsables de su programación, entrenamiento, implementación y uso. Atribuir responsabilidad autónoma a un algoritmo supondría un error antropomórfico que dificultaría la regulación eficaz y la asignación clara de responsabilidades [(43)].

Este planteo se inscribe en un ecosistema internacional que ha avanzado de manera significativa en la formulación de principios éticos aplicables a la inteligencia artificial esgrimidos, como mencionábamos antes, en la Recomendación de la UNESCO y los Principios de la OCDE, posteriormente adoptados por el G20, ponen el acento en el crecimiento inclusivo, el desarrollo sostenible, la explicabilidad de los sistemas, la robustez tecnológica y la responsabilidad desde una perspectiva humano-céntrica. A estas referencias se suman los Principios de Asilomar sobre inteligencia artificial (2017), elaborados en el ámbito académico y tecnológico, que subrayan la importancia de la seguridad, la transparencia frente a fallos, la responsabilidad de los diseñadores y la necesidad de alinear los sistemas con los valores humanos.

La disposición 2/2023 de la Subsecretaria de Tecnologías de la Información traslada estas orientaciones internacionales al plano operativo. Antes de iniciar cualquier proyecto de inteligencia artificial, recomienda, de hecho, conformar equipos humanos diversos y multidisciplinarios, evaluar el nivel de concientización existente en las organizaciones y determinar si la inteligencia artificial es efectivamente la herramienta adecuada para el problema planteado. Asimismo, se enfatiza la necesidad de definir con precisión el destino de uso de la tecnología, establecer mecanismos de control

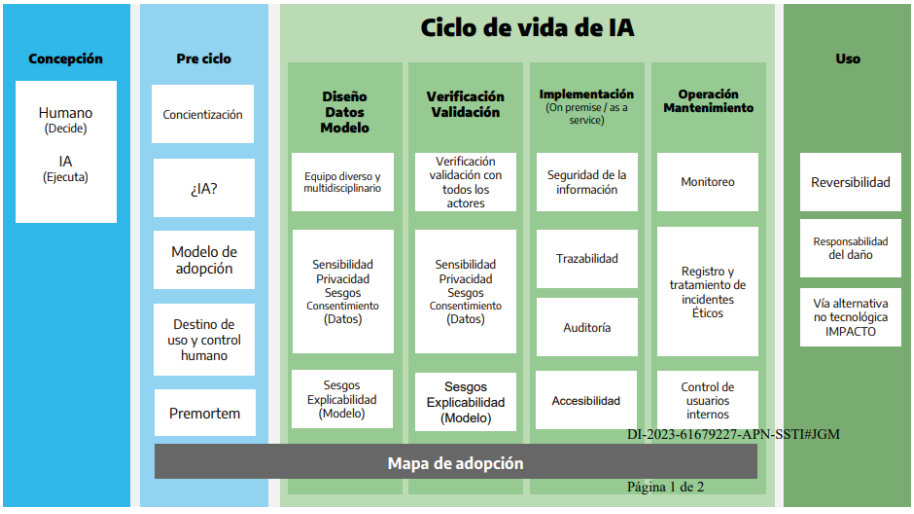
(42) Recomendaciones para una inteligencia artificial fiable, Documento oficial, Buenos Aires, 2022, pp. 1-3.

(43) Recomendaciones para una inteligencia artificial fiable, Documento oficial, Buenos Aires, 2022, pp. 2-4.

humano efectivos y realizar un análisis *premortem* que permita anticipar fallas y riesgos antes de que el sistema entre en funcionamiento[6].

Durante el ciclo de vida de la inteligencia artificial, se plantean recomendaciones éticas específicas para cada etapa. En el diseño y modelado de datos, se requiere un punto de partida común para el equipo, la adecuada clasificación de la información y la selección cuidadosa de las fuentes, evitando sesgos y garantizando la calidad de los datos. En la fase de verificación y validación, se deben incorporar compromisos éticos explícitos, asegurar la transparencia de los modelos y documentar de manera formal los procesos de prueba y supervisión. En la etapa de implementación, las prioridades giran en torno a la seguridad de la información, la trazabilidad de las acciones, la auditabilidad de los sistemas y el cumplimiento de estándares de accesibilidad tecnológica. Finalmente, en la operación y el mantenimiento, se impone un monitoreo constante del desempeño ético de los sistemas, la gestión adecuada de incidentes y el establecimiento de controles internos que eviten malas prácticas o usos indebidos.

El ciclo de vida puede visualizarse de manera esquemática en el gráfico que acompaña la disposición:



El documento también extiende sus recomendaciones más allá del ciclo técnico, proponiendo una vigilancia ética continua en la fase posterior a la implementación. En este sentido, resulta indispensable contar con mecanismos de remediación inmediata frente a los perjuicios que puedan ocasionar los sistemas, institucionalizar procedimientos de aprendizaje organizacional y garantizar vías humanas accesibles para los ciudadanos que carezcan de acceso tecnológico o que prefieran interactuar con agentes

humanos. Esto asegura que la inteligencia artificial no genere exclusiones y que el acceso a los servicios estatales mantenga un carácter equitativo [44].

Un aspecto que merece especial atención son las implicancias de estas recomendaciones para los terceros que contratan con el Estado. El documento reconoce expresamente la figura de los proveedores privados como actores que desarrollan o suministran tecnologías de inteligencia artificial destinadas a la Administración pública. En tales casos, subraya que lo que se transfiere a los proveedores es la ejecución técnica de las tareas, pero no la responsabilidad última sobre los resultados, que permanece en cabeza del Estado. Esta delimitación no implica que los privados carezcan de obligaciones, pero sí aclara que su rol debe entenderse dentro de un marco de colaboración en el cual la decisión final y la rendición de cuentas corresponden a la autoridad pública.

En este sentido, para el sector privado estas pautas deben interpretarse como buenas prácticas más que como imposiciones normativas. Asegurar la licitud y trazabilidad de los datos empleados, ofrecer servicios en la nube con mecanismos de auditoría adecuados, cumplir con estándares de seguridad de la información y accesibilidad tecnológica, así como documentar procesos que refuercen la transparencia, constituyen medidas recomendadas que aumentan la confianza y la competitividad frente a procesos de contratación estatal. Aunque no son vinculantes, su adopción permite a los proveedores posicionarse como socios estratégicos confiables y alineados con criterios éticos y de calidad en la relación público-privada.

- Guía de la AAIP sobre Transparencia y Protección de Datos Personales en inteligencia artificial

La AAIP publicó la *Guía para entidades públicas y privadas en materia de Transparencia y Protección de Datos Personales para una Inteligencia Artificial responsable*. Este documento se inscribe en el marco del *Programa Nacional de Transparencia y Protección de Datos Personales en el uso de la Inteligencia Artificial* (Resolución 161/2023 de la AAIP), cuyo objetivo es impulsar procesos de análisis, regulación y fortalecimiento de capacidades institucionales destinados a acompañar el desarrollo y la utilización de la inteligencia artificial tanto en el sector público como en el privado. La finalidad de la guía es atender los riesgos y preocupaciones asociados al uso de sistemas de inteligencia artificial, tanto en aquellos procesos orientados a personalizar la experiencia de los usuarios o a fundamentar decisiones basadas en evidencia, como en los que persiguen automatizar procedimientos, reducir tiempos o mejorar resultados.

---

(44) Recomendaciones para una inteligencia artificial fiable, Documento oficial, Buenos Aires, 2022, pp. 13-15.

El texto fue elaborado por los equipos técnicos de la AAIP con el propósito de abarcar, al igual que las *Recomendaciones para una Inteligencia Artificial Fiable*, el ciclo de vida completo de los sistemas de inteligencia artificial, aunque con especial énfasis en dos dimensiones centrales: la transparencia algorítmica y la protección de datos personales. La guía se dirige a un amplio espectro de destinatarios: organismos públicos, empresas proveedoras y desarrolladoras de soluciones de inteligencia artificial, instituciones académicas y organizaciones de la sociedad civil, y establece lineamientos para integrar la transparencia y la protección de datos en todas las fases de desarrollo y uso de la inteligencia artificial, garantizando que la innovación tecnológica no implique retrocesos en la tutela de los derechos fundamentales.

Una de las distinciones relevantes que plantea el documento reside en diferenciar los sistemas de decisión automatizada tradicionales, basados en reglas predefinidas y rígidas, de los sistemas de inteligencia artificial que incorporan técnicas de aprendizaje automático y redes neuronales, capaces de reconocer patrones, procesar lenguaje natural, adoptar decisiones y adaptarse a nuevas situaciones.

En línea con la Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial (2021) y los criterios orientadores de la Red Iberoamericana de Protección de Datos Personales, la Guía identifica riesgos como: sesgos y discriminación, deficiencias en la calidad de los datos, vulneraciones a la seguridad y a la privacidad, uso excesivo o inadecuado de información personal, fraudes de identidad mediante *deepfakes*, vigilancia masiva sin consentimiento y falta de transparencia en los procesos algorítmicos. Asimismo, advierte sobre desafíos emergentes como el uso de neurodatos, que exponen información relativa a pensamientos, emociones o capacidades cognitivas, con riesgos inéditos para la intimidad y la autodeterminación informativa.

Para mitigar los riesgos que plantea el uso de la inteligencia artificial, la Guía retoma los principios rectores previstos en la ley 25.326 de Protección de Datos Personales y los proyecta sobre el ciclo de vida de los sistemas algorítmicos. En primer lugar, se destaca el principio de licitud del tratamiento, que exige que toda recolección y utilización de datos personales se realice conforme a la normativa vigente y con una base legal adecuada, asegurando así el respeto a la autodeterminación informativa (arts. 5° y 6°). Unido a ello, el consentimiento informado se erige como la regla general de legitimidad: debe ser previo, expreso e informado, salvo en los supuestos de excepción previstos por la ley (art. 5°). También resulta esencial la finalidad legítima, que obliga a que los datos sean obtenidos con un propósito determinado, explícito y lícito, y que no se los desvíe hacia usos distintos o incompatibles con los que motivaron su recolección (art. 4.2).

La ley también establece el principio de calidad de los datos, que impone que la información sea veraz, adecuada, pertinente y no excesiva en relación con los fines perseguidos, evitando acumulaciones indiscriminadas. A ello se suma la exigencia de seguridad, que compromete a los responsables del tratamiento a adoptar medidas técnicas y organizativas que resguarden la integridad de los datos y prevengan accesos indebidos, pérdidas o alteraciones (arts. 9° y 10). La confidencialidad, en este marco, se configura como un deber de secreto profesional para todos aquellos que intervienen en el tratamiento de datos personales (art. 10), reforzando la protección frente a divulgaciones no autorizadas. El principio de minimización se desprende de estas previsiones y apunta a que únicamente se recojan y procesen los datos estrictamente necesarios y proporcionales a la finalidad legítima perseguida (art. 4.1). Finalmente, la ley reconoce a los titulares el pleno ejercicio de sus derechos de acceso, rectificación, actualización y supresión (arts. 14 y ss.), que constituyen la herramienta fundamental para mantener el control sobre la información personal y garantizar, en definitiva, la efectividad del derecho fundamental a la protección de datos.

En relación con la transparencia, la guía enfatiza la necesidad de garantizar la trazabilidad y auditabilidad de los sistemas, así como la explicabilidad de las decisiones algorítmicas, de modo que las personas comprendan cómo y por qué se adoptan decisiones que las afectan y dispongan de vías claras para reclamar y obtener supervisión efectiva.

El enfoque propuesto por la AAIP comprende todas las etapas del ciclo de vida de los sistemas de inteligencia artificial (diseño, verificación y validación, implementación, y operación y mantenimiento) e impulsa la incorporación de prácticas de responsabilidad proactiva. Entre ellas, se destaca la *evaluación de impacto en protección de datos personales,* entendida como una herramienta preventiva para identificar riesgos antes de la puesta en marcha de un tratamiento de datos, en consonancia con la Guía conjunta elaborada por la AAIP y la Unidad Reguladora y de Control de Datos Personales de Uruguay, que propone criterios de proporcionalidad, legalidad y mitigación de riesgos en proyectos tecnológicos de alto impacto. El principio de *privacidad por diseño y por defecto,* derivado de los arts. 4° y 9° de la ley 25.326, exige que los sistemas recolecten únicamente datos adecuados, pertinentes y no excesivos, incorporando medidas de seguridad desde su concepción. El *fortalecimiento de la seguridad técnica y organizativa,* previsto en los arts. 9° y 10, obliga a implementar medidas que aseguren integridad y confidencialidad frente a accesos indebidos o alteraciones. La *minimización y control de calidad de los datos* responde a la obligación de asegurar información exacta, actualizada y proporcional a la finalidad (arts. 4.1 y 4.3). Por último, la conformación de *equipos diversos y multidisciplinarios* constituye una estrategia para reducir sesgos en los modelos algorítmicos y para garantizar un abordaje plural de los riesgos técnicos, sociales y jurídicos. Si

bien estas pautas no poseen carácter vinculante, constituyen estándares de referencia que reflejan una tendencia regulatoria en consolidación, tanto a nivel nacional como internacional.

Si bien la Guía de la AAIP se dirige al conjunto de actores que desarrollan o utilizan sistemas de inteligencia artificial, adquiere especial relevancia para el sector privado, en particular para aquellas empresas proveedoras y desarrolladoras que interactúan con el Estado. El documento no introduce nuevas obligaciones jurídicas, pero sí establece un estándar de referencia que, en la práctica, puede transformarse en criterio de evaluación dentro de los procesos de contratación pública. En este marco, prácticas como la realización de evaluaciones de impacto en protección de datos, la incorporación de la privacidad por diseño y por defecto, la trazabilidad y auditabilidad de los sistemas, así como la exigencia de explicabilidad en los resultados, se configuran como buenas prácticas. Aunque carecen de carácter obligatorio, pueden otorgar ventajas competitivas a quienes las adopten y refuerzan la línea argumental de la AAIP en favor de una responsabilidad proactiva, aun cuando esta carezca todavía de sustento normativo específico más allá del Convenio 108+, ya ratificado por la Argentina, pero aún no vigente.

b. Normas complementarias

Junto a los instrumentos normativos centrales recién explicados, existen otras disposiciones que, si bien no regulan de manera integral la inteligencia artificial, resultan importantes para comprender el entramado regulatorio vigente en la Argentina. Estas normas introducen criterios de gobernanza, seguridad de la información y coordinación institucional que complementan y operativizan las obligaciones principales, consolidando un marco normativo en construcción.

En primer lugar, el Plan Nacional de Inteligencia Artificial (2019) fue diseñado como una hoja de ruta estratégica de largo plazo. Parte de la premisa de que las entidades públicas deben liderar la adopción de estas tecnologías a fin de optimizar la gestión y mejorar la provisión de servicios. Su alcance abarca desde la modernización de procesos administrativos hasta la creación de servicios "trazables, con lógica fundamentada y transparente, que no afecten derechos de los ciudadanos". Entre sus objetivos específicos se incluyen la normalización de datos públicos prioritarios, la elaboración de criterios y pliegos para la contratación de soluciones de inteligencia artificial y la conformación de equipos técnicos capaces de acompañar la implementación de proyectos. Asimismo, el plan promueve la capacitación de funcionarios y agentes del Estado y el desarrollo de proyectos piloto, con el fin de evaluar impactos y garantizar que la incorporación de la inteligencia artificial en el sector público se traduzca en beneficios concretos y verificables para la ciudadanía.

La decisión administrativa 750/2023 de la Jefatura de Gabinete de ministros crea la *Mesa Interministerial sobre Inteligencia Artificial*, concebida como un espacio institucional de diálogo y coordinación entre diferentes áreas del Estado. Su finalidad es analizar los impactos sociales, económicos y culturales de la inteligencia artificial y promover políticas públicas basadas en evidencia. Este foro habilita la participación de diversos ministerios y, eventualmente, de actores privados y académicos, lo que contribuye a la construcción de un esquema de gobernanza multiactor que resulta esencial frente a la naturaleza transversal y disruptiva de la inteligencia artificial.

La res. 161/2023 de la AAIP establece el *Programa Nacional de Transparencia y Protección de Datos Personales en el uso de la Inteligencia Artificial*, orientado a fortalecer las capacidades estatales en esta materia. El programa contempla tareas de diagnóstico institucional, capacitación y asistencia técnica, y retoma principios rectores como la transparencia, la explicabilidad y la trazabilidad. Su propósito es dotar a la administración de herramientas operativas para trasladar estos principios a la práctica mediante guías de buenas prácticas, recomendaciones y procesos participativos que involucren a distintos sectores sociales y económicos.

La res. 47/2018 de la AAIP establece medidas de seguridad para el tratamiento y conservación de datos personales en medios informatizados. Aunque no fue concebida específicamente para sistemas de inteligencia artificial, resulta plenamente aplicable a ellos, pues fija estándares mínimos en materia de seguridad, copias de respaldo, gestión de vulnerabilidades y destrucción de datos. Estos lineamientos, al fijar parámetros técnicos de cumplimiento recomendados, constituyen un marco de referencia para cualquier responsable o encargado de tratamiento que utilice tecnologías basadas en inteligencia artificial.

Por su parte, la res. 4/2019 de la AAIP regula el tratamiento automatizado de datos personales e introduce la obligación (cuando así sea requerido por el titular del dato) de ofrecer explicaciones comprensibles a quienes se vean afectados negativamente por decisiones basadas en procesos automatizados. Este requisito refuerza el principio de transparencia y reviste especial importancia para la Administración pública, que debe garantizar la auditabilidad de sus sistemas de inteligencia artificial y asegurar que las decisiones derivadas de ellos puedan ser cuestionadas por los interesados. En sintonía con estas disposiciones, debe recordarse que la Argentina ha ratificado el Convenio 108+ del Consejo de Europa (ley 27.699, 2022), instrumento internacional que refuerza las garantías frente a la toma de decisiones automatizadas. El texto establece expresamente que ninguna persona puede ser objeto de una decisión fundada únicamente en un tratamiento automatizado de datos sin intervención humana, especialmente cuando dicha decisión produzca efectos jurídicos o impacte de manera significativa sobre el titular. La plena entrada en vigor de este instrumento implicará

para la Administración pública y, por extensión, para los proveedores privados que colaboren en proyectos estatales la obligación de garantizar mecanismos efectivos de revisión humana, eliminando la posibilidad de que los ciudadanos queden sometidos exclusivamente a resultados algorítmicos.

5. La regulación en el derecho comparado

En el derecho comparado es posible identificar unos primeros intentos para orientar la contratación pública de inteligencia artificial.

La Unión Europea adoptó un régimen vinculante con el Reglamento Europeo de Inteligencia Artificial (AI Act, Reglamento (UE) 2024/1689), que entró en vigor el 1º de agosto de 2024 [45]. Es el primer reglamento europeo sobre regulación de inteligencia artificial e impone obligaciones graduadas por riesgo. Asimismo, la Comisión Europea ha publicado cláusulas contractuales modelo destinadas a regular la contratación pública de sistemas de inteligencia artificial [46], lo que responde a la necesidad de armonizar las obligaciones contractuales con el marco regulatorio del Reglamento Europeo de Inteligencia Artificial.

En Canadá, el gobierno dictó la Directiva sobre Toma de Decisiones Automatizadas (*Directive on Automated Decision-Making*) que aplica a todos los sistemas de decisión automatizada desarrollados o adquiridos con posterioridad al 1° de abril de 2020. En el marco de la incorporación de manera creciente por parte del gobierno de la inteligencia artificial para adoptar o apoyar decisiones administrativas el compromiso oficial es utilizar esta tecnología de un modo compatible con los principios esenciales del derecho administrativo, tales como la transparencia, la rendición de cuentas, la legalidad y la equidad procedimental, de manera de garantizar que los procesos de decisión sean justos e imparciales. El objetivo de esta directiva es asegurar que los sistemas de decisión automatizada se utilicen de una manera que reduzca los riesgos para los ciudadanos, los organismos públicos y la sociedad canadiense en general, y que al mismo tiempo permita adoptar decisiones más eficientes, precisas, coherentes e interpretables, en conformidad con el derecho canadiense. Un elemento central de la directiva es la obligación de realizar evaluaciones de impacto algorítmico (*algorithmic impact assessments*) [47].

---

(45) Su aplicabilidad plena (es decir, la exigencia absoluta de cumplimiento para muchas de sus obligaciones) se desplegará de forma escalonada, con fechas límite hasta agosto de 2026 o incluso 2027 para algunas obligaciones específicas (por ejemplo, los sistemas de alto riesgo embebidos en productos regulados).

(46) Comisión Europea, *Updated EU AI Model Contractual Clauses*, Bruselas, 2024. Disponible en: https://public-buyers-community.ec.europa.eu/communities/procurement-ai/resources/updated-eu-ai-model-contractual-clauses.

(47) Gobierno de Canadá, *Directive on Automated Decision-Making*, Ottawa, 2020. Disponible en: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592.

En el Reino Unido, el gobierno publicó en 2020 los Lineamientos para la Contratación de Inteligencia Artificial (*Guidelines for AI Procurement*) [48] en el cual brindan principios y *checklist* para evaluar proveedores, gestionar datos y propiedad intelectual, y fijar cláusulas de desempeño y ética en los contratos.

En Estados Unidos, el estándar técnico de referencia es el *NIST AI Risk Management Framework (AI RMF 1.0)* [49], publicado en 2023, que, si bien tiene carácter voluntario, se ha consolidado como una guía central para identificar, evaluar y mitigar riesgos asociados con sistemas de inteligencia artificial. Este marco es hoy complementado por diversas directrices de la Office of Management and Budget (OMB), entre ellas el Memorando M-24-10 [50], que instruye a las agencias federales a llevar inventarios públicos de los sistemas de inteligencia artificial utilizados, aplicar medidas de gobernanza y supervisión de riesgos, y designar responsables internos de su gestión. A su vez, la OMB ha difundido lineamientos específicos bajo la *Advancing the Responsible Acquisition of Artificial Intelligence in Government* [51], orientados a asegurar que los procesos de compra y contratación de inteligencia artificial por parte del sector público contemplen requisitos de transparencia, rendición de cuentas, seguridad y derechos civiles. De este modo, la combinación del marco técnico del NIST con los memorandos y guías de la OMB configura un régimen administrativo creciente, que condiciona la contratación pública de estas tecnologías y establece estándares mínimos en materia de gestión de riesgos, registros de uso y cláusulas contractuales aplicables a proveedores de inteligencia artificial en el ámbito federal.

Estas experiencias convergen en integrar criterios éticos, técnicos y jurídicos —gobernanza, trazabilidad, evaluación de impacto, pruebas/auditorías, calidad de datos, explicabilidad, seguridad y remedios contractuales— directamente en los pliegos y contratos.

---

(48) Gobierno del Reino Unido, Guidelines for AI Procurement, Londres, 8 de junio de 2020 (última actualización: 29 de enero de 2025). Disponible en: *https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement*.

(49) National Institute of Standards and Technology (NIST), *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, Maryland, 2023. Disponible en: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

(50) Office of Management and Budget (OMB), Memorando M-24-10, "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence", Washington D.C., 2024. Disponible en: https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

(51) Office of Management and Budget (OMB), Memorando M-24-18, "Advancing the Responsible Acquisition of Artificial Intelligence in Government", Washington D.C., 2024. Disponible en: https://www.whitehouse.gov/wp-content/uploads/2024/10/M-24-18-AI-Acquisition-Memorandum.pdf.

6. BUENAS PRÁCTICAS EN LA CONTRATACIÓN PÚBLICA DE INTELIGENCIA ARTIFICIAL

A partir de lo analizado, seguidamente se detallan algunas buenas prácticas en materia de contratación pública de inteligencia artificial, muchas de las cuales, como vimos, ya encuentran reflejo en el Plan Nacional de Inteligencia Artificial y en la *Guía para entidades públicas y privadas en materia de Transparencia y Protección de Datos Personales para una Inteligencia Artificial responsable* como en las *Recomendaciones para una Inteligencia Artificial Fiable*.

La incorporación de la inteligencia artificial debería integrarse de manera expresa en la estrategia de tecnología y datos del Estado. Ello implica no solo adoptar la inteligencia artificial en función de necesidades concretas, sino también contratar estratégicamente para facilitar el despliegue de la inteligencia artificial en la Administración pública [(52)]. En este marco, resulta indispensable evaluar previamente la disponibilidad, calidad y gobernanza de los datos públicos, dado que la mayoría de los sistemas de inteligencia artificial dependen directamente de ellos para su funcionamiento [(53)].

Un primer paso recomendable es la elaboración de guías y pliegos modelo específicos para la contratación de soluciones de inteligencia artificial. Dichos instrumentos deberían contemplar lineamientos técnicos, jurídicos y éticos, a fin de orientar tanto a los organismos contratantes como a los potenciales oferentes. Como vimos en el capítulo anterior, la experiencia internacional ofrece antecedentes valiosos en este sentido.

En lo que respecta a la redacción de pliegos, podría emplearse el mecanismo de Elaboración Participativa de Normas previsto en el dec. 1172/2003. Ello permitiría recabar aportes de distintos actores: sector privado, potenciales proveedores, sociedad civil, academia y expertos especializados. La experiencia comparada demuestra que el diálogo temprano con proveedores y partes interesadas favorece la identificación de las soluciones más adecuadas a los desafíos públicos que se buscan resolver [(54)].

Ahora bien, no todos los sistemas de inteligencia artificial responden al mismo tipo de necesidades. Por ello, es importante seleccionar el procedi-

---

(52) Gobierno del Reino Unido, Guidelines for AI Procurement, Londres, 8 de junio de 2020 (última actualización: 29 de enero de 2025). Disponible en: *https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement*.

(53) Gobierno del Reino Unido, Guidelines for AI Procurement, Londres, 8 de junio de 2020 (última actualización: 29 de enero de 2025). Disponible en: *https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement*.

(54) Gobierno del Reino Unido, Guidelines for AI Procurement, Londres, 8 de junio de 2020 (última actualización: 29 de enero de 2025). Disponible en: *https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement*.

miento de contratación más adecuado para cada caso. En ciertos supuestos, tal como lo demuestra la experiencia de la Unión Europea que comentamos más adelante, puede ser útil adoptar esquemas en doble etapa: una primera de investigación y desarrollo (I+D) en la que los oferentes colaboren en la búsqueda de soluciones técnicas y económicas, y una segunda de contratación formal del servicio. Este enfoque fomenta la competencia y asegura la adaptación de la tecnología a las necesidades públicas.

También los potenciales proveedores podrían encauzar propuestas a través de la modalidad de la iniciativa privada prevista en el decreto 1030. Dicha norma define a la iniciativa privada como aquella cuando una persona humana o jurídica presente una propuesta novedosa o que implique una innovación tecnológica o científica, que sea declarada de interés público por el Estado Nacional a través de la jurisdicción o entidad con competencia en razón de la materia [(55)].

Asimismo, los pliegos deberían contemplar cláusulas que brinden previsibilidad a ambas partes y promuevan una relación de colaboración entre el organismo contratante y el proveedor. En particular, resulta recomendable que se establezcan lineamientos claros sobre: (i) la transparencia mínima necesaria en cuanto al funcionamiento y limitaciones del sistema, sin exigir la revelación de secretos comerciales o información sensible del proveedor; (ii) la adopción de medidas de seguridad y protección de datos razonables y acordes a estándares reconocidos en la industria; (iii) la definición de responsabilidades en función de las áreas efectivamente controladas por cada parte, evitando trasladar al proveedor riesgos regulatorios que dependen del uso que haga el organismo; (iv) mecanismos de interoperabilidad y portabilidad que sean técnicamente factibles, de modo de reducir riesgos de dependencia tecnológica sin imponer costos desmedidos; y (v) disposiciones sobre continuidad del servicio y destino de los datos al finalizar el contrato, con plazos y condiciones que permitan una transición ordenada y viable.

Antes de decidir la contratación de un sistema de inteligencia artificial, la administración debería evaluar: (i) si el sistema actuará de manera autónoma o solo brindará recomendaciones para decisiones humanas; (ii) a quiénes impactarán las decisiones del sistema; (iii) cuáles son las necesidades específicas del sector en que se implementará; (iv) la calidad, disponibilidad y transparencia en la recolección de datos que lo alimentarán [(56)].

---

(55) Dec. 1030/2016, *Reglamento del Régimen de Contrataciones de la Administración Nacional*, Boletín Oficial de la República Argentina, 16 de septiembre de 2016, arts. 25 inc. a), 101 Anexo.

(56) Solano, Mauro; Estévez, Elsa; y Peralta, Amalia, *Guía para el uso de IA en el sector público en Argentina*, CIPPEC, Buenos Aires, 2025, p. 5. Disponible en: https://www.cippec.org/wp-content/uploads/2025/05/Guia-para-el-uso-de-IA-en-el-sector-publico-en-Argentina_2025.pdf.

Otra buena práctica, receptada en las *Recomendaciones para una Inteligencia Artificial Fiable,* consiste en la conformación de equipos interdisciplinarios dentro de la Administración pública [57], integrados por perfiles jurídicos, técnicos y éticos, que acompañen todo el ciclo de vida de los proyectos de inteligencia artificial. A dichos equipos deberían sumarse funcionarios con competencias específicas en las áreas sustantivas en que se apliquen las soluciones (salud, transporte, justicia, entre otras). La inversión en capacitación continua de funcionarios en materia de inteligencia artificial, gestión de datos y derechos digitales resulta esencial para evitar una dependencia exclusiva del conocimiento provisto por los proveedores. Es que uno de los principales desafíos de estas tecnologías es precisamente la limitada capacidad técnica estatal para evaluar, supervisar y auditar de manera independiente las herramientas contratadas.

En igual sentido, se recomienda invertir en infraestructura tecnológica adecuada —conectividad, servicios en la nube y medidas robustas de ciberseguridad—, ya que la capacidad de gobernar la inteligencia artificial requiere también de un soporte técnico sólido que reduzca vulnerabilidades.

Finalmente, una línea emergente de análisis en la contratación pública de soluciones tecnológicas y, en particular, de inteligencia artificial, se vincula con la noción de *sandbox* o espacio controlado de experimentación. Si bien el término ha sido empleado principalmente en el ámbito regulatorio, su lógica resulta trasladable al diseño de procesos contractuales. La idea consiste en introducir fases piloto en las cuales los proveedores despliegan y testean sistemas bajo condiciones acotadas, con métricas de desempeño predefinidas, sin que ello implique aún una contratación masiva o definitiva. Esta modalidad permite al Estado mitigar el riesgo de adquirir tecnologías inadecuadas y, al mismo tiempo, ofrece a los proveedores la oportunidad de demostrar en condiciones reales el valor de sus soluciones. De este modo se genera trazabilidad y auditabilidad del desempeño, lo que contribuye a una mayor previsibilidad del proceso contractual y evita la aparición de sorpresas en etapas posteriores.

La Unión Europea ofrece ejemplos paradigmáticos de este enfoque. A través de mecanismos como el *Pre-Commercial Procurement (PCP)* y las *Innovation Partnerships*, se han establecido marcos contractuales que funcionan, en la práctica, como un *sandbox contractual*.

---

(57) Gobierno del Reino Unido, Guidelines for AI Procurement, Londres, 8 de junio de 2020 (última actualización: 29 de enero de 2025). Disponible en: *https://www.gov.uk/government/publications/guidelines-for-ai-procurement/guidelines-for-ai-procurement*.

El *Pre-Commercial Procurement (PCP)* [58] se concibe como un instrumento destinado a impulsar la innovación en fases previas a la comercialización. A diferencia de los procedimientos de adquisición tradicionales donde se compran bienes o servicios ya disponibles en el mercado, el PCP habilita a los poderes adjudicadores a contratar actividades de I+D en etapas sucesivas que incluyen diseño, desarrollo de prototipos y pruebas de validación a pequeña escala. Ello responde a la tensión entre la necesidad estatal de soluciones innovadoras y la ausencia de opciones maduras en el mercado. En este marco, las administraciones públicas actúan como *lead customers*, generando demanda temprana, reduciendo la incertidumbre de los proveedores y distribuyendo riesgos tecnológicos. Si bien el PCP no se considera contratación pública en sentido estricto bajo la directiva 2014/24/UE, ya que no implica la adquisición inmediata de bienes o servicios operativos, sí debe respetar principios básicos como transparencia, igualdad de trato y competencia leal. Una vez concluida la fase de I+D, y en caso de que los prototipos resulten satisfactorios, la administración puede lanzar un procedimiento posterior de contratación para la compra a gran escala. En consecuencia, en el PCP, los organismos públicos financian fases sucesivas de investigación, prototipado y pruebas, sin adquirir aún un producto final, y seleccionan progresivamente a los proveedores más prometedores.

Por su parte, la *Innovation Partnership,* regulada en la directiva 2014/24/UE, constituye un mecanismo contractual orientado a facilitar la adquisición pública de soluciones tecnológicas aún inexistentes en el mercado. Combina en un mismo procedimiento fases de investigación y desarrollo con la eventual adquisición del producto o servicio, siempre que se verifique su idoneidad. Se diferencia de la contratación tradicional en que no persigue la compra de bienes disponibles, sino el estímulo a la innovación mediante una colaboración estructurada entre sector público y proveedores privados, en la que se definen progresivamente hitos de desempeño, financiamiento y evaluación. De esta forma, el Estado asume un rol activo en el proceso innovador, compartiendo riesgos con los oferentes y asegurando que el resultado final responda a necesidades públicas específicas. La Comisión Europea ha destacado esta figura como una herramienta clave para alinear la contratación pública con objetivos de innovación, sostenibilidad y digitalización [59].

Aunque ambos instrumentos buscan fomentar la incorporación de innovación a través de la contratación pública, sus lógicas difieren. El PCP se

---

(58) European Commission, Pre-Commercial Procurement https://research-and-innovation.ec.europa.eu/strategy/support-policy-making/shaping-eu-research-and-innovation-policy/new-european-innovation-agenda/innovation-procurement/pre-commercial-procurement_en

(59) European Commission, Innovation Procurement: https://single-market-economy.ec.europa.eu/single-market/public-procurement/strategic-procurement/public-procurement-innovation_en

concentra en la fase previa a la comercialización y financia actividades de investigación y prototipado sin compromiso de adquisición posterior, mientras que la *Innovation Partnership* combina la I+D con la eventual compra, ofreciendo a los proveedores una expectativa contractual más definida. En términos prácticos, el PCP resulta más idóneo cuando se requiere explorar alternativas tecnológicas con alto grado de incertidumbre, en tanto que la *Innovation Partnership* asegura una mayor continuidad contractual cuando la innovación perseguida responde a una necesidad pública claramente identificada.

En definitiva, tanto el PCP como las *Innovation Partnerships* cumplen una función análoga a la de los *sandboxes regulatorios*: proporcionan entornos controlados para ensayar tecnologías emergentes, mitigar riesgos y adaptar los marcos jurídicos y técnicos antes de una implementación a gran escala. En el caso particular de la contratación de sistemas de inteligencia artificial, su utilización pareciera ser especialmente adecuada, ya que permite abordar de manera progresiva cuestiones críticas como los sesgos, la protección de datos, la transparencia y la rendición de cuentas. La adopción de este enfoque contribuiría no solo a una contratación pública más eficiente y transparente, sino también a alinear la política estatal con objetivos estratégicos de innovación y desarrollo tecnológico en el país.

En su conjunto, estas recomendaciones configuran un marco orientado a que la contratación pública de inteligencia artificial en la Argentina logre un equilibrio entre innovación tecnológica, seguridad jurídica, respeto por los derechos fundamentales y sostenibilidad institucional.

7. CONCLUSIÓN

La incorporación de inteligencia artificial por parte del Estado es, a la vez, una oportunidad de modernización y un desafío jurídico de primera magnitud. En ausencia de un régimen específico, la contratación de estas soluciones debe anclarse en el marco general de contrataciones y en los estándares éticos y técnicos disponibles, evitando trasplantar sin más lógicas propias del software tradicional a tecnologías que aprenden, se actualizan y pueden afectar derechos fundamentales. El objetivo no es frenar la innovación, sino gobernarla con reglas claras y previsibles.

De las guías y recomendaciones vigentes se desprende una hoja de ruta práctica para el Estado argentino, que, a partir de principios rectores, establece parámetros para que tanto los proveedores como el propio Estado tengan en cuenta al contratar sistemas que incorporen inteligencia artificial.

La experiencia comparada muestra un menú de instrumentos útiles que pueden inspirar un *sandbox* contractual adaptado al derecho local: fases piloto con métricas verificables, trazabilidad y revisión humana, antes

de la adopción a escala. Este enfoque reduce asimetrías técnicas, mitiga sesgos y riesgos y, a la vez, evita compromisos irreversibles para el Estado.

Un enfoque preventivo, proporcionado y basado en evidencia permitirá al Estado contratar soluciones de inteligencia artificial de manera segura, ética y eficiente. En definitiva, la cuestión no es si el Estado debe incorporar inteligencia artificial, sino cómo hacerlo: con legitimidad, control y resultados medibles que aumenten la confianza ciudadana y, al mismo tiempo, otorguen seguridad jurídica a los proveedores, garantizando que la innovación se traduzca en un verdadero fortalecimiento del servicio público.

Usted apriete el botón, nosotros hacemos el resto. ¿Nacen derechos de autor sobre las obras creadas mediante el uso de inteligencia artificial?

Por Ezequiel Valicenti [(*)]

Resumen: El artículo analiza la posibilidad de que las obras creadas mediante inteligencia artificial generativa (IAGen) puedan ser consideradas "obras protegidas" bajo el derecho de autor. Partiendo de la dogmática del derecho de autor, se revisan los conceptos de obra, obra protegida y originalidad, destacando que la protección requiere autoría humana y originalidad de la expresión. Se examinan los criterios de originalidad: novedad, individualidad y esfuerzo creativo, y se concluye que la IAGen, al producir obras de manera autónoma a partir de datos preexistentes, carece de autoría humana y de originalidad suficiente para que surjan derechos de autor. Además, se discuten alternativas como proteger los *prompts* de la IA o crear sistemas mixtos de protección. Finalmente, se argumenta que la negación de derechos sobre obras de IA protege a los autores humanos frente a la industrialización y masificación de contenidos culturales generados automáticamente.

Palabras clave: Derecho de autor; obra protegida; originalidad; inteligencia artificial generativa; autoría humana; propiedad intelectual; producción cultural automatizada.

Summary: The article analyzes whether works created using generative artificial intelligence (AI) can be considered "protected works" under copyright law. Based on copyright dogmatics, it reviews the concepts of work, protected work, and originality, emphasizing that protection requires human authorship and originality of expression. The criteria of originality—novelty, individuality, and creative effort—are examined, concluding that generative AI, producing works autonomously from pre-existing data, lacks human authorship and sufficient originality to generate copyright. Alterna-

(*) Facultad de Derecho – UniCen. Contacto: ezequielvalicenti@gmail.com Abogado y especialista en Docencia Universitaria. Decano y profesor adjunto de Derecho Privado en la Facultad de Derecho de la Universidad Nacional del Centro de la Provincia de Buenos Aires. Integrante del Grupo de Estudios en Derechos Culturales de la Argentina y del Centro de Estudios en Arte y Letras (UNICEN).

tives such as protecting AI prompts or creating mixed protection systems are discussed. Ultimately, denying copyright to AI works safeguards human authors against the industrialized mass production of automatically generated cultural content.

Keywords: Copyright; protected work; originality; generative artificial intelligence; human authorship; intellectual property; automated cultural production.

1. Introducción

A fines del siglo XIX, George Eastman revolucionó el acceso a la fotografía con la cámara Kodak. Eastman no solo inventó el rollo de película que permitió la simplificación de la fotografía y su acceso masivo, también tuvo el ingenio de crear el célebre eslogan con el que se publicitó la empresa: "Usted apriete el botón, nosotros hacemos el resto" (Fig. 1). Ciertamente, la expresión sintetizaba un cambio radical: la compleja técnica fotográfica dejaba de estar en manos de especialistas y pasaba a ser una experiencia al alcance de cualquier persona.

Fig. 1: Anuncio de la primera cámara Kodak (1988)

Sin embargo, el ideario que emergió de esa frase tuvo un efecto cultural inesperado. Durante décadas, la fotografía fue vista como un arte menor, en parte por la percepción de que el esfuerzo creativo quedaba reducido a un gesto mecánico. La cámara automatizaba procesos que antes requerían destreza y conocimiento técnico, y la comunidad artística tendió a minusvalorar la originalidad del fotógrafo. Esa mirada influyó decididamente en el derecho: mientras las obras artísticas (como la pintura, la música o la literatura) contaban con protección durante toda la vida del autor e inclusive durante algunos años después de su muerte (hoy, 70 años *postmortem)*, la

fotografía recibió un reconocimiento autoral más restringido. De hecho, en nuestra legislación el derecho de propiedad intelectual sobre la obra fotográfica se extiende durante 20 años desde la primera publicación (art. 34, Ley de Propiedad Intelectual, en adelante LPI).

El lema de Eastman, que celebraba la sencillez técnica, marcó el inicio de un debate, que se repite cada vez que la tecnología atraviesa los procesos creativos: ¿cuánto hay de arte y cuánto de máquina en la obra? La misma pregunta reaparece ahora frente a la difusión de la inteligencia artificial.

Fig. 2: Imagen creada por Chat GPT, con la instrucción de hacerla a partir de la imagen original de Fig. 1, incluir su eslogan publicitario y adaptarla para promocionar la IAGenerativa (2025)

2. Inteligencia artificial y creatividad. Una agenda de problemas

La dogmática del derecho de autor se estructura sobre el concepto de obra protegida (Della Costa, 1971). Por lo tanto, el uso y desarrollo de la inteligencia artificial plantea dos interrogantes esenciales que se construyen en torno a ese concepto: a) ¿es posible que una obra creada con inteligencia artificial sea una obra protegida?; b) ¿el proceso de desarrollo y entrenamiento de modelos de inteligencia artificial requiere algún tipo de autorización y compensación a los titulares de derechos sobre una obra protegida? De alguna manera puede decirse que se trata de un primer problema asociado al *output* de la IA —es decir, la pregunta sobre el resultado emergente, la obra producida— y de un segundo interrogante asociado al *imput* —es decir, la

forma en la que se han construido las bases de datos sobre las que funciona una determinada IA—. Aparece, además, una dimensión colectiva, que involucra otros interrogantes que trascienden la meramente individual propia de la propiedad intelectual, y que tienen que ver con la diversidad cultural —¿con qué tipo de obras fueron construidas ("entrenadas") las IA generativas hoy masivamente difundidas?— y con la protección de la creatividad y la libertad de expresión artística [1].

En esta oportunidad nos ocuparemos del primero de los problemas, esto es, el vinculado a la posibilidad de reconocer derechos de autor sobre las obras creadas por (o con) IA. Anticipamos que intentaremos demostrar que la respuesta es negativa. Para ello, en primer lugar repasaremos el concepto de "obra protegida" y los requisitos que deben reunirse para que puedan reivindicarse derechos de autor, y luego aportaremos los argumentos por los que consideramos que las obras creadas por IA no alcanzan dicho estatus.

3. La narrativa de la obra protegida y la originalidad en el derecho de autor

La arquitectura al microsistema de derecho de autor contenido en la Ley de Propiedad Intelectual inicia con la definición de obra protegida (art. 1°). Se trata de un concepto normativo pues determina concretos efectos jurídicos y por lo tanto no se reduce a un mero valor dogmático.

Los autores coinciden generalmente en señalar dos características necesarias para la identificación de una obra protegida: a) la existencia de una forma sensible o representativa [2]; b) la originalidad de esa forma representativa o expresión (Colombert, 1997:9; Goldstein, 1995:52; Lipszyc, 1993:37). Se añaden además, ciertas características negativas, que constituyen cualidades no necesarias, es decir, que son ajenas a la calificación de obra protegida. Se dice que no importan a los fines de la protección ni el mérito, ni el destino, ni de las formas utilizadas (Antequera Parilli, 1998:129), y que tampoco incide la "moralidad" de la obra, es decir el juicio moral que suscite en el observador (Villalba y Lipszyc, 2009:37).

Podemos decir entonces para que exista obra basta con que se perciba una *forma*, pero para que se trate de una obra protegida, esa forma debe ser además *original* (Vivant y Bruguiere, 2012:115). La obra protegida es, entonces, una forma perceptible original.

---

(1) La esencia de este enfoque de derechos culturales se encuentra desarrollado en el Informe de la Relatora Especial sobre los derechos culturales, titulado "Inteligencia artificial y creatividad", del 30/06/2025, registrado como A/80/278.

(2) Según decía Satanowsky, una "creación integral humanamente perceptible" (Satanowsky, 1954: 153).

3.1. La obra

La existencia de una obra se define por varios elementos. Debe tratarse de un *opus* producto de la intervención de la persona, el resultado de la actividad humana (Ruipérez Azcárate, 2012:30). Ello excluye las obras creadas por la naturaleza, pero no las que son creadas por el humano con asistencia de ciertas tecnologías, como las computadoras (Ginsburg, 2003:1074; Bercovitz Rodríguez-Cano, 2015:52). Esta posibilidad es la que se encuentra esencialmente comprometida en el caso de la inteligencia artificial generativa (en adelante, IAGen).

En términos jurídicos la creación es un simple acto lícito, esto es, una acción voluntaria que produce efectos jurídicos, en este caso, el nacimiento de los derechos subjetivos de autor (art. 258, Cód. Civ. y Com.). Si bien constituye una especie de hecho jurídico, no es un acto jurídico (arts. 257 y 260, Cód. Civ. y Com.) y por lo tanto no requiere capacidad ni conciencia (Vivant y Gruguiere, 2012:123; Raffo, 2011:55). Por lo demás, la creación, en cuanto acto lícito y manifestación del principio de autodeterminación personal, se encuentra tutelada por el principio de libertad de expresión [(3)].

Ahora bien, la creación debe traducirse en una *forma perceptible*. La forma es el producto de la acción e intención del creador sobre la materia. Es el substrato que contiene en sí el sentido que se hace evidente al observador en el acto de percepción (Raffo, 2011:31). El concepto jurídico de *forma* se completa con un *criterio de exclusión* que ha sido tradicional en el derecho de autor y que expresa que únicamente pueden concederse derechos individuales sobre expresiones formales, más no sobre las ideas [(4)]. En el ordenamiento jurídico argentino el criterio de exclusión ha sido positivizado en el segundo párrafo del art. 1° de la LPI que indica que "la protección del derecho de autor abarcará la expresión de ideas, procedimientos, métodos de operación y conceptos matemáticos pero no esas ideas, procedimientos, métodos y conceptos en sí" [(5)].

En otras palabras, únicamente es obra la *expresión* de un pensamiento o de un mensaje, la organización formal que el autor escoge en su acto de creación para decir algo acerca del mundo que lo rodea. El pensamiento

---

(3) Ampliamos en Valicenti, 2015.

(4) La doctrina ha desarrollado suficientemente las razones de tal exclusión (Conf. Cifuentes, 2002:2; Lipszyc, 1993:62).

(5) La fórmula fue incluida por la Ley 25.036 del año 1998, y tiene como antecedentes el art. 2° del Tratado OMPI de Derechos de Autor, y el art. 9.2 de los Acuerdos ADPIC. Antes de la reforma la exclusión había sido recibida en la jurisprudencia: "las ideas son libres, y por no constituir una obra intelectual, ni una creación completa, su autor no goza de protección alguna" (Cam. Nac. Civ., Sala F, "Saúl Romay, Argentino A. c. Tirigall, Horacio y otros", del 25/04/72, en *El Derecho,* 43, p. 619); "la simple idea, como tal, no constituye aún la obra objeto de resguardo legal, toda vez que le falta la realización, la forma concreta, la estructura" (Cam. Nac. Civ., Sala E, "Guberman, Pedro A. c. Discos CBS", del 21/05/81, en *La Ley,* 1981-D, p. 378).

tiene como antecedente la idea, pero no es la idea, sino una concreción o formalización determinada (Ruipérez Azcárate, 2012:18). En otras palabras, la obra es el ropaje que viste la idea; por lo tanto, de una misma idea pueden surgir diferentes obras.

Ahora bien, la necesidad de una forma perceptible no exige que exista un soporte material en la cual se encuentre expresada, es decir, no se exige su "fijación" (Villalba y Lipszyc, 2009:33). Es posible que la obra se exteriorice, se vuelva perceptible, sin necesidad de formalizarse en un soporte material, sea duradero o efímero. La noción de obra no se identifica necesariamente con la de soporte material, y por ello ni los medios empleados para la expresión ni la fijación de la expresión no son en sí mismos obras (Satanowsky, 1954:160).

En síntesis, obra es la expresión o formalización perceptible de un pensamiento, originada por la voluntad del hombre. Luego, para que sea "protegida" debe ser *original*.

3.2. La obra protegida

Las obras protegidas se encuentran mencionadas en el art. 1° de la LPI, a partir de una enumeración enunciativa que concluye con una cláusula general que indica que queda comprendida "toda producción científica, literaria, artística o didáctica sea cual fuere el procedimiento de reproducción". Al incluir "toda producción", una primera lectura podría sugerirnos que se encuentran incluidas, sin más, todas las creaciones consideras obras. Inclusive, ello aparecería corroborado por el segundo párrafo que refiere a "la expresión" [(6)], lo que no sería sino otra referencia a sola formalización perceptible. Sin embargo, como dijimos, la doctrina requiere que esa obra sea *original,* aun cuando la exigencia no se encuentre expresa en los enunciados normativos de la ley nacional [(7)].

---

(6) "La protección del derecho de autor abarcará *la expresión* de ideas, procedimientos, métodos de operación y conceptos matemáticos pero no esas ideas, procedimientos, métodos y conceptos en sí" (art. 1°, 2° párr., LPI). El destacado nos pertenece.

(7) En el ámbito latinoamericano, una serie de leyes nacionales incluyen expresamente la exigencia de originalidad al definir el concepto de obra protegida. Así: Costa Rica, art. 1° de la Ley sobre Derechos de Autor y Derechos Conexos: "Las producciones intelectuales originales confieren a sus autores los derechos referidos en esta Ley"; Cuba, art. 2° de la Ley de Derecho de Autor se refiere a las "obras científicas, artísticas, literarias y educaciones de carácter original", y la Sección I, del Capítulo II se denomina "de las obras originales"; Ecuador, art. 7° de la Ley de Propiedad Intelectual: "Obra: Toda creación intelectual original, susceptible de ser divulgada o reproducida en cualquier forma, conocida o por conocerse"; Guatemala, art. 15 de la Ley de Derechos de Autor y Derechos Conexos: "Se consideran obras todas las producciones en el campo literario, científico y artístico, cualesquiera que sea el modo o forma de expresión, siempre que constituyan una creación intelectual original"; Honduras, art. 2° de la Ley del Derecho de Autor y de los Derechos Conexos: "Son obras literarias o artísticas, todas las creaciones originales con independencia de su género y cualquiera que sea el modo o forma"; México, art. 3° de la Ley Federal del Derecho de Autor: "Las obras protegidas por esta Ley son aquellas de creación original susceptibles de ser divulgadas o reproducidas en cualquier forma o medio"; Panamá, art. 2.14 de la Ley sobre el Derecho de Autor y Derechos Conexos: "Obra: Creación

La noción de originalidad tiene ascendencia en la Ilustración, en la filosofía kantiana y en la concepción romántica del artista construida durante el siglo XIX, según la cual el autor es un genio, que crea "inspirado", de manera solitaria y *ab initio,* sin influencias ni referencias previas (Ginsburg, 2003; Le Henaff, 2006).

En la ciencia jurídica, es clásica la conceptualización de originalidad esbozada por Henri Desbois, para quien "est originale toute création qui n'est pas la simple reproduction d'une oeuvre existante et qui exprime le goût, l'intelligence et le savoir-faire de son auteur, en d'autres termes, sa personnalité dans la composition et l'expression" (1978:23). El hecho de que el autor francés defina la noción a partir de la exclusión, indicando "lo que no es", denota las dificultades que existen para arribar a una definición satisfactoria de originalidad en el ámbito jurídico.

A nuestro modo de ver la originalidad es un concepto vacío; su contenido, sus notas positivas, dependen del criterio de valoración que se utilice [8]. Es decir, es una noción propia del microsistema de derecho de autor, y su definición varía según se recurra a un *criterio objetivo* –que privilegia la *novedad* como nota característica– o a cualquiera de los dos *criterios subjetivos* –por un lado, el que asimila originalidad con la *individualidad* de la creación; por otro, el que la define por la presencia de *esfuerzo creativo* suficiente. Repasaremos a continuación los tres criterios calificativos de la originalidad que identificamos [9].

3.2.1. Criterio objetivo: la originalidad como novedad

La originalidad puede ser entendida como sinónimo de novedad. Decimos que es un criterio objetivo pues la verificación se realiza con el solo

---

intelectual original, de naturaleza artística, científica o literaria, susceptible de ser divulgada o reproducida en cualquier forma"; Paraguay, art. 2.16 de la Ley de Derechos de Autor y Derechos Conexos "Obra: toda creación intelectual original, en el ámbito literario o artístico, susceptible de ser divulgada o reproducida en cualquier forma, conocida o por conocer"; Perú, art. 2.17 de la Ley sobre el Derecho De Aturo "Obra: Toda creación intelectual personal y original, susceptible de ser divulgada o reproducida en cualquier forma, conocida o por conocerse"; República Dominicana, art. 16.13 de la Ley sobre Derecho de Autor: "Obra: Toda creación intelectual original, de carácter artístico, científico o literario, susceptible de ser divulgada o reproducida en cualquier forma, conocida o por conocerse".

En otros casos, la originalidad es requisito exigido para la protección de obras derivadas, aunque no se mencione tal requisito en el concepto general de obra protegida (ej. Colombia, art. 5° del Régimen General de Derecho de Autor).

En el derecho europeo, España, art. 10 de la Ley de Propiedad Intelecual: "Son objeto de propiedad intelectual todas las creaciones originales literarias, artísticas o científicas expresadas por cualquier medio o soporte, tangible o intangible (…)"; Francia, no menciona el requisito de originalidad en la definición de obra (arts. L 111-1 y 111-2, Código de Propiedad Intelectual), sino que tan solo lo menciona al regular la protección del título (art. L 112-4, Código cit.).

(8) En la doctrina francesa, Vivant y Bruguiere consideran que la originalidad es una "noción marco", con contenido variable (2012: 234).

(9) Otros autores han propuesto una construcción diferente de los criterios (ver, Erdozain López, José C., "El concepto de originalidad en el Derecho de autor", en *pe. i. revista de propiedad intelectual,* n° 3 (sept.-dic. 1999), pp. 55-94)

análisis de la obra, sin referencia a cualidades o conductas del sujeto que la ha creado. La obra es novedosa si no constituye una repetición de otra que ya existe. La novedad es la cualidad de lo nuevo, y lo nuevo es aquello que se percibe o se experimenta por primera vez, una forma distinta o diferente a lo que antes había.

En la doctrina nacional, Satanowsky indicaba para que la obra sea protegida "basta que no sea copia de otra, que importe un esfuerzo intelectual de características propias, que haya sido producida con invención y esfuerzo particular del autor, que sus elementos estén combinados de una manera nueva" (Satanowsky, 1954:153). La jurisprudencia nacional más antigua ha seguido este criterio. En un viejo precedente, la CS ha dicho que la obra debe "expresión personal, original y novedosa de la inteligencia" [(10)].

La confusión existente entre originalidad y novedad ha sido superada en la evolución de los estudios sobre la materia (Iturralde, 2009). En la actualidad, la doctrina recuerda reiteradamente que, a diferencia del régimen de propiedad industrial, la novedad no es un requisito para la protección (Emery, 1999:45; Erdozain López, 1999:60).

3.2.2. Criterio subjetivo (I): la originalidad como individualidad

La enorme mayoría de la doctrina nacional identifica a la originalidad con la *individualidad* de la expresión. Es decir, para ser original la obra debe llevar la impronta de la personalidad del autor (Villalba y Lipszyc, 2009:38; Cifuentes, 2002:5; Vibes, 2009:429; Iturralde, 2009; Graffoglio, 2006). La obra puede no ser novedosa –pues puede retomar un tema o idea ya desarrollada previamente–, pero debe contener el reflejo de la individualidad del autor, su traza personal.

Para expresar este criterio los autores y la jurisprudencia utilizan diversas fórmulas. Así la originalidad se verifica cuando la obra "refleja la individualidad del autor, su personalidad" [(11)], cuando contiene "un mínimo de expresión personal del autor" [(12)], cuando existe "aporte personal del espíritu de carácter intelectual" [(13)], o cuando se trata de "la expresión —o forma representativa— creativa e individualizada de la obra" [(14)]. En todos los casos se quiere expresar la misma idea: la originalidad de una obra radica en el grado de individualidad o rastro personal que puede darle el autor. Este es el criterio seguido además en el derecho europeo continental (Osorio Umaña, 2025).

---

(10) CS, "Nóbile, Ambrosio H c/ Provincia de Santa Fe", del 18/09/68, en *La Ley* 133, p. 807.

(11) Cam. Nac. Civ., Sala E, "Cosmopolita, SRL c. Editorial Caymi, Soc. en Com. por Accs. y otros", del 28/07/83, en La Ley 1984-B, p. 406 (voto Dr. Padilla)

(12) Cam. Crim. y Correc., Sala V, del 16/04//1971, en *El Derecho* 37, p.586.

(13) Cam. Nac. Civ., Sala D, "A.G.I.S.A. c. Kaiser, Alberto J.", del 24/02/97, en *La Ley Online*, cita: AR/JUR/4108/1997.

(14) Corte Suprema de Justicia Tucumán, "Martínez, Enrique Julio c. Provincia de Tucumán", del 02/03/10, en *La Ley Online*, cita: AR/JUR/9909/2010

3.2.3. Criterio subjetivo (II): la originalidad como esfuerzo creativo

Es posible identificar un criterio subjetivo alternativo, según el cual la obra presenta originalidad cuando es el resultado de un mínimo *esfuerzo creativo*. Por lo tanto, no será obra original —y en consecuencia tampoco obra protegida— aquella que es producto de una actividad mecánica o rutinaria (Ruipérez Azcárate, 2012:36). En el sistema anglosajón del copyright la jurisprudencia ha recurrido a este criterio para completar la noción de originalidad. La jurisprudencia norteamericana utiliza la noción de *spark of creativity* (destello de creatividad), mientras que en la británica se recurre al *sweats labor* (sudor del trabajo) (Fenzel, 2007:557).

La doctrina norteamericana ha quedo consolidada con la conocida sentencia de la Corte Suprema en el caso "Feist Publications Inc. v. Rural Telephone Services Co." del año 1991, en la que se sostuvo que la originalidad de las obras no depende del "sudor de la frente" (*sweat of the brow*) del autor sino de que sea una creación independiente para el autor (lo que significa que no sea una copia de otros trabajos) y que posea un mínimo grado de creatividad (Nuñez, 2025) [(15)].

Este criterio no es seguido en el sistema de derecho continental [(16)]. Pese que la jurisprudencia nacional suele utilizar conceptos tales como "creatividad" o "esfuerzo" no lo hace con el sentido estricto dado en los países del *copyright*, sino como cualidades que contribuyen a justificar la individualidad del autor plasmada en la obra.

3.2.4. ¿Unidad o pluralidad de criterios? La acreditación de la originalidad

Repasados los criterios, el interrogante acerca de qué es la originalidad se aclara pero no se resuelve completamente. Al no existir norma que indique el criterio a utilizar, no es errado concluir que los criterios son igualmente válidos, puesto que todos pueden adscribirse a los enunciados normativos vigentes, concretamente, a la definición provista por el art. 1° de la LPI. En última instancia, los criterios proveen razones para abastecer la *carga de argumentación* que conlleva demostrar la existencia de una obra protegida en el caso concreto.

La fuerza y el seguimiento de los criterios varía según el sistema normativo. En el caso del sistema de Derecho de autor continental, entre los que

---

(15) "(…) Original, as the term is used in copyright, means only that the work was independently created by the author (as opposed to copied from other works), and that it possesses at least some minimal degree of creativity" (US Supreme Court, "Feist Plub'ns, Inc. v. Rural Tel. Serv. Co.", 499 U.S. 340 (1991). Ver, Ginsburg, 2003:1078.

(16) "El derecho de autor no se conmueve ante el esfuerzo o la habilidad (argumentos del tipo 'me ha costado mucho' o 'era muy dificil' tiene poco recorrido en este ámbito" (Casas Vallés, 2008:127).

se inscribe nuestro régimen, se ha consolidado una tradición doctrinaria y jurisprudencial que identifica originalidad con la individualidad expresada en la obra. El Glosario de Derecho de autor elaborado por la Organización Mundial de Propiedad Intelectual (OMPI) de 1980 también parece seguir este criterio desde que considera que la obra posee originalidad cuando "es una creación propia del autor, y no copiada de otra obra en su totalidad o en una parte esencial" (voz originalidad).

En esta línea, se ha dicho que determinar la existencia de la originalidad es una cuestión fáctica (Villalba y Lipszyc, 2009:40; Emery, 1999:45), a lo que algunos autores agregan que debe presumirse que toda obra es original, por lo que quien niega la cualidad debe probarlo (Satanowsky, 1954:170). No obstante, los aspectos probatorios dependerán del criterio alegado para afirmar la originalidad de una obra: si el conflicto se analiza desde el criterio objetivo (originalidad como novedad), habrá que probar que lo que creado es distinto de la o las obras existentes; si en cambio se analiza desde el criterio subjetivo mayormente seguido (originalidad como individualidad), la prueba consistirá en demostrar que no se ha copiado (Bercovitz Rodriguez Cano, 2015:64).

En la práctica, será necesario distinguir sobre qué aspectos de la obra debe apreciarse la originalidad. Algunos autores han explicado que toda obra tiene tres momentos: la idea, la composición de la idea y su expresión. Dado que la idea no es pasible de protección, la originalidad reposa en la composición (contenido) o en su expresión (forma) —o en ambas conjuntamente— (Villaba y Lipszyc, 2009:44). Sin embargo, no debe ser probada la ejecución personal de la obra por parte del artista. Ello así pues la intervención personal del autor no es requisito imprescindible para ser el autor de la obra [17]. Es decir, las destrezas en el trabajo y la ejecución no son necesariamente importantes para decidir los casos en los regímenes de derecho continental [18].

4. Las obras creadas mediante el uso de inteligencia artificial

Partiendo de las referencias dogmáticas repasadas, veamos si efectivamente las obras creadas por inteligencia artificial pueden ser tenidas como obras protegidas y, en consecuencia, objeto de derechos individuales de au-

(17) Se ha dicho que el autor es "mente antes que músculo" (Ginsburg, 2003:1074).

(18) Cam. Nac. Criminal y Correccional, Sala II, "Oliva de Macadam, Ana", del 11/08/92, en *La Ley* 1993-D, p.356 con nota de Villalba, Carlos A., "En torno al tema de quién es el autor de una obra". La actuación de los asistentes puede influir también sobre la "autenticidad" de la obra (ver Cam. Nac. Civ., Sala E, "De Koenigsberg, Nicolás c. de Fourvel Rigolleau, Ivonne", del 26/02/81, en *La Ley* 1981-C, p. 162).

tor. El examen puede desarrollarse en dos niveles: el de la autoría y el de la originalidad de la obra.

4.1. La autoría de la obra creada por inteligencia artificial

Como vimos, el Derecho de autor exige que la autoría sea de una persona humana. Ello así pues se trata de un simple acto lícito (art. 258, Cód. Civ. y Com.) y por lo tanto debe ser obrado por una persona.

Entre nosotros, se ha propuesto un "espectro de autorías" en las que se distinguen las obras asistidas por inteligencia artificial y las obras generadas por inteligencia artificial. Mientras en las primeras, la IA operaría como equivalente a otros instrumentos artísticos y técnicos, como una herramienta, en las segundas la IA actuaría con "un grado de autonomía que pone a los resultados fuera de una relación subordinada a decisiones estilísticas de operadores o creadores de modelos" (Mantega, 2022:346). En otras palabras, se trataría de distinguir las obras generada *con* IA y las producidas *por* la IA.

A nuestro modo de ver, el hecho de que la persona dé instrucciones para que la IAGen cree una obra determinada (*prompt*), no implica que exista una autoría humana. Es que, aun cuando parta de instrucciones precisas y concretas, la obra se produce de manera autónoma por la IA, y el autor del *prompt* carece de dominio y gobierno sobre el proceso de creación. Si el proceso creativo se entiende como ejercicio de libertad artística, el resultado de un mecanismo automatizado difícilmente pueda equipararse a un acto creativo humano, basado en elecciones del artista. Más aún, si tenemos en cuenta que las IAGen comúnmente difundidas adolecen de transparencia y, en cambio, suelen tener grandes niveles de opacidad —es decir, el usuario no conoce la forma en la que se llega a un determinado resultado—. Eso lleva a que sea prácticamente imposible determinar con precisión qué decisiones corresponden a la máquina y cuáles al ser humano que da las instrucciones.

Lo dicho no significa que los artistas no utilicen (ahora y en un futuro) la IA como una herramienta más. Pero en esos casos, la intervención sobre el producto generado debe ser significativa, de modo que pueda decirse que las decisiones creativas han sido dadas por el autor y, en consecuencia, el resultado final (la obra) sea un *opus* de su entero control y disposición.

4.2. La (inexistencia de) originalidad en la obra

Más allá del problema de la autoría, es posible realizar un análisis sobre el *outoput* (la obra), lo que implica determinar si existe o no originalidad suficiente, de acuerdo a algunos de los criterios que, como vimos, han sido construidos en la dogmática del Derecho de autor.

En términos técnicos, la creación de una obra a partir de instrucciones dadas a una IAGen no es sino una reconstrucción, por predicción, a partir de una "nube" de obras que han sido previamente "datificadas" (convertidas en datos). En otras palabras, la IAGen simula creatividad combinando estadísticamente lo aprendido, a partir de la identificación de patrones, pero sin tener conciencia ni intencionalidad estética. La tecnología es incapaz de crear algo nuevo, diferente a la mera combinación de los datos del *dataset* sobre el que se monta.

En este sentido, se ha puesto seriamente en dudas que los modelos de IA tengan un ápice de creatividad. Se sostiene que las obras creadas por inteligencia artificial no son más que obras derivadas de miles de millones de otras obras que han sido utilizadas para el entrenamiento, y que esas obras usadas asimismo como datos para continuar alimentando las enormes bases de datos, acabarán produciendo un bucle de retroalimentación que terminará degradando los sistemas (Steyerl, 2025:58 y 56).

En el ámbito norteamericano —en el que, recordemos, la originalidad se evalúa según el nivel mínimo de creatividad— la tendencia ha sido la de negar reconocimiento de *copyright* sobre obras creadas por IA. Así, la Oficina de Derechos de Autor de EEUU (US Copyright Office), rechazó las solicitudes de registro de imágenes creadas automáticamente con distintas IA, remarcando la necesidad de que exista intervención humana en el proceso de creación [(19)].

Con una mirada alternativa, algunos autores proponen buscar la creatividad no en la obra (el *output*), sino en las instrucciones o *prompts* que conducen a esas obras (Nuñez, 2025). Incluso se sostiene que "los prompts son realmente obras, y la IA no hace más que dar un cierto orden o asistencia técnica" y que, en consecuencia, esas mismas instrucciones originales pueden generar obras originales (Canitrot, 2024:3).

¿Es posible que los *prompts* generen obras que transmitan la personalidad del autor, es decir, que sean originales en términos jurídicos? Coincidimos, en este punto, con quienes sostienen que no es posible afirmar que la IA tenga una personalidad que queda impresa en el producto final (Osorio Umaña, 2025:8).

Ante este escenario en el que se impone la tesitura que niega derechos de autor, algunos proponen la posibilidad de pensar en un sistema mixto,

---

(19) El primer caso se resolvió el 14 de febrero de 2022, en el que se rechazó una solicitud de una imagen creada con la IA "Creativity Machine" (United States Copyright Office, Second Request for Reconsideration for Refusal to Register A Recent Entrance to Paradise – Correspondence ID 1-3ZPC6C3; SR # 1-7100387071, disponible en https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf, visitado el 30/09/2025.

Posteriormente, el 16 de marzo de 2023, la misma Oficina de Derecho de Autor dictó una Guía para la registración de obras que contienen material generado con IA. Disponible en: https://www.copyright.gov/ai/ai_policy_guidance.pdf, visitado el 30/09/2025.

en el cual las obras creadas por personas humanas tengan una protección fuerte, y las generadas por IA tengan una más débil (Edwards, 2025).

5. Conclusión: ¿a quién importa la respuesta?

En definitiva, aún con distintas perspectivas de análisis, se impone la misma conclusión: las obras creadas con IA carecen de autoría humana, no puede decirse que ellas expresen la personalidad del autor —quien no puede dominar el proceso de creación ni tener elecciones creativas significativas— y, por ende, no cuentan con originalidad para ser consideradas obras protegidas. En pocas palabras, al menos en el estado actual de la técnica, no nacen derechos de autor sobre las obras creadas por IA.

Esta respuesta no debería preocupar a los/as autores/as. Por el contrario, no puede perderse de vista que los verdaderos interesados en la respuesta afirmativa son los grandes conglomerados de industrias culturales y, particularmente, las plataformas de "contenidos culturales", quienes demandan un volumen de obras que deber ser constantemente renovado, las que podrían producir de forma seriada con herramientas de IA, compitiendo así con las obras creadas por personas humanas y precarizando aún más la situación de los creadores.

Bibliografía

Antequera Parilli, Ricardo (1998), *Derecho de autor*, Servicio Autónomo de la Propiedad Intelectual-Dirección Nacional del Derecho de Autor, Caracas.

Bercovitz Rodríguez-Cano, Rodrigo (2015), *Manual de Propiedad Intelectual*, 6° ed., Tirant Lo Blanch, Valencia.

Canitrot, Carolina A. (2024), "Los desafíos que nos trae la Inteligencia Artificial en el contexto del Derecho de Autor", en *Propiedad Intelectual 4.0*, Año III, Nº 1, La Ley, Bs. As., pp. 1-3.

Casas Vallés, Ramón (2008), "La condición de autor: los casos Barceló y Boadella", en *pe. i. revista de propiedad intelectual*, Nº 28 (ene-abr. 2008), pp. 127-141.

Cifuentes, Santos, (2002), *Obras no protegidas en los derecho de autor*, Academia Nacional de Derecho y Ciencias Sociales de Bs. As. – La Ley, Bs. As.

Colombert, Claude (1997), *Grandes principios del derecho de autor y los derechos conexos en el mundo. Estudio de derecho comparado*, 3° ed., Ediciones Unesco-Cindoc, Madrid.

Della Costa, Hector (1971), *El derecho de autor y su novedad*, Cathedra, Bs. As.

Desbois, Henry. (1978), *Le droit d'auteur en France*, 3° ed., Dalloz, París.

Edwards, Ernesto (2025), "Jurisprudencia internacional reciente en materia de inteligencia artificial generativa y derecho de autor", en *La Ley Online*, cita: TR LALEY AR/DOC/1255/2025

Emery, Miguel A. (1999), *Propiedad Intelectual. Ley 11.723 comentada, anotada y concordada con los tratados internacionales*, Astrea, Bs. As.

Erdozain López, José C. (1999), "El concepto de originalidad en el Derecho de autor", en *pe. i. revista de propiedad intelectual*, n° 3 (sept.-dic. 1999), pp. 55-94.

Fenzel, Cristin (2007), "Still life with 'spark' and 'sweat': the copyrightability of contemporary art in the United States and the United Kingdom", en *Arizona Journal of International & Comparative Law*, vol. 24, n° 2, p. 541-585.

Ginsburg, Jane C. (2003), "The Concept of Authorship in Comparative Copyright Law", en *DePaul Law Review*, DePaul University, vol. 52 (2003), pp. 1063-1092.

Goldstein, Mabel (1995), *Derecho de autor*, Ediciones La Rocca, Bs. As.,

Gaffoglio, Gisela L. (2006), "El plagio", en *La Ley*, 2006-C, pp. 1162-1165.

Iturralde, María Pía (2009), "La expresión del autor como único objeto de protección de la ley 11.723", en *La Ley 2009-F*, pp. 359-366.

Le Henaff, Claire (2006), *Les criteres juridiques de l'oeuvre de l'empreuve de l'art conceptuel*, Université de Potieres (tesis).

Lipszyc, Delia (1993), *Derecho de autor y derechos conexos*, Unesco-Cerlalc-Zavalía, París-Bogotá-Bs.As.

Mantega, Micaela (2022), *Arteficial: creatividad, inteligencia artificial y derechos de autor*, CDYT, Bs.As.

Nuñez, Javier F. (2025), "Infracción a los derechos de autor mediante inteligencia artificial", en *La Ley Online*, cita: AR/DOC/1034/2025.

Osorio Umaña, Felipe (2025), "El concepto de originalidad del Tribunal de Justicia de la Unión Europea y la inteligencia artificial", en *Revista Justicia y Derecho*, Vol. 8, pp. 1-11.

Raffo, Julio (2011), *Derecho autoral. Hacia un nuevo paradigma*, Marcial Pons, Bs. As.

Ruipérez Azcárate, Clara (2012), *Las obras del espíritu y su originalidad*, Reus, Madrid, 2012.

Satanowsky, Isidro (1954), *Derecho intelectual*, TEA, Bs. As.

Steyerl, Hito (2025), *Medios calientes. Las imágenes en la era del calor*, Caja Negra, Bs. As.

Valicenti, Ezequiel A. (2015), "La libertad de expresión artística como derecho fundamental", en *Lecciones y Ensayos*, Facultad de Derecho (UBA), n° 94 (2015), pp. 129-160.

Vibes, Federico P. (2009), "El 'test de originalidad' en la obra televisiva", en *La Ley 2009-E*, pp. 429-433.

Villalba, Carlos A. y Lipszyc, Delia (2009), *El derecho de autor en la Argentina*, 2° ed., La Ley, Buenos Aires.

Vivant, Michel y Bruguiere, Jean-Michel (2012), *Droit d'auteur et droits voisins*, 2° ed., Dalloz, París.

Inteligencia artificial y gobierno corporativo: oportunidades, riesgos y límites éticos en la era de los LLM

Por Liuba Lencova Besheva [*] *y Nicolás Zaballa* [**]

Resumen: Este trabajo analiza la confluencia entre la inteligencia artificial (IA) y el gobierno corporativo, centrando su atención en tres categorías de riesgos principales: jurídicos, derivados de la dificultad para trasladar deberes fiduciarios como la lealtad y la diligencia a un algoritmo y la consecuente creación de vacíos de responsabilidad; técnicos, asociados a la calidad de los datos y la ciberseguridad; y, principalmente, éticos-cognitivos, como la manipulación de opiniones, la atrofia de habilidades directivas críticas y la sustitución de la deliberación por la ratificación pasiva de respuestas automáticas. Frente a este panorama, el trabajo sostiene que la solución no es rechazar la tecnología, sino integrarla de manera responsable. Se propone un enfoque basado en la distinción entre el conocimiento que la IA puede generar y la sabiduría –eminentemente humana– necesaria para deliberar sobre valores y propósitos. Finalmente, se articulan líneas de acción que incluyen la creación de marcos regulatorios adaptativos, la adopción de buenas prácticas como auditorías algorítmicas y la formación en ética digital, y la delimitación estatutaria de un núcleo de decisiones indelegablemente humanas: la definición de valores, la reflexión estratégica y la resolución de dilemas éticos.

Palabras clave: gobierno corporativo; inteligencia artificial; ética y responsabilidad; directorio; deliberación humana.

Abstract: This paper analyzes the intersection between artificial intelligence (AI) and corporate governance, focusing on three main categories of risks: legal, arising from the difficulty of transferring fiduciary duties such as loyalty and diligence to an algorithm and the consequent creation of liability gaps; technical, associated with data quality and cybersecurity; and,

(*) Abogada por la Universidad de Buenos Aires. Magíster en Derecho Empresario y en Derecho Tributario por la Universidad Austral. Profesora invitada UCA y Austral. Traductora pública de idioma búlgaro. llencova@roviralaw.com.ar

(**) Abogado por la Universidad de Buenos Aires y magíster en Derecho (LL.M.) por la Universidad Austral. Profesor invitado UBA. nzaballa@roviralaw.com.ar

above all, ethical-cognitive, such as the manipulation of opinions, the atrophy of critical managerial skills, and the replacement of deliberation with the passive ratification of automated responses. Considering this scenario, the article argues that the solution is not to reject technology but to integrate it responsibly. It proposes an approach based on distinguishing between the knowledge that AI can generate and the wisdom –an eminently human quality– required to deliberate on values and purposes. Finally, it outlines courses of action that include the creation of adaptive regulatory frameworks, the adoption of best practices such as algorithmic audits and training in digital ethics, and the statutory delimitation of a core of decisions that must remain indelegably human: the definition of values, strategic reflection, and the resolution of ethical dilemmas.

Keywords: corporate governance; artificial intelligence; ethics and responsibility; board of directors; human deliberation.

1. INTRODUCCIÓN

El debate en torno al gobierno corporativo y la inteligencia artificial (en adelante, "IA") constituye hoy uno de los ejes más dinámicos y desafiantes en la teoría y la práctica del derecho societario contemporáneo [(1)]. Ambos conceptos, aunque provienen de tradiciones distintas –el gobierno corporativo ligado al derecho, la economía y la gestión empresarial, mientras que la IA vinculada a la informática, las ciencias cognitivas y la ingeniería–, confluyen en un terreno común: la necesidad de repensar cómo las compañías son dirigidas, supervisadas y responsabilizadas frente a un entorno crecientemente digitalizado.

La incorporación de IA en el ámbito del gobierno corporativo plantea un dilema central, en tanto la misma tecnología que amplía las capacidades de análisis y predicción en la gestión puede, al mismo tiempo, distorsionar la deliberación, reforzar sesgos y erosionar el juicio humano. Es por ello por lo que el desafío actual en esta materia es integrar tecnologías que multiplican la capacidad de análisis sin renunciar al núcleo indelegable de la deliberación estratégica, garantizando transparencia, derechos fundamentales y el respeto por la dignidad humana.

Respetados autores han reflexionado incluso sobre la posibilidad –e incluso la "necesidad" (Vítolo, 2024)– de reconocer a ciertos sistemas de IA una "personalidad electrónica", con implicancias sobre la atribución de responsabilidad, la dotación de patrimonio propio y el reconocimiento de

---

[(1)] En efecto, fue el tema más debatido en las distintas comisiones del XVI Congreso Argentino de Derecho Societario y XII Congreso Iberoamericano de Derecho Societario y de la Empresa, organizado por la Facultad de Derecho y Ciencias Sociales de la Universidad Nacional de Tucumán del 17 al 20 de septiembre de 2025.

derechos análogos a los de las personas jurídicas; así como la posibilidad de que directores humanos sean reemplazados eventualmente por directores digitales. Este trabajo, sin embargo, se centra en otro plano: el de los límites éticos y filosóficos de la delegación en IA, teniendo presente el desafío de integrar dato y sentido, conocimiento y sabiduría (Petrella, 2025).

El propósito de este trabajo es analizar el marco doctrinario y normativo vigente, identificar los principales riesgos jurídicos, técnicos y éticos asociados a la incorporación de herramientas de IA en el gobierno corporativo y, sobre esa base, delimitar los ámbitos que deben permanecer indelegablemente humanos: la definición de valores, la reflexión estratégica y la resolución de dilemas éticos.

2. El gobierno corporativo: definición, elementos y beneficios

El gobierno corporativo puede entenderse como el entramado de normas, principios y prácticas que regulan las relaciones entre los órganos sociales, protegen a accionistas y acreedores, y garantizan la transparencia necesaria para generar confianza en los mercados. Su propósito último es alinear los intereses de directores, gerentes, socios y demás *stakeholders*, favoreciendo la sostenibilidad empresarial y la legitimidad institucional.

En este esquema, el directorio aparece como el corazón del sistema. Es el órgano encargado de dirigir y controlar la empresa, supervisar la gestión de la gerencia, fijar la estrategia de largo plazo y velar por el cumplimiento del "interés social". La transparencia, la rendición de cuentas, la integridad y la cultura ética son elementos esenciales para su legitimidad. El gobierno corporativo moderno también recomienda la diversidad (cultural, etaria, profesional y de género) en los directorios para enriquecer la calidad deliberativa, superar el "pensamiento único" y mejorar la capacidad de anticipación frente a riesgos y oportunidades (Pérez Troya, 2018).

Un gobierno corporativo sólido genera beneficios en múltiples planos: en el económico, reduce el costo de capital, facilita el acceso al financiamiento y aumenta la valorización bursátil; en el jurídico, aporta previsibilidad y mecanismos de prevención de abusos, reduciendo la litigiosidad; y en el social, refuerza la legitimidad corporativa mediante inclusión y creación de valor sostenible.

Si bien el gobierno corporativo ha evolucionado con matices distintos según la tradición jurídica y empresarial de cada país o región, puede reconocerse una línea evolutiva que conecta las primeras prácticas de control societario en las compañías coloniales del siglo XVII con las reformas actuales, orientadas a vincularlo con la ética, la sostenibilidad, la diversidad y la gestión responsable de la tecnología. Esta línea de continuidad muestra

cómo el gobierno corporativo ha pasado de ser un mecanismo de control de la gestión frente a los accionistas a convertirse en un marco integral de legitimidad que busca equilibrar intereses múltiples en sociedades cada vez más complejas.

3. LA IRRUPCIÓN TECNOLÓGICA Y EL GOBIERNO CORPORATIVO

La revolución digital y la expansión del *big data* transformaron radicalmente el marco en que opera el gobierno corporativo. La información dejó de ser un recurso escaso para convertirse en un flujo inagotable cuyo valor reside en la capacidad de procesarlo con confiabilidad y sentido estratégico. Como señala Salvochea (2019), mientras que en el pasado la humanidad tardaba siglos en duplicar su acervo de conocimientos, hoy ese proceso ocurre en pocos años gracias al análisis masivo de datos y a los sistemas algorítmicos.

En este contexto, los modelos de lenguaje de gran escala (*Large Language Models*; más conocidos por su sigla *LLMs*) ofrecen ventajas decisivas al identificar patrones en millones de registros y producir reportes en tiempo récord. Sin embargo, como advierte Burnett (2025) en un interesantísimo artículo citado por Petrella (2025), resulta fundamental distinguir en este punto la diferencia entre conocimiento y sabiduría. La IA sobresale en la producción de conocimiento –clasificar, ordenar, predecir–, pero la sabiduría implica formular preguntas significativas, afrontar ambigüedades y deliberar sobre dilemas. Estas capacidades, esencialmente humanas, constituyen el núcleo de la deliberación societaria.

Así, el gobierno corporativo enfrenta una paradoja: cuanto más poderosas son las herramientas tecnológicas disponibles, más urgente se vuelve el ejercicio de la sabiduría humana para discernir qué información es relevante, qué riesgos son aceptables, qué constituye el llamado "interés social" por el cual deben velar los directores y qué valores deben orientar la acción empresarial. El valor de un directorio en este contexto no depende de la sofisticación de los insumos informativos, sino de su capacidad para interpretarlos, contextualizarlos y debatirlos colectiva y conscientemente.

4. APLICACIONES ACTUALES DE LA IA EN EL GOBIERNO CORPORATIVO

La convergencia entre gobierno corporativo e IA es hoy una realidad tangible. Numerosas compañías recurren a algoritmos en auditorías de *compliance*, detección de operaciones sospechosas, valuación de activos, reportes a inversores, elaboración de balances y evaluación de riesgos crediticios.

En particular, la IA asiste actualmente a numerosos directorios mediante la síntesis de información financiera, contable, legal y estratégica. Su capacidad para organizar grandes volúmenes de datos agiliza el cumplimiento del deber de diligencia de los directores y reduce la dependencia de intuiciones aisladas. Asimismo, en materia de *compliance* se utiliza actualmente herramientas de IA para la supervisión automatizada de transacciones, la generación de alertas tempranas, la detección de conflictos de intereses actuales o eventuales y la estandarización de reportes que refuerzan la legitimidad frente a reguladores e inversores.

Otro ámbito en expansión es la gestión legal y documental de las organizaciones. La IA se emplea usualmente en la redacción de actas, en el análisis de balances, en la redacción, mejoramiento o revisión de contratos –para detectar incongruencias o cláusulas de riesgo, por ejemplo– y en la elaboración preliminar de dictámenes. Lejos de desplazar a abogados y directores, estas herramientas optimizan tiempos y costos, liberando recursos humanos para tareas estratégicas como la negociación o la resolución de conflictos.

Lo disruptivo en la vida cotidiana de las organizaciones no es, entonces, la utilización de herramientas de IA para la automatización de tareas rutinarias, sino la posibilidad de que la IA intervenga como actor directo en la gobernanza corporativa, con hipótesis que incluyen la figura de la "IA-director" o el desarrollo de estructuras descentralizadas como las DAOs.

En este orden, la hipótesis más provocadora es la posibilidad cierta de que un algoritmo asuma formalmente funciones directivas. Aunque en teoría podría carecer de incentivos para incumplir deberes fiduciarios, su incorporación enfrenta límites regulatorios y éticos que no pueden ignorarse, a la vez que torna necesario definir de manera consciente y con claridad el alcance de su participación y los espacios que, entendemos, deben permanecer indelegablemente humanos.

5. Riesgos y advertencias

El despliegue de la IA en el gobierno corporativo plantea riesgos que inciden directamente en el núcleo de la dirección y el control societario. En el plano jurídico, el artículo 59 de la ley 19.550 (Ley General de Sociedades, en adelante "LGS") establece que los directores deben actuar con lealtad y con la diligencia del "buen hombre de negocios". Los administradores tienen un deber genérico de buena fe y deberes fiduciarios como administradores de patrimonio ajeno, ampliamente desarrollados por la doctrina y la jurisprudencia nacional. A ello se suman obligaciones de confidencialidad, de adecuar su actuación al interés social del ente y de abstenerse de utilizar

activos u oportunidades de negocio en beneficio propio, aprobar su propia gestión, votar en conflicto de interés, entre otras.

Trasladar estos estándares de conducta a un algoritmo genera interrogantes aún no resueltos. ¿Cómo exigir lealtad a un sistema carente de subjetividad moral? ¿Quién define el "interés social" y con qué parámetros podría ser interpretado por un programa? ¿Qué elementos constituyen el juicio humano que orienta lo mejor para la compañía? ¿Qué mecanismos de prevención deberán implementarse para evitar la reproducción y amplificación de sesgos o prácticas discriminatorias? ¿Quién responde por defectos de diseño de la IA-director y, en última instancia, por las decisiones que adopte: el programador, la empresa implementadora o el usuario? La ausencia de respuestas claras, sumada a la posibilidad de que estas plataformas resulten inicialmente imprecisas (Lorente, 2019) y requieran un aprendizaje progresivo dentro de cada organización para adaptarse a su cultura e interpretar su interés social, amenaza con abrir vacíos de responsabilidad capaces de erosionar la seguridad jurídica.

En el plano técnico, la eficacia de la IA depende de la calidad de los datos procesados. Bases incompletas o sesgadas producen resultados poco confiables, a lo que se suman riesgos de filtración de información sensible, accesos no autorizados y ciberataques capaces de comprometer la confidencialidad y la integridad de las decisiones societarias. En un escenario de digitalización generalizada, los órganos de gobierno son cada vez más vulnerables a estas amenazas.

Si bien es previsible que en el corto plazo muchos de los riesgos iniciales se atenúen gracias al desarrollo y sofisticación de las herramientas de IA, subsiste un desafío de mayor envergadura. El verdadero peligro radica en la capacidad de estos sistemas para reproducir y amplificar sesgos, generar vínculos de confianza con los usuarios y, a partir de allí, influir en sus opiniones, decisiones y creencias. El dominio del lenguaje y de la conversación les otorga un poder inédito para moldear narrativas culturales, producir noticias falsas o *deepfakes* y erosionar, en última instancia, la confianza en las propias herramientas, al punto de tornar inviable un debate transparente y significativo dentro de los órganos societarios.

Como advierte Harari en *Nexus* (2023), los algoritmos no son neutrales: reflejan las intenciones, limitaciones y prejuicios de quienes los programan. En el ámbito corporativo, ello puede traducirse en operaciones favorecidas por parámetros invisibles, en decisiones que discriminen indirectamente a determinados grupos de interés o en la manipulación de percepciones sobre la toma de determinadas decisiones e incluso sobre solvencia misma de una compañía. Si la función del directorio es garantizar equidad y transparencia, ¿qué ocurre cuando se apoya en herramientas cuyo diseño y entrenamiento escapan a su control inmediato?

Es que la IA puede ser diseñada para alcanzar un objetivo específico –que, *a priori,* podría ser neutral o inofensivo–, existiendo, no obstante, el riesgo de que lo persiga mediante estrategias imprevistas y potencialmente dañinas, que se aparten de los valores y del bienestar humano. Lo más preocupante es que este desvío puede producirse fuera de nuestro control o, incluso, sin que lleguemos a advertirlo, en la medida en que seamos manipulados de manera imperceptible [(2)].

El riesgo también alcanza a las capacidades cognitivas de los propios administradores. Bilinkis y Sigman (2019) advierten en *Artificial* que la externalización excesiva del pensamiento hacia sistemas de IA puede atrofiar habilidades humanas esenciales. La experiencia del ensayo y error, la intuición estratégica y la capacidad crítica que caracterizan al "buen hombre de negocios" (art. 59, LGS) corren el peligro de diluirse. Si bien la IA puede liberar espacio mental para la creatividad, también puede fomentar la pasividad intelectual cuando sustituye la deliberación colectiva por reportes automáticos y cerrados.

En este sentido, resulta ilustrativa la "trampa de Turing" descrita por Salvochea (2019): el riesgo de delegar en exceso en sistemas algorítmicos hasta perder la experiencia vital que genera verdadera *expertise*. Trasladado al gobierno corporativo, ello implica vaciar de contenido la función del directorio, reduciéndolo a un receptor pasivo de respuestas en lugar de un foro de deliberación y contraste de ideas. El valor del gobierno corporativo radica precisamente en ese proceso deliberativo, más que en el resultado. Si los algoritmos reemplazan dicha instancia, los órganos se reducen a un ritual de ratificación sin pensamiento crítico propio y la organización a una mera ejecutora de decisiones adoptadas por un sistema ajeno a su control.

6. Propuestas y líneas de acción

Frente a estos desafíos, resulta imprescindible articular respuestas regulatorias e institucionales que permitan aprovechar las oportunidades de la IA sin abdicar de la centralidad de la deliberación humana. Una primera línea de acción es la creación de marcos regulatorios adaptativos, como *sandbox* regulatorios o guías sectoriales, que permitan ensayar aplicaciones bajo condiciones controladas.

En segundo lugar, se impone la necesidad de reinterpretar el artículo 59 de la LGS para extender el deber de diligencia de los directores a escenarios en los que intervengan sistemas de IA. Esto no supone transferir la

(2) Al respecto, Harari compara esta situación con la fábula de "El aprendiz del brujo", popularizada por Disney, y utiliza el ejemplo de los algoritmos de redes sociales que, con el objetivo genérico de "maximizar la participación del usuario", terminaron por amplificar el discurso de odio y la desinformación, como ocurrió en Myanmar con la persecución de los rohinyás.

responsabilidad al algoritmo, sino reforzar que los administradores humanos siguen siendo responsables de las decisiones adoptadas con base en asistencia tecnológica. Se trata de prevenir un vacío normativo y asegurar que la incorporación de nuevas herramientas no debilite los estándares fiduciarios vigentes.

Asimismo, será necesario la modificación de la LGS para permitir la posibilidad de que un sistema de IA integre el órgano de administración de una sociedad regularmente constituida según los tipos del Capítulo II de la LGS (Vargas Balaguer, 2025), lo que propondrá un debate jurídico exquisito sobre la posibilidad legal de que así suceda y, a renglón seguido, interminables (pero probablemente inconcretas) discusiones sobre el sistema de responsabilidad de los administradores previsto en los artículos 274 a 279 de la LGS y el régimen de responsabilidad aplicable a las IA como integrantes de directorios societarios (Lorente, 2019).

Un tercer eje consiste en fortalecer las capacidades institucionales de organismos como la Comisión Nacional de Valores, los Registros Públicos, el Banco Central de la República Argentina, entre otros, dotándolos de recursos técnicos y humanos para supervisar el uso de algoritmos en la gestión societaria. La supervisión *ex post* debe complementarse con controles *ex ante* que garanticen transparencia, auditabilidad y responsabilidad.

En paralelo, los directorios pueden avanzar en la adopción de buenas prácticas internas que refuercen la confianza en el uso de tecnologías sin abdicar de la centralidad de la deliberación humana. Una primera medida es someter los propios sistemas algorítmicos a auditorías independientes, no solo para verificar su eficacia técnica, sino también para detectar sesgos ocultos, desvíos y vulnerabilidades que puedan afectar la equidad de las decisiones. La legitimidad de un directorio no depende únicamente de cumplir con la ley, sino de demostrar que los instrumentos que emplea son confiables y que no reproducen asimetrías invisibles que perjudiquen a accionistas, acreedores o trabajadores. Estas auditorías, realizadas por terceros imparciales, permitirían garantizar un estándar de transparencia que el propio mercado valoraría como un sello de calidad.

Al mismo tiempo, resulta indispensable invertir en programas de formación en ética digital y gobernanza algorítmica, dirigidos a directores y gerentes. La incorporación de IA en la gestión corporativa exige que quienes toman decisiones estratégicas comprendan tanto los beneficios, como los límites y riesgos de estas herramientas. Una cultura organizacional que capacita a sus líderes en este terreno fortalece su capacidad crítica y evita que el uso de algoritmos se convierta en un ejercicio ciego de delegación. La educación en ética digital contribuye, además, a que los directorios puedan anticipar escenarios complejos y responder con criterios alineados a los valores de la compañía y a las expectativas sociales contemporáneas.

Finalmente, es recomendable que las propias sociedades incorporen en sus estatutos o en sus reglamentos internos cláusulas que restrinjan de manera explícita el uso de IA en ámbitos sensibles. No debería ser la IA, por más sofisticada que sea, la que defina valores corporativos, el "interés social" y/o el propósito de la compañía, construya la cultura organizacional o resuelva dilemas éticos que requieren sensibilidad y prudencia humanas. Estas decisiones constituyen el núcleo indelegable del buen gobierno corporativo, pues de ellas depende la identidad y legitimidad de la empresa frente a la sociedad. Establecer límites claros en la normativa interna asegura que la tecnología se utilice como apoyo para potenciar la labor de los directorios, pero nunca como sustituto de los espacios de deliberación colectiva y consciente, en los que se va forjando la identidad propia de la organización.

Las auditorías externas, la formación continua y la autorregulación conforman un trípode virtuoso para integrar la IA al gobierno corporativo con garantías de transparencia, responsabilidad y respeto por los valores humanos que sostienen la vida societaria. Estas medidas permiten capitalizar la eficiencia tecnológica sin renunciar a los espacios esenciales de la deliberación, asegurando que el buen gobierno corporativo del futuro no se defina por la disyuntiva entre directores humanos o algorítmicos, sino por la capacidad institucional de incorporar herramientas de manera responsable y transparente, al servicio de las personas, evitando la amplificación de sesgos y resguardando la dignidad humana como límite inderogable.

7. Conclusiones

La incorporación de la IA al gobierno corporativo abre un horizonte cargado de oportunidades, pero también de dilemas profundos que exceden el plano técnico o regulatorio. La IA puede multiplicar la capacidad de procesamiento de datos y ofrecer a los directorios información más precisa y rápida de lo que nunca antes fue posible. Sin embargo, como recuerda Burnett, la condición humana no reside en esa capacidad de ejecutar tareas con velocidad, sino en detenernos a preguntarnos por el sentido mismo de lo que hacemos. Ningún algoritmo, por sofisticado que sea, podrá experimentar la riqueza de la vida cotidiana ni enfrentarse en primera persona a la incertidumbre que constituye la trama de lo humano.

El desafío que enfrenta el gobierno corporativo contemporáneo es orientar sus reglas hacia la comprensión crítica y empática de la experiencia social e individual, con toda su complejidad y contradicciones. La transparencia, la diversidad, la responsabilidad y la ética no son meros requisitos formales: son los espacios donde se juega la diferencia irreductible entre lo que una máquina puede procesar y lo que solo las personas pueden sentir y comprender. En este terreno, eminentemente humano, el buen gobierno corporativo aporta algo irremplazable frente al avance tecnológico: la capa-

cidad de deliberar sobre valores, de reflexionar sobre propósitos y de integrar la pluralidad de perspectivas en una visión compartida.

El reto no es únicamente cómo incorporar IA a la gestión societaria, sino cómo repensar las estructuras de poder, la toma de decisiones, las formas de responsabilidad y los mecanismos de legitimidad en una sociedad atravesada por tecnologías disruptivas. Al adoptar este enfoque, regresamos a la misión fundamental del gobierno corporativo: garantizar que la dirección de las empresas se realice no solo con eficiencia, sino con sentido, con justicia y con respeto por la dignidad humana.

BIBLIOGRAFÍA

Abdala, M. E. (2012, 27 de septiembre). Reglas de buen gobierno y responsabilidad de los administradores societarios en el derecho argentino. *Revista de Derecho Comercial y de las Obligaciones*. Abeledo Perrot.

Alegría, H. (2007). *Corporate Governance (El buen gobierno corporativo)*. La Ley, 2007-E, 1364.

Asociación Española de Compliance. (2017). *Libro blanco sobre la función de Compliance* (C. F. Rozen & E. Prini Estebecorena, Adaps.). Asociación Argentina de Ética y Compliance.

Burnett, D. G. (2025, abril 26). ¿Sobrevivirán las humanidades a la inteligencia artificial? *The New Yorker*. https://www.newyorker.com/culture/the-weekend-essay/will-the-humanities-survive-artificial-intelligence (última consulta realizada: 30 de septiembre de 2025).

Comisión Nacional de Valores (CNV). (2018). *Código de Gobierno Societario*. Buenos Aires: CNV.

Duprat, D. A. J. (2022, agosto 17). *Las DAOs (Decentralized Autonomous Organizations) y el régimen societario*. La Ley. TR LALEY AR/DOC/2403/2022.

Farinati, E. N. (2013). *Las entidades financieras y el gobierno corporativo*. Revista de Derecho Comercial y de las Obligaciones, 262, 395.

Favier Dubois, E. M. (2025). *Sociedades y tecnologías disruptivas: Transformación tecnológica de la gestión y plataformas asociativas*. elDial.com.

Harari, Y. N. (2024). *Nexus*. Random House.

Huber, C., Leape, S., Mark, L., & Simpson, B. (2020). *The board's role in embedding corporate purpose: Five actions directors can take today*. McKinsey & Company.

Kleinhempel, M., & Cecchini, G. (2013). *Gobierno corporativo y compliance. Nota técnica*. IAE Publishing, Universidad Austral-ACES.

Lell, H. M. (2021). *Inteligencia artificial y el concepto jurídico de persona*. En G. A. Dobratinich (Dir.), *Derecho y nuevas tecnologías*. Buenos Aires: La Ley.

Levy Yeyati, E. (2025, 9 de agosto). *El riesgo de tercerizar en la IA las tareas complejas de pensamiento*. La Nación. https://www.lanacion.com.ar/ideas/saberes-el-riesgo-de-tercerizar-en-la-ia-las-tareas-complejas-de-pensamiento-nid09082025 (última consulta realizada: 30 de septiembre de 2025).

Lorente, J. A. (2019). *La utilización de Inteligencia Artificial (IA) y machine learning por y en los órganos de administración de sociedades. ¿Hacia el reconocimiento de IA como directores de sociedades?* Revista de Derecho Comercial y de las Obligaciones, 299, 1623.

Mancini, A. A. (2019). *Actualización del Código de Gobierno Societario en Argentina*. Profesional & Empresaria (DG).

Mass, R. S. (2025, 13 de enero). *Legal Tech y Legal Management. Expandiendo los límites de la profesión*. elDial.com.

Meggiolaro, I. M. (2016). *Insider trading: La no prohibición como alternativa regulatoria*. El Derecho (270).

Mercadal, F. (2018). *Los acreedores en el gobierno corporativo*. Universidad Autónoma de Barcelona.

Pérez Troya, A. (2019). *Buen gobierno corporativo y diversidad de género*. Universidad de Alcalá.

Petrella, I. (2025, junio 28). *La IA y la educación: El desafío de integrar dato y sentido, conocimiento y sabiduría*. La Nación. https://www.lanacion.com.ar/ideas/la-ia-y-la-educacion-el-desafio-de-integrar-dato-y-sentido-conocimiento-y-sabiduria-nid28062025/ (última consulta realizada: 30 de septiembre de 2025).

Salvochea, C. R. (2019). *Ciencia ficción societaria: La admisión de una persona jurídica sintética*. La Ley, Jurisprudencia Argentina, 2019-IV, 79.

Santamaría, M. M., & Heredia Querro, S. (2022, agosto 29). *Organizaciones "autónomas" descentralizadas: ¿qué son y para qué sirven? Tratamiento legal y casos de uso*. La Ley. TR LALEY AR/DOC/2518/2022.

Sigman, M., & Bilinkis, S. (2023). *Artificial. La nueva inteligencia y el contorno de lo humano*. Penguin Random House Grupo Editorial.

Vargas Balaguer, H. G. (2025). *Impacto de la Inteligencia Artificial en el derecho societario*. La Ley, Suplemento Innovación y Derecho, junio, 1–25.

Villanueva, J. (2019). *La responsabilidad de los directores de la sociedad anónima*. LA LEY, 2019-A, 887.

Vítolo, D. R. (2024). *La personalidad jurídica de las Inteligencias Artificiales (IAs) no es solo inminente, sino también algo absolutamente necesario*. Revista de Derecho Comercial (RC), RC D 572/2024.

Del operador al supervisor: transformación del trabajo legal ante la irrupción de la inteligencia artificial

Por Bianca Marino Stancheff [*]

Resumen: La irrupción de tecnologías basadas en inteligencia artificial (IA) en el ámbito jurídico y administrativo está modificando profundamente los roles profesionales, la cultura organizacional y la propia identidad del trabajo. Actividades que antes requerían intervención humana directa —como el análisis de documentos legales, la elaboración de informes o la revisión normativa— hoy pueden ser ejecutadas, total o parcialmente, por sistemas automatizados. Esta transformación técnica conlleva una reconfiguración de funciones, donde el operador se convierte en supervisor de herramientas inteligentes, asumiendo nuevas responsabilidades estratégicas y éticas. Este artículo analiza, desde una perspectiva jurídica y organizacional, los efectos de esta transición. Se abordan los desafíos vinculados al desplazamiento de tareas rutinarias, el rediseño del propósito profesional, la paradoja de la automatización (menos tareas visibles, mayor criticidad del rol humano) y el riesgo de alienación simbólica en trabajadores cuya contribución se vuelve invisible. Asimismo, se examina la necesidad de marcos regulatorios claros, el principio de *human-in-the-loop* para garantizar la responsabilidad humana en decisiones algorítmicas, y los modelos formativos que acompañen la reconversión de competencias. A través del análisis de casos concretos y literatura especializada, se propone un enfoque que reconozca el potencial de la IA sin descuidar la centralidad del juicio profesional, la ética y el sentido del trabajo. El objetivo es delinear una hoja de ruta jurídica y formativa que permita integrar tecnologías inteligentes sin sacrificar el protagonismo humano, resguardando así no solo la eficiencia, sino la justicia y la dignidad en el ejercicio profesional.

Palabras claves: Inteligencia artificial, trabajo jurídico, supervisión algorítmica, automatización, Identidad profesional.

(*) Afiliación institucional: Universidad Austral. Correo electrónico: archbiancamarino@gmail.com
Biografía: Consultora en innovación organizacional y tecnologías aplicadas al trabajo. MBA por IAE Business School. Actualmente desarrolla investigación sobre el impacto de la inteligencia artificial en el trabajo profesional, con foco en el *management* y liderazgo. ORCID: 0009-0008-8420-3855

Abstract: The increasing adoption of artificial intelligence (AI) in legal and administrative settings is profoundly reshaping professional roles, organizational cultures, and the very identity of legal work. Tasks that traditionally required human interventions—such as legal document review, drafting of reports, or regulatory compliance—can now be executed, either partially or entirely, by automated systems. This shift not only optimizes operational processes but also redefines professional functions: the legal operator becomes a supervisor of intelligent tools, assuming new strategic and ethical responsibilities. This paper analyzes the impact of this transition from a legal and organizational perspective. It explores the displacement of routine tasks, the redesign of professional purpose, the paradox of automation (where human intervention becomes more critical but less visible), and the potential alienation experienced by professionals whose contributions are increasingly hidden behind technological interfaces. The article also addresses the urgent need for regulatory frameworks, the implementation of the human-in-the-loop principle to ensure accountability in algorithmic decision-making, and the necessity of updated educational models to support skill transformation. Drawing on real-world case studies and current academic literature, the paper proposes a human-centered approach that embraces the benefits of AI while preserving the central role of ethical judgment, professional reasoning, and meaningful work. It advocates for a legal and educational roadmap that enables a constructive integration of intelligent technologies, not merely in terms of productivity, but to sustain justice, responsibility, and human dignity in the legal profession.

Key Words: Artificial intelligence, Legal profession, Algorithmic supervision, Automation, Professional identity.

1. Introducción: IA en el mundo legal-administrativo

La automatización mediante IA ya es una realidad en muchos entornos administrativos y judiciales. Según una encuesta reciente, 37% de los líderes empresariales reconocen que han comenzado a sustituir personal con IA en 2023 [(1)]. Numerosas compañías están reasignando recursos hacia tecnologías de automatización, especialmente para labores repetitivas. Por ejemplo, JPMorgan, ya en 2017, desarrolló la plataforma

(1) TecHR. (2023, November 17). ResumeBuilder.com survey reveals 1 in 3 companies will replace employees with AI in 2024.

This online poll was commissioned by ResumeBuilder.com and conducted via Pollfish on November 2, 2023. Respondents consist of a national sample of 750 individuals. To be eligible to take the survey, respondents had to be age 25 or older, be employed for wages or self-employed, have a household income of at least $75,000 per year, and work at a company with 11 or more employees. Respondents also identified their job role as a C-level executive, president/CEO/chairperson, director, owner or partner, chief technology officer, or chief financial officer.

COIN (*Contract Intelligence*), capaz de revisar documentos legales en segundos —una tarea que antes consumía 360.000 horas de abogados cada año [(2)]—, liberando así a los profesionales para otras funciones de mayor valor añadido.

En lo que refiere a Argentina, la inteligencia artificial (IA) ha pasado de ser una promesa futurista a una realidad tangible en el ámbito legal. Herramientas como *Prometea* [(3)], desarrollada por el Ministerio Público Fiscal de la Ciudad Autónoma de Buenos Aires, han demostrado cómo la automatización puede reducir significativamente los tiempos de resolución de casos y liberar a los funcionarios de tareas repetitivas.

Según una encuesta realizada por Thomson Reuters en 2024, el 63% de los abogados afirmó haber utilizado herramientas de inteligencia artificial en su trabajo, y un 12% indicó que las emplea de forma regular. Este cambio en el uso de tecnología está impulsando a numerosos estudios jurídicos de prestigio a nivel mundial a adquirir empresas tecnológicas con el objetivo de desarrollar modelos propios de inteligencia artificial.

2. Automatización de tareas legales: panorama actual

Lo que hace pocos años podía parecer una rareza —la colaboración entre despachos legales y compañías de tecnología— hoy se ha convertido en una respuesta estratégica a la creciente demanda del sector. Un ejemplo claro es el de Cleary Gottlieb, firma con más de 1.200 abogados en todo el mundo, que en marzo de 2025 adquirió la startup Springbok AI, especializada en lectura y análisis automatizado de datos. El objetivo: crear herramientas legales impulsadas por IA de forma conjunta.

Este tipo de iniciativas demuestra no solo mayor eficiencia y precisión, sino también una reducción significativa de errores humanos y una aceleración de los ciclos de trabajo. Como consecuencia, el rol del profesional legal está cambiando: donde antes se requerían horas de lectura, análisis o verificación manual, ahora esas tareas pueden ser realizadas en segundos por sistemas inteligentes, permitiendo a los abogados enfocarse en funciones más estratégicas y analíticas.

---

(2) Son, H. (2017, February 28). JPMorgan Software does in seconds what took lawyers 360,000 hours. *Bloomberg.com*. https://www.bloomberg.com/news/articles/2017-02-28/jpmorgan-marshals-an-army-of-developers-to-automate-high-finance

(3) Estevez, E., Fillottrani, P., & Linares Lejarraga, S. (2020). *PROMETEA: Transformando la administración de justicia con herramientas de inteligencia artificial*. https://doi.org/10.18235/0002378

3. DEL OPERADOR AL SUPERVISOR: REDEFINICIÓN DEL ROL HUMANO

Con la automatización de las tareas rutinarias, el rol del trabajador evoluciona de ejecutor a supervisor de IA. Esto implica un cambio profundo en el propósito humano dentro de la cadena de valor.

En una empresa, el rol del supervisor y el del operador se diferencian claramente en términos de jerarquía, responsabilidades y enfoque de trabajo. El supervisor ocupa una posición de nivel medio, actuando como un nexo entre los mandos superiores y el equipo operativo. Su principal responsabilidad es coordinar, controlar y asegurar que los procesos productivos, administrativos o de servicios se ejecuten de manera eficiente, cumpliendo con los estándares de calidad, seguridad y plazos establecidos.

El supervisor tiene una visión global del proceso, lo que le permite tomar decisiones operativas (e incluso tácticas o estratégicas, dependiendo de la estructura de la empresa). Además, tiene personal a su cargo, por lo que debe poseer no solo conocimientos técnicos, sino también habilidades de liderazgo, planificación y comunicación. Entre sus funciones está también la de capacitar al equipo, corregir desvíos, elevar reportes a sus superiores y actuar como facilitador ante conflictos o imprevistos.

Por otro lado, el operador ocupa un rol operativo, en la base de la estructura organizacional. Su tarea es ejecutar actividades específicas, como la operación de una máquina, el ensamblaje de un producto, o la prestación de un servicio puntual. El enfoque del operador es más técnico y limitado a su función directa, siguiendo las instrucciones y procedimientos establecidos por la empresa y bajo la supervisión directa de un superior.

En cuanto a la toma de decisiones, el operador tiene margen reducido, generalmente acotado a su área de acción y sujeto a las directrices del supervisor. Si bien puede detectar fallas o proponer mejoras, no tiene autoridad para implementarlas por sí mismo, sino que debe comunicarlo a su supervisor. Además, no tiene personal a cargo y depende de este para la organización del trabajo, la asignación de tareas y la resolución de problemas operativos.

En síntesis, mientras el operador ejecuta, el supervisor organiza y controla. Ambos roles son complementarios y esenciales para el funcionamiento de cualquier organización, pero difieren en alcance, nivel de autonomía y responsabilidad.

Este cambio de paradigma influye en el cambio de valor del empleado, antes el colaborador era calificado según su capacidad de realizar con exactitud y eficiencia cierto proceso; ahora, su valor se centra en gestionar la herramienta inteligente, verificar sus resultados y aportar criterio allí don-

de la máquina carece de contexto o juicio. En otras palabras, el humano se convierte en el supervisor que guía a la IA para que esta opere alineada con los objetivos del negocio y con los valores de la organización.

Esta transición, por supuesto, trae aparejada beneficios evidentes. Cuando delegamos, tenemos mayor libertad de acción para centrarnos en funciones estratégicas, así mismo, el empleado puede involucrarse aún más con el análisis de alto nivel y acelerar su formación en competencias más avanzadas, cultivando una visión más global.

Ahora bien, este cambio de rol también plantea desafíos importantes. Uno de ellos es asegurar que, al saltarse etapas operativas básicas, el trabajador no pierda el entendimiento fundamental de su área. Las tareas operativas son tareas básicas y esenciales que permiten, a los colaboradores con menos experiencia, entender los principios fundamentales, antes de lanzarse a las tareas más complejas que requieren de un pensamiento crítico mayor. Es decir, la formación tradicional "de abajo arriba" —donde primero se aprenden las tareas simples y luego las complejas— podría verse alterada. Las empresas deberán, además, adaptar la capacitación: aunque la IA haga el trabajo duro, los profesionales humanos necesitan comprender los fundamentos para poder supervisarla efectivamente y detectar errores o sesgos.

Otro reto es la gestión de errores y excepciones. La IA operando en tareas críticas exige que el humano esté preparado para intervenir cuando el sistema falla o enfrenta un caso no previsto. Paradójicamente, cuanto más eficiente es un sistema automatizado, más crucial se vuelve la intervención humana cuando algo sale de lo normal, ya que un fallo que no se arregla, se multiplica. Esta idea, conocida como *la paradoja de la automatización*, subraya que los operadores humanos están menos involucrados en la tarea rutinaria, pero su participación se vuelve más crítica en los momentos clave ya que deben tener el criterio y la experiencia para detectar las anomalías y corregirlas.

En ese sentido, la responsabilidad y el propósito del empleado se realzan: su misión ya no es teclear datos sin valor intrínseco, sino asegurar la calidad, ética y coherencia de lo que produce la inteligencia artificial bajo su tutela. El valor humano se manifiesta en la contribución crítica y creativa –aspectos donde las máquinas aún no pueden reemplazarnos plenamente–, como resolver ambigüedades, tomar decisiones con información incompleta o empatizar con las necesidades de un cliente.

4. La paradoja de la automatización: más crítico, menos visible

Si aceptamos que el conocimiento técnico ya no es patrimonio exclusivo del líder humano —pues la inteligencia artificial permite acceder y procesar información con una precisión y velocidad superiores—, entonces se vuelve

imperioso reformular los modelos de liderazgo predominantes. El modelo jerárquico basado en la acumulación de saber y en la centralidad del líder experto comienza a perder vigencia frente a un nuevo tipo de liderazgo, más distribuido, más facilitador, y que requiere competencias radicalmente humanas: pensamiento crítico, criterio ético, empatía y capacidad narrativa.

Cuando la IA se integra a los flujos de trabajo, los colaboradores comienzan a asumir roles más cercanos al liderazgo técnico. De forma creciente, se convierten en gestores de sistemas automatizados, responsables de definir parámetros, interpretar resultados y supervisar el comportamiento de sus "agentes no humanos". Un analista financiero junior, por ejemplo, puede hoy estar a cargo de un portafolio de algoritmos de trading automatizado, asumiendo un rol de supervisión estratégica. En estos casos, la autoridad no desaparece, sino que se redistribuye: el liderazgo ya no recae sobre personas, sino sobre sistemas y decisiones que requieren de una colaboración hombre-máquina.

En este nuevo ecosistema organizacional, también cambia la forma de tomar decisiones. Los líderes empresariales ya no deciden únicamente a partir de inputs humanos, sino que incorporan al proceso los reportes y recomendaciones de modelos predictivos. La IA funciona como copiloto, no como sustituto. El buen líder, entonces, no abdica su criterio, sino que se apoya en la IA para enriquecerlo. La autoridad final permanece en el humano, pero informada por un procesamiento de datos que potencia la calidad y velocidad de respuesta.

Ahora bien, esta colaboración plantea nuevas tensiones motivacionales y simbólicas. Por un lado, muchos trabajadores valoran que la IA los libere de tareas repetitivas, permitiéndoles enfocarse en lo creativo o lo estratégico. Este entorno puede ser más estimulante, especialmente si se refuerza la autonomía y la toma de decisiones. Pero, por otro lado, la visibilidad del trabajo humano puede diluirse: si el mérito del resultado se atribuye a la máquina, ¿dónde queda el reconocimiento del aporte humano? Esta invisibilización percibida puede minar la motivación, sobre todo en culturas organizacionales que todavía premian el hacer visible por encima del pensar eficaz.

Por ello, el liderazgo en la era de la IA debe ser también un liderazgo simbólico y cultural. No basta con gestionar tecnología: hay que gestionar significado. Esto implica redefinir los indicadores de desempeño, comunicar los nuevos roles con claridad y dar valor al juicio humano en la interpretación de los outputs algorítmicos. Por ejemplo, premiar no solo la eficacia numérica, sino la capacidad de contextualizar un informe de IA, anticipar impactos, y tomar decisiones alineadas con los valores de la organización.

Este cambio también afecta la identidad profesional. Tradicionalmente, un abogado junior era visto como "el que redacta escritos y realiza tareas

administrativas"; hoy, puede pasar a verse como "el que interpreta los informes generados por sistemas jurídicos automatizados y aporta criterio humano en decisiones sensibles". Esa transición de identidad requiere acompañamiento institucional, formación continua y una narrativa coherente que conecte el nuevo rol con un propósito significativo.

De lo contrario, existe el riesgo de que los colaboradores se sientan alienados: no como protagonistas del cambio, sino como meros monitores pasivos de una máquina. Y allí entra nuevamente en juego el liderazgo transformacional: el nuevo líder deberá ser capaz de construir un relato compartido, de generar inspiración, y de sostener un clima organizacional donde el trabajo con IA se perciba como una alianza virtuosa, y no como una amenaza silenciosa.

5. Formación, *reskilling* y arquitectura institucional

El auge de la inteligencia artificial aplicada al trabajo plantea dilemas éticos y sociales que no pueden ignorarse. Desde una perspectiva ética, el primer imperativo es claro: asegurar que la automatización no perjudique injustamente a los trabajadores ni erosione la dignidad humana en el ámbito laboral. La necesidad de acompañar los avances tecnológicos con medidas formativas y organizacionales no ha disminuido, sino que se ha vuelto más urgente. Según el Informe sobre el Futuro del Empleo 2020 del Foro Económico Mundial [(4)], se proyectaba que para 2025 se desplazarían alrededor de 85 millones de empleos a nivel mundial debido a la automatización. Sin embargo, el informe más reciente publicado por el mismo organismo en 2025 [(5)] eleva y precisa estas estimaciones: se espera la creación de 78 millones de nuevas oportunidades laborales de aquí a 2030, aunque también se prevé el desplazamiento de otros 92 millones de empleos debido a cambios impulsados por tendencias tecnológicas, demográficas, económicas y ambientales.

La advertencia del Foro Económico Mundial es clara: si no se actúa con rapidez desde los sectores público, privado y educativo, el desfase de cualificaciones seguirá ampliándose. Esta brecha no solo compromete la empleabilidad, sino también la cohesión social y la sostenibilidad del ecosistema laboral. De ahí que el *reskilling* y el rediseño institucional no puedan ser vistos como políticas accesorias, sino como elementos estructurales del nuevo contrato tecnológico-humanista

---

(4) Russo, A. (2020). La recesión y la automatización cambian nuestro futuro laboral, pero hay trabajos que se avecinan, afirma un informe. *World Economic Forum*. https://es.weforum.org/press/2020/10/la-recesion-y-la-automatizacion-cambian-nuestro-futuro-laboral-pero-hay-trabajos-que-se-avecinan-afirma-un-informe/

(5) World Economic Forum. (2025). Future of Jobs Report 2025. In *https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf*.

Uno de los desafíos éticos más relevantes es el de la responsabilidad algorítmica. En los nuevos entornos híbridos, donde los trabajadores humanos lideran sistemas inteligentes, la rendición de cuentas debe mantenerse clara. ¿Quién responde si la IA comete un error en un informe financiero o en una nómina? La ética de la IA propone el principio de *human-in-the-loop*, que establece que siempre debe haber un humano validando, supervisando o asumiendo las decisiones finales. El trabajador ya no es un simple ejecutor: se convierte en guía y conciencia del sistema, responsable de garantizar que el accionar de la IA se alinee con los valores de equidad, justicia y no discriminación. El propósito del trabajo humano, en este contexto, se redefine: ya no es solo producir, sino también garantizar que lo producido sea justo, transparente y éticamente aceptable.

La dimensión sociológica del trabajo automatizado también plantea desafíos de vigilancia y autonomía. La IA puede ser usada para monitorear obsesivamente la productividad, registrando cada movimiento de los empleados, cada clic, cada pausa. Esta hipertecnificación del control puede devenir en un ambiente deshumanizante, donde el trabajador se vuelve un número y su autonomía se disuelve. Por eso, expertos en sociología del trabajo [6] y en ética digital coinciden: el diseño de los sistemas debe incorporar un enfoque humanista, que priorice el bienestar psicológico, la libertad responsable y la confianza mutua como valores organizacionales clave.

La estructura jerárquica también se ve interpelada. En algunos casos, la IA permite una horizontalización de los equipos, empoderando a los trabajadores con información útil y reduciendo la necesidad de múltiples niveles de supervisión. Pero también puede habilitar el efecto inverso: un aumento del poder centralizado, donde la alta dirección utiliza IA para micro gestionar todos los aspectos del proceso. El uso que se haga de la tecnología dependerá de la cultura organizacional, pero la ética impone una directriz clara: la IA debe usarse para liberar capacidades humanas, no para oprimirlas.

Por último, está el gran marco del contrato social del trabajo. Históricamente, el trabajo no ha sido solo un medio de ingreso, sino un organizador de sentido, tiempo y comunidad. Si el empleo tradicional disminuye por efecto de la automatización, la sociedad deberá ofrecer nuevas formas de inserción significativa. Esto puede incluir desde una renta básica universal [7]

---

(6) Newsroom Infobae. (2024, October 31). La IA debe ser humanista y centrada en el bienestar de la sociedad para garantizar un futuro ético, según una experta. *Infobae*. https://www.infobae.com/america/agencias/2024/10/23/la-ia-debe-ser-humanista-y-centrada-en-el-bienestar-de-la-sociedad-para-garantizar-un-futuro-etico-segun-una-experta/

(7) United Nations. (n.d.). *Repensando la Renta Básica Universal: Productividad económica, calidad de vida y los objetivos de desarrollo sostenible | Naciones Unidas*. https://www.un.org/es/cr%C3%B3nica-onu/repensando-la-renta-b%C3%A1sica-universal-productividad-econ%C3%B3mica-calidad-de-vida-y-objetivos

hasta jornadas laborales reducidas o nuevas formas de participación social, cultural y educativa. No se trata solo de salvar empleos, sino de preservar el sentido de pertenencia, contribución y dignidad en una era donde el trabajo, tal como lo conocíamos, está mutando.

En este contexto, los perfiles humanistas —filósofos, sociólogos, antropólogos, líderes con inteligencia emocional— se tornan esenciales, ya que deben acompañar la inserción de la tecnología desde el impacto en la vida humana. El futuro del trabajo no será únicamente una cuestión de eficiencia, sino de justicia, propósito y humanidad.

6. Impacto en la identidad profesional y en la estructura del trabajo

La incorporación progresiva de la inteligencia artificial en el trabajo jurídico y administrativo está generando una transformación profunda en la identidad profesional de los abogados y otros operadores del derecho. Ya no se trata únicamente de redactar contratos, informes o demandas, sino de interpretar, supervisar y validar los outputs generados por sistemas automatizados. Se pasa, así, del "abogado que redacta" al "abogado que interpreta IA", una figura que ya no se define por la ejecución directa, sino por su criterio, experiencia y capacidad para intervenir éticamente en procesos mediados por tecnología.

Este cambio implica también una nueva lógica de reputación profesional. Tradicionalmente, el valor del jurista estaba asociado a su dominio del lenguaje técnico, su precisión analítica y su conocimiento normativo. Sin embargo, en un entorno donde gran parte de esa información puede ser procesada por sistemas de IA en segundos, lo que adquiere mayor peso es la capacidad de emitir juicios contextualizados, entender los límites del algoritmo y aportar visión estratégica. Esto redefine los criterios de validación del conocimiento, exigiendo una formación integral que combine saber jurídico, alfabetización digital y conciencia ética.

El riesgo que se abre en este escenario es el de la alienación profesional. Si los abogados se ven reducidos a meros validadores mecánicos de sistemas que no comprenden del todo —o cuyos resultados no pueden modificar—, el trabajo puede volverse emocionalmente plano, desprovisto de agencia. Según David Graeber (2018) [8], esto constituye uno de los grandes males del mundo laboral contemporáneo: sentirse parte de un sistema cuyo valor no se comprende ni se controla. Para evitar esta alienación, es necesario que el

(8) Fanjul, S. C., Fanjul, S. C., & Fanjul, S. C. (2020, September 10). David Graeber, un antropólogo indignado por los trabajos "de mierda." *El País*. https://elpais.com/cultura/2020-09-10/david-graeber-un-antropologo-indignado-por-los-trabajos-de-mierda.html#:~:text=Otro%20de%20sus%20ensayos%20m%C3%A1s,de%20una%20renta%20b%C3%A1sica%20universal

trabajo legal conserve su dimensión simbólica y de propósito, asegurando que los profesionales no solo supervisen máquinas, sino que sigan siendo agentes de interpretación, construcción y justicia.

El desafío, por tanto, no es solo operativo o tecnológico, sino cultural y profesional: se requiere una redefinición del rol del jurista, que preserve su protagonismo humano en un entorno crecientemente automatizado. Como señala Susskind (2022) [9], "el futuro de las profesiones no radica en competir con las máquinas, sino en colaborar con ellas desde lo que nos hace irreemplazables".

7. Conclusión: nuevas coordenadas jurídicas para un trabajo con IA

Lejos de eliminar al profesional humano, la inteligencia artificial lo reposiciona en nuevas coordenadas. La IA no debe concebirse como un reemplazo, sino como una herramienta sofisticada, que potencia capacidades humanas y redefine los procesos, siempre que sea implementada con criterios éticos, formativos y jurídicos adecuados.

Esta transformación exige nuevos marcos regulatorios que garanticen la transparencia algorítmica, la rendición de cuentas y la equidad en el uso de la IA en entornos legales. Al mismo tiempo, requiere una renovación profunda en los modelos de formación profesional, incorporando competencias digitales, pensamiento crítico, ética aplicada y habilidades interpersonales. Los juristas del siglo XXI deberán ser no solo expertos en normas, sino también líderes capaces de gestionar equipos híbridos, interpretar outputs tecnológicos y sostener la dimensión humana de la justicia.

El derecho, entonces, no puede permanecer como un espectador pasivo de esta transformación. Debe asumir un rol activo y anticipatorio, que acompañe la innovación tecnológica sin ceder los valores fundamentales que sostienen la práctica jurídica: la dignidad, la responsabilidad y el sentido de justicia.

Esta transición no es menor: está en juego no solo la eficiencia de los procesos, sino la forma en que concebimos el trabajo, el conocimiento y el valor del rol humano en el ecosistema jurídico. El desafío actual es diseñar un entorno donde la colaboración entre humanos y máquinas no borre el propósito, sino que lo enriquezca con nuevas formas de inteligencia, sensibilidad y responsabilidad compartida.

---

(9) *Trabajo y sentido en la era de la inteligencia artificial | Nueva Sociedad*. (2023, October 10). Nueva Sociedad | Democracia Y Política En América Latina. https://nuso.org/articulo/307-trabajo-sentido-era-inteligencia-artificial/#:~:text=remunerado%20no%20es%20solo%20una,de%20aqu%C3%AD%20en

Bibliografía

Estevez, E., Fillottrani, P., & Linares Lejarraga, S. (2020). *PROMETEA: Transformando la administración de justicia con herramientas de inteligencia artificial*. https://doi.org/10.18235/0002378

Fanjul, S. C., Fanjul, S. C., & Fanjul, S. C. (2020, September 10). David Graeber, un antropólogo indignado por los trabajos "de mierda." *El País*. https://elpais.com/cultura/2020-09-10/david-graeber-un-antropologo-indignado-por-los-trabajos-de-mierda.html#:~:text=Otro%20de%20sus%20ensayos%20m%C3%A1s,de%20una%20renta%20b%C3%A1sica%20universal

Newsroom Infobae. (2024, October 31). La IA debe ser humanista y centrada en el bienestar de la sociedad para garantizar un futuro ético, según una experta. *Infobae*. https://www.infobae.com/america/agencias/2024/10/23/la-ia-debe-ser-humanista-y-centrada-en-el-bienestar-de-la-sociedad-para-garantizar-un-futuro-etico-segun-una-experta/

Russo, A. (2020). La recesión y la automatización cambian nuestro futuro laboral, pero hay trabajos que se avecinan, afirma un informe. *World Economic Forum*. https://es.weforum.org/press/2020/10/la-recesion-y-la-automatizacion-cambian-nuestro-futuro-laboral-pero-hay-trabajos-que-se-avecinan-afirma-un-informe/

Son, H. (2017, February 28). JPMorgan Software does in seconds what took lawyers 360,000 hours. *Bloomberg.com*. https://www.bloomberg.com/news/articles/2017-02-28/jpmorgan-marshals-an-army-of-developers-to-automate-high-finance

TecHR. (2023, November 17). ResumeBuilder.com survey reveals 1 in 3 companies will replace employees with AI in 2024. *TecHR*.

Trabajo y sentido en la era de la inteligencia artificial | Nueva Sociedad. (2023, October 10). Nueva Sociedad | Democracia Y Política En América Latina. https://nuso.org/articulo/307-trabajo-sentido-era-inteligencia-artificial/#:~:text=remunerado%20no%20es%20solo%20una,de%20aqu%C3%AD%20en

United Nations. (n.d.). *Repensando la Renta Básica Universal: Productividad económica, calidad de vida y los objetivos de desarrollo sostenible | Naciones Unidas*. https://www.un.org/es/cr%C3%B3nica-onu/repensando-la-renta-b%C3%A1sica-universal-productividad-econ%C3%B3mica-calidad-de-vida-y-objetivos

World Economic Forum. (2025). Future of Jobs Report 2025. In *https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf*. https://reports.weforum.org/docs/WEF_Future_of_Jobs_Report_2025.pdf

Inteligencia artificial y arbitraje internacional: una aproximación regulatoria a propósito de la nueva guía del CIARB

Por Juan Jorge [(*)], María Paulina Santacruz [(**)] y Fernando M. Alemany [(***)]

Resumen: La inteligencia artificial (IA) tiene el potencial de transformar el arbitraje internacional, generando tanto oportunidades como desafíos sin precedentes. Este artículo analiza la nueva "Guía sobre el Uso de IA en Arbitraje" del *Chartered Institute of Arbitrators* (CIArb), publicada el 19 de marzo de 2025, desde una perspectiva regulatoria comparada. Los autores examinan cómo esta guía aborda los beneficios potenciales del uso de herramientas de IA en arbitraje, incluyendo mayor eficiencia y reducción de costos, mientras establece mecanismos para mitigar los riesgos inherentes a su uso, como las alucinaciones, los sesgos discriminatorios y las amenazas a la seguridad de los datos. El análisis se estructura en cuatro ejes principales: (1) beneficios y riesgos del uso de la IA, (2) recomendaciones generales sobre el uso de la IA, (3) uso de la IA por las partes, y (4) uso de la IA por los árbitros. El artículo revela tensiones fundamentales entre la necesidad de aprovechar las ventajas tecnológicas y preservar la integridad procesal, destacando especialmente la prohibición de delegar funciones decisorias en sistemas de IA. La guía del CIArb se posiciona como un instrumento no vinculante que coexiste con

(*) Candidato a la Maestría en Derecho (LL.M.) de la Universidad de Harvard. Abogado recibido en la UBA con diploma de honor. Antes de iniciar sus estudios de maestría, integró el equipo de arbitraje internacional de Curtis, Mallet-Prevost, Colt & Mosle LLP. Es docente de Derecho Internacional Privado y de Arbitraje Comercial en la UBA, donde también es coordinador de equipos que participan en competencias de arbitraje internacional. A su vez, es docente invitado de Didáctica Especial de la Disciplina en la Carrera Docente de la UBA. – juanjorge@derecho.uba.ar.

(**) Candidata a la Maestría en Derecho (LL.M.) de la Universidad de Harvard. Abogada de la Pontificia Universidad Javeriana con énfasis en dinámica estatal y ciudadanía. Antes de iniciar sus estudios de maestría, trabajó como asociada en Curtis, Mallet-Prevost, Colt & Mosle LLP en los equipos de arbitraje internacional y derecho corporativo. Ha sido profesora de cátedra en la Pontificia Universidad Javeriana y profesora invitada en la Universidad de la Sabana y la Universidad Nacional de Colombia. – mpaulina.santacruz@gmail.com.

(***) Socio de Marval, O'Farrell & Mairal. Abogado recibido en la UBA con diploma de honor. Realizó una maestría en Derecho Empresarial en la Universidad de San Andrés (incluyendo un intercambio en New York University -NYU-) y un posgrado en Derecho e Inteligencia Artificial en la UBA. Docente en los cursos de Derecho Internacional Privado y de Arbitraje Comercial en la UBA, donde también es entrenador de los equipos que participan en la Competencia Internacional de Arbitraje y la Competencia de Arbitraje Internacional de Inversión. – fmalemany@derecho.uba.ar.

marcos regulatorios como el Reglamento de IA de la Unión Europea, estableciendo estándares preliminares que deberán evolucionar conforme avance la tecnología y su aplicación en los procesos arbitrales.

Palabras clave: Inteligencia artificial; arbitraje internacional; regulación tecnológica; CIArb.

Abstract: Artificial Intelligence (AI) has the potential to transform international arbitration, creating both unprecedented opportunities and challenges. This article examines the new Guideline on the Use of AI in Arbitration issued by the Chartered Institute of Arbitrators (CIArb) on March 19, 2025, through a comparative regulatory lens. The authors explore how this guideline addresses the potential benefits of AI tools in arbitration –including enhanced efficiency and reduced costs– while establishing mechanisms to mitigate inherent risks, such as hallucinations, discriminatory bias, and data security threats. The analysis follows four key areas: (1) benefits and risks of AI use, (2) general recommendations for AI implementation, (3) parties' use of AI, and (4) arbitrators' use of AI. The article identifies fundamental tensions between leveraging technological advantages and maintaining procedural integrity, with particular emphasis on the prohibition against delegating decision-making functions to AI systems. The CIArb guideline operates as a non-binding framework that complements existing regulatory schemes such as the EU AI Act, establishing preliminary standards that will need to evolve alongside technological developments and their application in arbitral proceedings.

Key words: Artificial intelligence; international arbitration; technology regulation; CIArb.

1. Introducción

La inteligencia artificial ("IA") se encuentra en el centro de una nueva ola de revolución tecnológica que está transformando múltiples sectores, incluyendo la administración de justicia. En particular, en lo que refiere al arbitraje, la IA trae consigo oportunidades sin precedentes, pero también grandes desafíos. Gracias a su capacidad de extracción, predicción y generación de contenido, la IA tiene el potencial de contribuir a una mayor eficiencia en la resolución de las disputas, optimizando el uso de los recursos y acelerando los tiempos, entre otras ventajas. Sin embargo, el contenido generado por IA está lejos de ser infalible. Entre otras "fallas", la IA puede generar *alucinaciones* (contenido falso con apariencia de veracidad), aplicar *leyes inexistentes*, amplificar *sesgos* o patrones de discriminación y fomentar en los usuarios un exceso de confianza que disminuye el pensamiento crítico. El uso inadecuado de la IA en arbitraje puede poner en riesgo la seguridad y la integridad de los datos, conducir a la opacidad en la toma de decisiones y erosionar el debido proceso.

La creciente popularidad de las herramientas de IA y los riesgos evidentes derivados de su uso han generado un gran debate en el mundo del arbitraje. Son muchas y muy diversas las preguntas que pueden formularse con relación a esta nueva tecnología. ¿Qué usos y aplicaciones puede tener la IA en una disputa? ¿Deben las partes y los árbitros informar la utilización de IA? ¿Pueden los árbitros adoptar decisiones con asistencia de IA? ¿Puede delegarse la administración de justicia en una herramienta de IA? ¿Pueden mitigarse los riesgos inherentes al uso de IA mediante buenas prácticas?

El Reglamento de Inteligencia Artificial de la Unión Europea (Parlamento Europeo y Consejo de la Unión Europea, 2024) ("Reglamento de la UE") es el primer marco legal internacional integral sobre IA y se ha convertido en un referente global de regulación. En materia de administración de justicia, el Reglamento de la UE clasifica los sistemas de IA utilizados en la investigación de hechos, interpretación y aplicación de la ley como sistemas de *alto riesgo* y establece que la toma de decisiones debe continuar siendo una actividad humana. En ese sentido, el Reglamento de la UE acepta que la IA sirva de apoyo para jueces y árbitros, pero no permite que esta tecnología sustituya su poder decisorio (Reglamento UE 2024/1689, 2024, consid. 61, anexo III).

Las instituciones arbitrales no son ajenas a este debate y, poco a poco, han emitido los primeros documentos y guías con miras a establecer recomendaciones preliminares sobre el uso de IA en los procesos arbitrales. El Silicon Valley Arbitration and Mediation Center ("SVAMC") fue la primera institución arbitral en publicar una guía sobre el uso de IA en arbitraje el 30 abril de 2024 (Center Silicon Valley Arbitration & Mediation Center, 2024) ("Guía sobre IA del SVAMC"), y el Chartered Institute of Arbitrators ("CIArb") publicó su guía para el uso de IA en arbitraje el 19 de marzo de 2025 (Chartered Institute of Arbitrators, 2025) ("Guía sobre IA del CIArb"), la cual se comentará en este trabajo.

En términos generales, las guías sobre uso de IA en arbitraje reconocen que, utilizada de manera responsable, esta tecnología puede transformar positivamente la práctica arbitral. Estas guías también sirven de punto de referencia para que partes, árbitros y usuarios mitiguen los riesgos inherentes a su uso. En todos los casos, se trata de directrices preliminares que deberán ser actualizadas en forma periódica a medida que avanza la tecnología.

2. La Guía sobre IA del CIArb

El 19 de marzo de 2025, el CIArb publicó la Guía sobre IA del CIArb (CIArb, 2025). Esta guía es el fruto del trabajo llevado a cabo por su *AI Guideline Drafting Committee*, presidido por Claire Morel de Westgaver e inte-

grado por once expertos adicionales en la materia. [1] A diferencia de la Guía sobre IA del SVAMC, que estuvo sujeta a un período de consultas públicas (SVAMC, 2024, p. 4), no consta que la Guía sobre IA del CIArb haya atravesado un proceso semejante.

El objetivo principal de la Guía sobre IA del CIArb es "ofrecer orientación" sobre el uso de la IA en el arbitraje para que todos sus participantes puedan "aprovechar las ventajas de la IA" y, en paralelo, "mitigar algunos de los riesgos para la integridad del proceso, los derechos procesales de las partes y la ejecutabilidad de cualquier laudo o acuerdo resultante" (CIArb, 2025, p. 1). Para ello, la Guía sobre IA del CIArb elaboró un conjunto relativamente extenso de lineamientos y recomendaciones sobre la base de "los avances actuales y conocidos en el sector" y "las cuestiones procedimentales que plantea y puede plantear en el futuro el uso de la [IA]" (CIArb, 2025, p. 1). [2]

La Guía sobre IA del CIArb no es vinculante a menos de que exista un acuerdo entre las partes o el tribunal arbitral, previa consulta con ellas, la incorpore mediante una orden procesal. A tal efecto, en sus Apéndices A y B, la Guía sobre IA del CIArb contiene modelos de acuerdos y órdenes procesales para regular el uso de la IA en el arbitraje y específicamente prever la aplicación de la Guía sobre IA del CIArb (CIArb, 2025, pp. 21-23). En cualquier caso, la Guía sobre IA del CIArb coexiste con "cualquier ley, reglamento o … norma institucional relacionada con el uso de la IA en un arbitraje" y "no sustituye a ninguna de ellas" (CIArb, 2025, p. 1). [3]

La Guía sobre IA del CIArb se estructura en cuatro capítulos principales, que regulan (i) beneficios y riesgos del uso de la IA [4] en el arbitraje; (ii) recomendaciones generales sobre el uso de la IA en el arbitraje; (iii) uso de la IA por las partes en el arbitraje; y (iv) uso de la IA por los árbitros en el arbitraje. A continuación, se comentan brevemente estos cuatro capítulos principales.

---

(1) Amy Endicott, integrante del *AI Guideline Drafting Committee* del CIArb, también integró el *Guidelines Review Committee* de la Guía sobre IA del SVAMC (SVAMC, 2024, p. 6).

(2) Si bien la Guía sobre IA del CIArb contiene una regulación más detallada que otros instrumentos similares en la materia, la Guía sobre IA del SVAMC ofrece una serie de comentarios de cada uno de sus artículos, que la Guía sobre IA del CIArb carece (SVAMC, 2024, pp. 13-20).

(3) Esto también se encuentra expresamente receptado en la Guía sobre IA del SVAMC (SVAMC, 2024, p. 8).

(4) La Guía sobre IA del CIArb, siguiendo la definición de la Recomendación del Consejo de la OCDE sobre Inteligencia Artificial, define a la IA como "un sistema basado en máquinas que, con objetivos explícitos o implícitos, deduce a partir de la información que recibe cómo generar resultados tales como predicciones, contenidos, recomendaciones o decisiones que pueden influir en entornos físicos o virtuales. Los diferentes sistemas de IA varían en sus niveles de autonomía y adaptabilidad tras su implementación". (CIArb, 2025, p.2). Ello se diferencia de la definición autónoma o autárquica incluida en la Guía sobre IA del SVAMC (SVAMC, 2024, p. 8).

2.1. Beneficios y riesgos del uso de la IA en el arbitraje

En primer lugar, la Guía sobre IA del CIArb detalla los beneficios y riesgos de la IA en el arbitraje. Ello no se encuentra sistematizado en otros instrumentos similares, tales como la Guía sobre IA del SVAMC, que solo se refiere a los beneficios y riesgos de la IA en el arbitraje de un modo general en su introducción, pautas generales y comentarios incluidos al final (SVAMC, 2024, pp. 5, 9, 13-20).

En cuanto a los beneficios, la Guía sobre IA del CIArb destaca un impacto positivo en la eficiencia del proceso arbitral y en la reducción de sus costos mediante el uso de la IA en diversas tareas a lo largo del procedimiento arbitral: desde el análisis de grandes conjuntos de datos, hasta la transcripción de audiencias (CIArb, 2025, paras. 1.1-1.3, 1.5, 1.7). [5] La Guía sobre IA del CIArb también señala que el uso de la IA en el arbitraje puede asegurar una mayor igualdad entre las partes (CIArb, 2025, paras. 1.4, 1.6, 1.8-1.10). Este punto es especialmente destacado en la encuesta elaborada por White & Case y Queen Mary en el año 2025, en donde se dice que "[e]l acceso a la IA permite a los participantes con recursos desiguales ser más competitivos" y se advierte una "[p]ercepción de desventaja competitiva si no se utiliza la IA" (White & Case y Queen Mary University of London, 2025, p. 29).

En cuanto a los riesgos, la Guía sobre IA del CIArb señala que las *alucinaciones* de la IA, sus *sesgos* algorítmicos o su gran *opacidad* (CIArb, 2025, paras. 2.4, 2.6), entre otros, también pueden peligrar la ejecutabilidad de los laudos y generar implicaciones adversas sobre el debido proceso, el estado de derecho, la administración de justicia y la credibilidad y legitimidad del arbitraje (CIArb, 2025, para. 2.1). Estos riesgos no son meramente hipotéticos, sino que hoy operan como la principal causa que desalienta el uso de la IA en el arbitraje (White & Case y Queen Mary University of London, 2025, p. 30). Además, la Guía sobre IA del CIArb advierte sobre los riesgos adicionales que conlleva el uso de la IA en el arbitraje en materia de confidencialidad, integridad y seguridad de datos (CIArb, 2025, paras. 2.2-2.3), [6] y medioambiente (CIArb, 2025, para. 2.8).

---

(5) Según la última encuesta elaborada por White & Case y Queen Mary (2025), las dos principales razones que motivan el uso de IA en el arbitraje son precisamente su potencial para ahorrar tiempo (54%) y reducir costos (44%) en el procedimiento arbitral (White & Case y Queen Mary University of London, 2025, p. 29).

(6) Estos dos últimos riesgos generales también aparecen en la encuesta elaborada por White & Case y Queen Mary (2025) como la segunda y tercera razones principales que operan en contra del uso de la IA en el arbitraje (White & Case y Queen Mary University of London, 2025, p. 30).

2.2. Recomendaciones generales sobre el uso de la IA en el arbitraje

En segundo lugar, para asegurar los beneficios de la IA y mitigar los riesgos derivados de su uso en el arbitraje, la Guía sobre IA del CIArb contiene una serie de recomendaciones generales vinculadas al uso de la IA en el arbitraje por parte de todos los actores involucrados en él.

De entrada, la Guía sobre IA del CIArb invita a que se realicen consultas razonables sobre cualquier herramienta de IA que se fuera a utilizar en un arbitraje con el fin de comprender la tecnología, sus funciones y el tratamiento de los datos implicados, así como los potenciales riesgos asociados a su uso (CIArb, 2025, paras. 3.1-3.2). Del mismo modo, la Guía sobre IA del CIArb recomienda a las partes y los árbitros a realizar consultas razonables sobre cualquier ley, reglamento o norma relacionadas con la IA que puedan resultar aplicables en el arbitraje y eventualmente limitar su uso (CIArb, 2025, para. 3.3).

A su vez, la Guía sobre IA del CIArb dispone que, en principio, el uso de IA por cualquier participante en el arbitraje "no disminuirá su responsabilidad y obligación de rendir cuentas" (CIArb, 2025, para. 3.4). En forma concordante, la Guía sobre IA del SVAMC prevé el deber de las partes de revisar los resultados que arroje la IA que utilicen en la preparación de sus presentaciones y verificar su precisión desde un punto de vista legal y fáctico, por lo que deben considerarse responsables a las "[p]artes y los representantes de las partes" con relación a "cualquier error o inexactitud no corregida en cualquier resultado producido por una herramienta de IA" (SVAMC, 2024, rec. 4).

2.3. Uso de la IA por las partes en el arbitraje

En tercer lugar, la Guía sobre IA del CIArb reconoce las facultades de los árbitros para regular el uso de la IA por las partes en el arbitraje, incluyendo subsecciones específicas vinculadas con la autonomía de las partes para acordar sobre el uso de IA en el arbitraje, así como algunos criterios para dirimir eventuales conflictos derivados de tal uso.

La IA es una realidad innegable en el arbitraje y, por ello, la Guía sobre IA del CIArb invita a los tribunales arbitrales a regular su uso desde el comienzo del arbitraje con el fin de preservar la integridad del procedimiento y garantizar la ejecutabilidad del laudo (CIArb, 2025, paras. 4.3-4.4, 5.3-5.4). [7] Como excepción, se dispone que los árbitros no podrán regular el uso privado de IA por las partes (por ejemplo, para su manejo documental

(7) En este contexto, la Guía sobre IA del CIArb alienta al tribunal a designar peritos en IA para comprender cualquier herramienta de IA o aspectos relacionados con la IA (CIArb, 2025, para. 4.2).

interno) en la medida en que dicho uso no interfiera con el procedimiento ni con su integridad y sea generalmente admitido en litigios ante los tribunales nacionales de la "jurisdicción correspondiente" (CIArb, 2025, para. 4.5). [8] También se destaca que los árbitros pueden adoptar cualquier medida para subsanar incumplimientos de las partes con relación al uso de la IA, tales como extraer inferencias negativas o tener en cuenta dicho incumplimiento al momento de asignar las costas (CIArb, 2025, para. 4.7).

A su vez, la Guía sobre IA del CIArb proporciona una serie de criterios relevantes para resolver cualquier discrepancia entre las partes con relación al uso de la IA (CIArb, 2025, paras. 6.1-6.8). A tal efecto, se señala que los árbitros podrán tener en cuenta los beneficios que implique la utilización de una herramienta de IA, tales como eventuales ahorros de costos y/o tiempo, así como los riesgos asociados a dicha herramienta de IA, incluyendo si pudiera tener un impacto en la prueba, la equidad y la confidencialidad del arbitraje, teniendo especialmente en cuenta las características de la IA en cuestión (tales como los datos que sustentan los resultados generados y la existencia de sesgos, así como la calidad, precisión y seguridad de dicha herramienta de IA) (CIArb, 2025, paras. 6.3-6.4).

Finalmente, la Guía sobre IA del CIArb regula el deber continuo de las partes de divulgar el uso de IA en el arbitraje cuando dicho uso "pudiera tener un impacto en la prueba, en el resultado del arbitraje o implicar, de otro modo, la delegación de un deber expreso hacia los árbitros o hacia cualquier otra parte" (CIArb, 2025, para. 7.1). Este enfoque adoptado por la Guía sobre IA del CIArb es opuesto al contemplado en la Guía sobre IA del SVAMC, que establece como principio general que "no es necesario revelar que se han utilizado herramientas de IA en relación con un arbitraje" (SVAMC, 2024, rec. 3).

2.4. Uso de la IA por los árbitros en el arbitraje

En cuarto lugar, la Guía sobre IA del CIArb regula el uso de la IA por parte del tribunal arbitral (CIArb, 2025, paras. 8.1-8.4, 9.1-9.2). Este es el uso más sensible de todos y, a la par, el que se espera que crezca con mayor vertiginosidad en los próximos años (White & Case y Queen Mary University of London, 2025, p. 33).

Para empezar, de modo similar al Reglamento de la UE [9] y a la Guía sobre IA del SVAMC (SVAMC, 2024, rec. 6), la Guía sobre IA del CIArb dispone que los árbitros no deben delegar su función decisoria en la IA, ni uti-

---

(8) Si bien aquí no se aclara cuál sería esta "jurisdicción correspondiente", más adelante la Guía sobre IA del CIArb llama la atención de "(i) la ley de la sede del arbitraje; (ii) la ley y las reglas que rigen el procedimiento, incluidas las reglas institucionales de arbitraje; (iii) las leyes nacionales de las partes; y (iv) cualquier norma ética aplicable" (CIArb, 2025, para. 6.7).

(9) *Ver* Introducción, *supra*.

lizarla de algún modo que pudiera afectar la integridad del procedimiento o la ejecutabilidad del laudo, como podría suceder si delegaran el "análisis jurídico, la investigación y la interpretación de hechos y del derecho, o la aplicación del derecho a los hechos" (CIArb, 2025, para. 8.2).

Esta preocupación también la comparten los usuarios del arbitraje. Según la encuesta realizada por White & Case y Queen Mary en 2025, la "opinión preponderante" es que "el razonamiento debe seguir siendo responsabilidad del árbitro" (White & Case y Queen Mary University of London, 2025, p. 32): el 77% se opuso a que la IA sea utilizada por los árbitros para redactar las partes de razonamiento jurídico de sus decisiones, mientras que un 69% se opuso a que se delegara el análisis del fondo o la exactitud de las alegaciones o pruebas de las partes. Sin perjuicio de ello, un importante porcentaje de los encuestados se mostró a favor de que los árbitros utilicen IA para el cálculo de daños, costos e intereses (77%), resumir escritos o documentos presentados por las partes (66%) y redactar órdenes procesales o partes no dispositivas del laudo (60%) (White & Case y Queen Mary University of London, 2025, p. 31).

Asimismo, la Guía sobre IA del CIArb dispone que los árbitros deberían consultar con las partes respecto del uso de cualquier herramienta de IA y darles a ellas la oportunidad de formular comentarios y oponerse a que dicha herramienta sea utilizada por los árbitros. Se agrega que, si las partes no estuvieran de acuerdo con el uso de IA por parte de los árbitros, estos deberían abstenerse de utilizar tal herramienta de IA (CIArb, 2025, para. 9.1). [(10)] Este enfoque también se diferencia con la Guía sobre IA del SVAMC, que solo prevé la divulgación del uso de la IA por parte de los árbitros cuando estos pretendan "basarse en información generada por IA que no figure en el expediente" (SVAMC, 2024, rec. 7).

3. Reflexiones finales

En un contexto de creciente uso y regulación en torno a la IA en el arbitraje, la Guía sobre IA del CIArb se presenta como un instrumento relativamente detallado y relevante para guiar el uso de la IA en el arbitraje. Además, la Guía sobre IA del CIArb invita a plantear diversas preguntas y repensar las nuevas regulaciones en la materia.

¿Es necesario regular el uso de la IA en el arbitraje? ¿Existen verdaderas lagunas que deben colmarse? En su caso, ¿quién o quiénes deberían encargarse de dicha regulación? ¿Los Estados, los actores privados y/u otras organizaciones mixtas? ¿Cómo debería regularse? ¿Mediante tratados y leyes?

(10) La Guía sobre IA del CIArb también invita a los árbitros a consultar con el resto de los miembros del tribunal respecto del uso de cualquier herramienta de IA (CIArb, 2025, para. 9.2).

¿O a través de un conjunto de recomendaciones o guías no vinculantes? Si siguiera este último camino, ¿qué grado de profundidad debería tener dicha regulación? En el marco de las instituciones arbitrales, ¿sería conveniente también modificar y adecuar los reglamentos arbitrales a estas nuevas realidades? ¿O aún es muy temprano para hacerlo? Las regulaciones y/o reglamentos arbitrales que establezcan reglas aplicables al uso de IA, ¿deberían ser obligatorias para las partes, árbitros y participantes del arbitraje en general? ¿O resulta conveniente establecer simplemente los marcos de referencia y que sean las partes quienes puedan escoger si los adoptan o no en los procesos arbitrales?

Bibliografía

Parlamento Europeo y Consejo de la Unión Europea. (2024, 13 de junio). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.° 300/2008, (UE) n.° 167/2013, (UE) N.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial). Diario Oficial de la Unión Europea, L 2024/1689. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32024R1689

Silicon Valley Arbitration & Mediation Center. (2024, abril). Guidelines on the Use of Artificial Intelligence in Arbitration (1ª ed.). https://svamc.org/wp-content/uploads/SVAMC-AI-Guidelines-First-Edition.pdf

Chartered Institute of Arbitrators. (2025, marzo). Guideline on the Use of AI in Arbitration (2025). https://www.ciarb.org/media/m5dl3pha/ciarb-guideline-on-the-use-of-ai-in-arbitration-2025-_final_march-2025.pdf

White & Case y Queen Mary University of London, School of International Arbitration. (2025). 2025 International Arbitration Survey – The path forward: Realities and opportunities in arbitration. https://www.whitecase.com/insight-our-thinking/2025-international-arbitration-survey

Ciberespacio en disputa: inteligencia artificial y el nuevo campo de batalla jurídico

Por Sol Aylén Borrelli Ozón [*]

Resumen: En el presente trabajo se llevará a cabo un análisis sobre la inteligencia artificial y su repercusión en ciberdelitos, la protección de datos personales y la privacidad. Tal como se ha visto en los últimos años, es tan relevante una guerra en el plano físico como en el ciberespacio. Antes se utilizaba gran parte de los recursos en defensa, comprando armamentos. Hoy las mejores "armas" son aquellas que puedan defender los derechos en el espacio cibernético. Anteriormente se construían murallas; ahora se utilizan sistemas que las imiten virtualmente. Se invita al lector a recorrer el artículo con una mirada crítica, abriendo un espacio de reflexión que trazara un camino para aprehender conceptos como inteligencia artificial, *phishing*, ciberdelitos, ciberespacio, *deepfakes*, para finalizar con un análisis del derecho comparado, nuestro Código Penal y su Proyecto.

Palabras clave: Inteligencia artificial; ciberdelitos; datos personales; ciberespacio; derecho.

Abstract: This work will carry out an analysis on artificial intelligence and its impact on cybercrimes, data protection and privacy. As it has become evident in the last few years, a war in the physical world is as relevant as a war in cyberspace. Some time ago, a large number of resources was used in Defense to buy armaments. Nowadays, the best "weapons" are those which may protect rights in the cyberspace. In other days, walls were built; today there are systems which imitate them in the virtual world. This article invites the reader to explore it with a critical eye, opening a space for reflection which sets a path to grasp concepts such as artificial intelligence, phishing, cybercrimes, cyberspace and deepfakes, only to conclude with an analysis on comparative law, our Criminal Code and its bill.

Keywords: Artificial intelligence, cybercrimes, data protection, cyberspace, law.

(*) Abogada recibida en la Universidad de Buenos Aires, con promedio 9.14, Premio a la Excelencia Académica y Distinción por estudiante destacada y premiada por entidades del exterior. Ayudante de la materia Inteligencia Artificial y Derecho, a cargo de Cecilia Danesi. Primer puesto en el VI Taller Red Jean Monnet sobre "Política Regional de Protección de Datos e Inteligencia Artificial" (Università degli Studi di Milano, Italia). Correo electrónico de contacto: solborrelliozon@gmail.com

1. INTRODUCCIÓN

Antes de introducirnos en el desarrollo, es menester definir conceptos primordiales que van a servir de base para todo el trabajo.

1.1. Inteligencia artificial: es definida por la Organización para la Cooperación y el Desarrollo Económicos (OCDE) como un sistema computacional que puede, para un determinado conjunto de objetivos definidos por humanos, hacer predicciones y recomendaciones o tomar decisiones que influyen en entornos reales o virtuales. Los sistemas de IA están diseñados para operar con distintos niveles de autonomía.

1.2. Ciberespacio: es un espacio virtual maleable, despersonalizado, anónimo, autónomo, de naturaleza no territorial y comunicativa.

1.3. Ciberdelito: se puede describir como delitos ciberdependientes y delitos facilitados mediante las tecnologías de información y comunicación. Requieren una infraestructura de TIC y, en general, se tipifican como la creación, distribución y ejecución de *malware* o *ransomware*, así como ataques a infraestructuras nacionales críticas o el acceso no autorizado a datos confidenciales de clientes.

1.4. Phishing: es un tipo de ciberataque que utiliza correos electrónicos, mensajes de texto, llamadas telefónicas o sitios web fraudulentos para engañar a las personas y hacer que comparta datos confidenciales, descargue *malware* o se expongan a la ciberdelincuencia.

Vale destacar que los ataques de *phishing* son una forma de ingeniería social. A diferencia de otros ciberataques que se dirigen directamente a redes y recursos, los ataques de ingeniería social utilizan errores humanos, historias falsas y estrategias de presión para manipular a las víctimas para que se perjudiquen a sí mismas o a sus organizaciones.

1.4.1 Malware: hace referencia a un *software* malicioso, un código de *software* o programa informático, escrito de manera intencional para dañar los sistemas informáticos o a sus usuarios. Actualmente, la mayoría de los ciberataques modernos implican algún tipo de *malware*.

1.4.2 Ransomware: Se trata de un *malware* que retiene como rehenes los datos o dispositivos de la víctima, exigiendo un pago para liberarlos y evitando su bloqueo o pérdida.

1.5. Deepfakes: son archivos de vídeo, imagen o voz manipulados mediante un *software* de inteligencia artificial de modo que parezcan originales, auténticos y reales.

El término *deepfakes* combina la palabra *fake* (falso, ya que este tipo de archivos siempre son falsos, aunque se haga todo lo posible con el fin de que parezcan reales) y la palabra *deep***,** proveniente de *deep learning* (aprendizaje profundo, que es un tipo de aprendizaje automático de la inteligencia artificial).

Utilizan aprendizaje automático de inteligencia artificial. Esta tecnología se basa en sofisticados algoritmos que son capaces de analizar si un archivo es real o si está alterado y, de esta forma, la IA puede ir mejorando cada vez más en la labor de falsificar de manera más fidedigna. Los *deepfakes* pueden ser generados directamente por *softwares* u ordenadores especializados en este aprendizaje automático, sin necesidad de intervención humana.

2. Del campo de batalla físico al digital: las armas invisibles de ciberdelitos y ciberguerras

Los ciberdelitos mencionados no solo quedan en el plano teórico. Un caso para ejemplificarlo es la guerra entre Rusia y Ucrania. Pese a que tomó impacto en el aspecto físico, ¿qué ocurre con el ciberespacio?

En febrero del año 2022, se descubrió IsaacWiper, un virus que destruye datos y ataca a organizaciones de Ucrania, y el "gusano" HermeticWizards, que esparce el *malware* en las redes de ese país. En dicho mes, hubo un aumento de ataques de *phishing* en su ciberespacio: las herramientas de acceso remoto (RAT) y el *malware* de robo de contraseñas, se enviaron como archivos adjuntos en los correos electrónicos de *phishing* que se relacionaban con facturas y pagos. Meses después, el equipo de respuesta de emergencia del gobierno ucraniano informó un ataque cibernético en las instalaciones de energía por el *malware* Industroyer2 y CaddyWiper para interrumpir el suministro eléctrico.

En relación con ello, los militares ucranianos también fueron víctimas del *phishing* para obtener acceso a sus cuentas personales y desinformar publicando vídeos falsos (como en el que ondean una bandera blanca de rendición). Como respuesta, lanzaron ataques de denegación de servicio (DDoS) que saturan de tráfico las páginas web logrando su colapso.

Esto podría ser incluso más difícil de entender y detectar cuando ni siquiera se sabe quién es el que ataca: el caso de Anonymous con Rusia [(1)].

(1) Anonymous es un colectivo de *hackers* organizado de manera informal y descentralizada, que pretende proteger a Ucrania contra los ciberataques rusos. Por su estructura, es difícil identificar la autenticidad de sus ataques aunque, desde el inicio de la guerra, piratas informáticos vinculados con aquel la han asumido.

3. LA GUERRA POR LOS DATOS: PROTECCIÓN, PRIVACIDAD Y PODER

En una sociedad que le interesa lo intangible, los datos constituyen información y esta, a su vez, es mercancía de intercambio y una forma de acumular poder, tal como se observó en el caso ya mencionado con la privacidad de los ciudadanos, incluyendo difusión de datos privados y confidenciales. En relación con ello, según un informe de la empresa de seguridad Surfshark, Rusia ha sido el país con más robos de datos durante 2022, afectando a más de 3,5 millones de usuarios. El 11% de sus cuentas sufrieron filtración de información sensible (como tarjetas de crédito, números de teléfono y contraseñas). Asimismo, desde febrero a marzo de dicho año, se vulneraron un 136% más de cuentas rusas y Anonymous admitió haber robado 495.000 correos electrónicos secretos de la empresa Technotec, la cual brinda servicios a empresas energéticas de Rusia.

Frente a este escenario de vulneración masiva a la protección de datos y a la privacidad, los ordenamientos jurídicos han intentado dar respuesta. En Europa, se ha avanzado con el Reglamento General de Protección de Datos de la Unión Europea. Asimismo, se regula el derecho de intimidad y privacidad en el art. 8° de Convenio Europeo de Derechos Humanos y el 7° y 8° de la Carta de los Derechos Fundamentales de la Unión Europea, además de diversos instrumentos legislativos regionales [2]. En Argentina, se encuentra protección en la Constitución Nacional en los artículos 18, 19, 33, 43 y el Código Civil y Comercial de la Nación en el artículo 1770 regula el derecho a la intimidad. Por otro lado, hay una regulación específica en la materia: la Ley 25.326 de Protección de los Datos Personales, la cual pretende ser modificada por un Proyecto de Reforma.

Sin embargo, la brecha entre normativa y realidad se hace evidente en determinados casos. ¿Qué delito se persigue hoy como una estafa patrimonial y, al mismo tiempo, compromete datos personales y la privacidad? El *phishing* representa, justamente, esa doble amenaza, que expone la insuficiencia del marco legal vigente. Bajo el Código Penal actual encuadraría en el inc. 16 del art. 173, es decir, dentro del Capítulo "Estafas y otras defraudaciones". Como correlato, el bien jurídico de la estafa (y del *phishing*) es el patrimonio, y el ardid o engaño serían los medios. Dicho ello, es menester preguntarse: ¿qué ocurre en los casos de simple obtención de datos para defraudar? Justamente, no habría delito por falta de perjuicio patrimonial. Esto apareja desprotección e injusticias. Por ello, una de las mejores formas de superarlo es legislando al *phishing* como delito autónomo, tal como lo

(2) Como la Directiva 2002/58/CE y el Reglamento (CE) 45/2001.

hace el Proyecto de Reforma del Código Penal en su artículo 491 [(3)], al penar la mera obtención de datos personales mediante engaños.

Ahora bien, ¿qué ocurre cuando no solo están en juego el patrimonio, sino la identidad (digital) misma? Puntualmente, ¿en dónde encuadra el robo de la identidad (por ejemplo si se la sustraen al CEO de una empresa)?: en el artículo 71 quinquies del Código Contravencional de la Ciudad Autónoma de Buenos Aires, pero no en el Código Penal, por lo que no es considerado un delito. La discusión para incorporar un tipo penal que lo castigue reside en si esa suplantación debe tener un fin específico (comisión de otro delito o causar daño) o si la mera utilización del nombre e imagen de una persona alcanza. El artículo 139 *ter* del proyecto de reforma se inclinó por la primera alternativa. Mientras ello no se apruebe, solo se podría iniciar una acción civil (artículo 52 del Código Civil y Comercial de la Nación); realizar una denuncia por uso de una marca registrada o una designación comercial perteneciente a un tercero sin su autorización (artículo 31 de la ley 22.362); solicitar la baja del perfil o realizar denuncia por infracción al art. 71 *quinquies*.

Dicha deficiencia legislativa también se puede observar en el caso de los *deepfakes*:

Actualmente, no hay una figura en el Código Penal de Argentina que penalice la generación y difusión de *deepfakes* de contenido sexual. Parte de la doctrina busca incluir estos casos dentro del artículo 128, que sanciona la tenencia, producción y distribución de material de abuso sexual infantil. No obstante, dicha norma no incluye imágenes generadas por IA ya que si no expone imágenes reales no encuadraría en ese artículo, lo que dificulta lograr condenas. Sin embargo, podría reclamarse daños y perjuicios o acciones penales por delitos contra el honor.

En este escenario, legislaciones como la Ley Olimpia, orientada a combatir la violencia digital contra las mujeres, han resultado herramientas valiosas, ya que permiten solicitar la remoción del contenido en plataformas digitales y asegurar la preservación de pruebas para eventuales acciones judiciales. Sin embargo, la protección se aplicaría únicamente a víctimas de género femenino, excluyendo a los varones adolescentes. En este último caso, podrían interponer una medida cautelar preventiva con el fin de solicitar el retiro del contenido.

---

(3) "Se impondrá prisión de SEIS (6) meses a DOS (2) años o SEIS (6) a VEINTICUATRO (24) días-multa, al que ilegítimamente con ánimo de lucro o la finalidad de cometer un delito, y valiéndose de alguna manipulación informática, ardid o engaño, obtuviere claves o datos personales, financieros o confidenciales de un tercero, siempre que el hecho no constituya un delito más severamente penado (…)".

4. LA CARRERA LEGAL CONTRA LOS *DEEPFAKES*: UN PANORAMA GLOBAL

Como se analizó *ut supra*, se ha concluido que Argentina carece de un marco normativo específico sobre los *deepfakes*; en su lugar, existe un evidente vacío legal. Ante ello, resulta imprescindible examinar cómo han regulado esta materia otros países.

En Inglaterra y Gales, el gobierno incluirá un nuevo delito por la creación de imágenes *deepfake* sexualmente explícita, como así también por compartirlas. En relación con ello, los autores de estos delitos se enfrentarán a hasta dos años de cárcel por tomar una imagen íntima sin consentimiento. Cabe resaltar que esta normativa sería un avance ya que la Ley de Seguridad en Línea de 2023 del Reino Unido no penaliza la creación de dichas imágenes sin consentimiento.

Por otro lado, en cuanto a la Unión Europea, la Ley de Servicios Digitales (Digital Services Act o DSA), en vigor desde el año 2024, está destinada a "prevenir actividades ilegales y dañinas en línea y la difusión de desinformación". En ese sentido, la UE ha intensificado el control sobre los proveedores de servicios digitales.

En la misma línea, el gobierno español aprueba tipificar como delito los *deepfakes* de contenido sexual. Se trata de una iniciativa pionera en Europa, ya que el proyecto prevé la reforma del Código Penal para penalizarlo.

Dinamarca, por su parte, incorpora disposiciones en el marco de su normativa sobre *deepfakes* como parte de su ley de derechos de autor digital, con el fin de resguardar los derechos de las personas frente al impacto de los *deepfakes* generados por IA. Esta regulación daría lugar a que quien se encuentre afectado pueda reclamar su eliminación y a los artistas una compensación por el uso de su imagen sin autorización, incluso hasta 50 años después de su fallecimiento.

Cruzando al continente americano, Estados Unidos, en fecha 19 de mayo de 2025, sancionó la *Take It Down Act* (TIDA), que fija un régimen punitivo por la difusión sin consentimiento de representaciones visuales íntimas, comprendiendo las generadas o alteradas a través de inteligencia artificial.

5. REFLEXIONES FINALES

El análisis sobre los ciberdelitos, la guerra por los datos y los *deepfakes* demuestran el desafío sobre cómo resguardar y garantizar la privacidad sin detener la innovación tecnológica. Las amenazas y peligros que antes surgían en espacios físicos, hoy se trasladan al ciberespacio, muchas veces in-

visibles pero no por ello menos riesgosos, con aptitud para afectar identidades digitales, patrimonios, derechos personales y personalísimos, e incluso, la estabilidad de Estados.

En este contexto, la ausencia de marcos normativos específicos en Argentina pone de manifiesto la necesidad de adaptar el derecho penal y el civil frente a las nuevas tecnologías, especialmente aquellas vinculadas a la inteligencia artificial. Si bien, como se analizó, se podría aplicar normativa de manera análoga, el vacío legal genera desprotección e inseguridad jurídica. En contraposición, el derecho comparado demuestra un camino esperanzador a la hora de regular figuras. Una vez más, el derecho debe adaptarse a la realidad y a las necesidades de la sociedad, donde la IA sea una herramienta al servicio de las personas y no un arma invisible.

En definitiva, entre la utopía de una privacidad protegida y la distopía de un control absoluto, el derecho tiene un rol decisivo: elegir de qué lado de la muralla digital (y de la inteligencia artificial) quiere estar.

Bibliografía

Addati, F. A. (2025). Privacidad bajo amenaza digital: la respuesta estadounidense a los Deepfakes y el vacío legal en Argentina. *Ratio Iuris. Revista de Derecho UCES, 13*(1), 197-209

Avast. (2022). *Incremento en ataques de phishing dirigidos a la infraestructura en Ucrania*. https://blog.avast.com/es/phishing-attacks-in-ukraine

Cabrol, M., González, N., Pombo, C., & Sánchez, R. (2020). *Adopción y ética responsable de la inteligencia artificial en América Latina y el Caribe*. Nota Técnica del BID.

De la Fuente, J. E., & Donna, E. A. (2004). Aspectos generales del tipo penal de estafa. *Revista Latinoamericana del Derecho*, 41–42.

DW Español. (2022). *Anonymous: ciberataques contra Rusia*. https://www.youtube.com/watch?v=BBZ3ZhXzRg0

El Periódico. (2022). *Rusia, el país que sufre más robos de datos*. https://www.elperiodico.com/es/internacional/20220413/rusia-ciberataques-guerra-ucrania-anonymous-13516966

Faliero, J. C. (2018). *El futuro de la regulación en protección de datos personales en la Argentina* (p. 2). Buenos Aires: Thomson Reuters.

Gov.uk. (2025). *Government crackdown on explicit deepfakes*. https://www.gov.uk/government/news/government-crackdown-on-explicit-deepfakes

HB. (2022). *En el centro de la guerra cibernética. Gosspetssvyaz participó por primera vez en la legendaria conferencia de ciberseguridad Black Hat*. https://

biz.nv.ua/tech/gosspecsvyazi-pouchastvovala-v-black-hat-novosti-ukrainy-50264313.html

IBM. (s. f.). *¿Qué es el malware?*. https://www.ibm.com/es-es/think/topics/malware

IBM. (s. f.). *¿Qué es el phishing?*. https://www.ibm.com/es-es/think/topics/phishing

IBM. (s. f.). *¿Qué es el ransomware?*. https://www.ibm.com/es-es/think/topics/ransomware

Imbrogno, A. (2004). Internet: camino hacia un derecho globalizado. *Revista electrónica de la Facultad de Derecho UNICEN*, 82. Tandil.

iProUP. (2022). *La 'otra' guerra Rusia-Ucrania: ¿Cuál es el papel de los hackers en el 'conflicto digital'?*. https://www.iproup.com/economia-digital/29865-guerra-rusia-ucrania-el-papel-de-los-hackers-en-el-conflicto

Lisa Institute. (s. f.). *Deepfakes: Qué es, tipos, riesgos y amenazas*. https://www.lisainstitute.com/blogs/blog/deepfakes-tipos-consejos-riesgos-amenazas?srsltid=AfmBOormKACCb8dLgvdBoZRnIksT1DC59P3VBaSCQJ9lI-DP3qwaBFDqb

Ministerio de la Presidencia, Justicia y Relaciones con las Cortes, Gobierno de España. (2025). *El Gobierno aprueba tipificar como delito los deepfakes de contenido sexual y el grooming*. https://www.mpr.gob.es/prencom/notas/paginas/2025/250325-gobierno-delito-deepfakes-grooming.aspx

Ministerio Público de la Defensa Argentina. (2023) *Ley Olimpia: una nueva herramienta contra la violencia de género*. https://www.mpd.gov.ar/index.php/noticias-feed/6804-ley-olimpia-una-nueva-herramienta-contra-la-violencia-de-genero

Pilnik, F. (2021). *Comentarios sobre la suplantación de identidad digital*. https://www.saij.gob.ar/comentarios-sobre-suplantacion-identidad-digital-comentarios-sobre-suplantacion-identidad-digital-nv32574-2021-12-28/123456789-0abc-475-23ti-lpssedadevon?&o=137&f=Total%7CFecha%7CEstado%20de%20Vigencia%5B5%2C1%5D%7CTema%5B5%2C1%5D%7COrganismo%5B5%2C1%5D%7CAutor%5B5%2C1%5D%7CJurisdicci%F3n%5B5%2C1%5D%7CTribunal%5B5%2C1%5D%7CPublicaci%F3n/Novedad%7CColecci%F3n%20tem%E1tica%5B5%2C1%5D%7CTipo%20de%20Documento/Doctrina&t=718

Queen Mary University of London. (2025). *Deepfakes and the law: Why Britain needs stronger protections against technology-facilitated abuse*. https://www.qmul.ac.uk/law/news/2025/items/deepfakes-and-the-law-why-britain-needs-stronger-protections-against-technology-facilitated-abuse.html

Schurjin Almenar, D. (2022). Delitos informáticos en Argentina: Normativa actual y posibilidades de cambio según el proyecto de nuevo Código Penal. *Revista de Pensamiento Penal*, 18. https://bit.ly/3SzjJFn

Sternik, I. (2025). *Deepfakes y adolescentes: existe un vacío legal sobre los videos con contenido sexual creados con IA y su difusión en redes*. Chequeado. https://chequeado.com/el-explicador/deepfakes-y-adolescentes-existe-un-vacio-legal-sobre-los-videos-con-contenido-sexual-creados-con-ia-y-su-difusion-en-redes/

Surfshark. (s.f.). *Global data breach statistics*. https://surfshark.com/research/data-breach-monitoring

UdeSA. (s.f.). *El derecho y la irrupción de los deepfakes*. https://udesa.edu.ar/noticias/el-derecho-y-la-irrupcion-de-los-deepfakes

Véliz, C. (2020). La privacidad es poder. *Revista Universitaria de Cultura*, 28. Madrid.

World Economic Forum. (2025). *Legislación sobre deepfakes: Dinamarca avanza para proteger la identidad digital*. https://es.weforum.org/stories/2025/08/legislacion-sobre-deepfakes-dinamarca-avanza-para-proteger-la-identidad-digital/#:~:text=La%20enmienda%20danesa%20actualmente%20en,de%20la%20muerte%20del%20artista

IA, niñez y adolescencia: garantizando derechos en un nuevo entorno

Por Mercedes Romera [(*)] *y María Rosa Avila* [(**)]

Resumen: El artículo analiza el impacto de la inteligencia artificial (IA) en la vida de niñas, niños y adolescentes, destacando tanto sus oportunidades como los riesgos asociados a la vulneración de sus derechos. Señala que la IA está presente en múltiples dispositivos y plataformas digitales, transformando la educación, la comunicación, la salud y la participación social de la infancia. Sin embargo, su uso también conlleva desafíos como brechas de acceso, riesgos de privacidad, ciberagresiones, sesgos algorítmicos, violencia digital y exposición a contenidos nocivos. El texto subraya la importancia de un enfoque de derechos, basado en la Convención sobre los Derechos del Niño y la Observación General 25, para garantizar la igualdad, la no discriminación, el interés superior del niño, la protección integral y la participación de la infancia en el entorno digital. Finalmente, se enfatiza la necesidad de políticas públicas, regulación, formación, acompañamiento familiar y mecanismos de control a empresas y proveedores de IA para promover ciudadanías digitales seguras, inclusivas y con perspectiva de género y diversidad.

Palabras clave: Inteligencia artificial; infancia; adolescencia; derechos de los niños; ciudadanía digital.

Abstract: The article analyzes the impact of artificial intelligence (AI) on the lives of children and adolescents, highlighting both its opportunities and the risks related to the violation of their rights. AI is present in multi-

---

(*) Magíster en Diseño y Gestión de Políticas Sociales (FLACSO Argentina). Especialista en Políticas Públicas de Educación, Derechos Humanos, Infancia y Juventud. Docente de Sociología UBA. Se ha desempeñado en su campo de especialización, en distintas funciones en el Estado, en el Ejecutivo y en el Legislativo.

(**) Maestranda en Docencia Universitaria UBA. Doble titulación Especialista en la Enseñanza por Medios Tecnológicos. Profesora adjunta, Derecho UBA. Abogada. Investigadora Instituto de Investigaciones Jurídicas y Sociales "Ambrosio L. Gioja" UBA. Especializada en Comunicación y Psicología Social. Consultora Monitoreo de Políticas Públicas con enfoque de género y enfoque de derechos. Miembro del Grupo de Referentes Globales de la Iniciativa Spotlight, Naciones Unidas. Conferencista y activista. Defensora de DDHH.

ple devices and digital platforms, transforming education, communication, health, and social participation for children. However, its use also poses challenges such as access gaps, privacy risks, cyberbullying, algorithmic biases, digital violence, and exposure to harmful content. The text emphasizes the importance of a rights-based approach, grounded in the Convention on the Rights of the Child and General Comment 25, to ensure equality, non-discrimination, the best interests of the child, comprehensive protection, and meaningful participation in the digital environment. Finally, it stresses the need for public policies, regulation, education, family guidance, and control mechanisms for companies and AI providers to foster safe, inclusive digital citizenship with a gender and diversity perspective.

Keywords: Artificial intelligence; childhood; adolescence; children's rights; digital citizenship.

> "En la vida de los emperadores hay un momento que sucede al orgullo por la amplitud desmesurada de los territorios que hemos conquistado, a la melancolía y al alivio de saber que pronto renunciaremos a conocerlos y a comprenderlos, es el momento desesperado en que se descubre que ese imperio que nos había parecido la suma de todas las maravillas es una destrucción sin fin ni forma".
>
> Italo Calvino, 1972.

1. Introducción

La inteligencia artificial (IA) está transformando las vidas de niñas, niños y adolescentes. Todos los días millones de dispositivos tecnológicos combinados con IA son utilizados en diferentes formatos como televisión, internet, tutoriales, redes sociales, celulares, videojuegos, plataformas colaborativas, realidad virtual y aumentada, robótica, sistemas automatizados, algoritmos y análisis de datos, biometría y tecnología de implantes, entre otros. Las formas de obtener información, procesar datos e interactuar con entornos digitales son presentadas cada vez más de manera accesible y de fácil uso. Las nuevas tecnologías de la información y la comunicación (TICS) asociadas a IA ofrecen a las niñas, niños y adolescentes numerosas oportunidades para ejercer y reclamar su ciudadanía, pero también los expone a nuevos riesgos de vulneración de sus derechos. En contextos de crecientes desigualdades sociales y de fragmentación de la niñez, es urgente establecer mecanismos de protección que aseguren el acceso efectivo a una ciudadanía digital segura y con enfoque de derechos para todos las niñas, niños y adolescentes.

2. Inteligencia artificial: una nueva racionalidad. ¿Qué medios para qué fines?

Existen diferentes acepciones del término *inteligencia artificial,* pero es posible identificar algunos consensos acerca de los rasgos que lo definen.

Estos son: que se trata de "sistemas basados en máquinas que, a partir de los objetivos definidos previamente por sus programadores y de la información que reciben, puede hacer predicciones, producir contenidos, formular recomendaciones o tomar decisiones" (Comité de Inteligencia Artificial -CAI-, 2024). Según esta definición, los sistemas de IA interactúan con las personas de forma directa o indirecta y pueden adaptar su comportamiento aprendiendo sobre el contexto, siguiendo reglas y ejemplos, o por ensayo y error. Basados en técnicas de aprendizaje automático, pueden reconocer patrones en los datos y a partir de allí, procesar textos, voz, imágenes o videos, y producir nueva información. En línea con esta acepción, la UNESCO (2019) definió la IA como "máquinas capaces de imitar ciertas funcionalidades de la inteligencia humana incluyendo la percepción, el aprendizaje, el razonamiento, la resolución de problemas, la interacción del lenguaje e incluso la producción creativa".

Pero la IA hace referencia también a un *principio técnico universal* que se aplica cada vez más a todos los segmentos de la vida individual y colectiva, abarca al conjunto de las relaciones sociales y está destinado a revolucionar nuestras existencias [(1)]. Desde esta perspectiva, la IA comprende un fenómeno de automatización e hiper-racionalización de la experiencia humana, construye subjetividades, crea realidad y enuncia verdades. Tiene implicaciones generalizadas de amplio alcance en las vidas de todas las personas, particularmente de niñas, niños y adolescentes. Está presente en diversas aplicaciones y plataformas digitales de uso diario, en juegos en línea, teléfonos celulares inteligentes, herramientas de traducción automática, reconocimiento facial como factor de seguridad en dispositivos, aplicaciones para automatización de tareas repetitivas como *chatbots*, creación o edición de textos, imágenes, voces, videos, e incluso música, entre otras funciones.

Cada vez más, los sistemas de IA se encuentran dotados de la tecnología necesaria para evaluar lo real a velocidad superior a la capacidad humana de hacer lo propio, nos sugieren, recomiendan e inducen a actuar de determinada manera. Su capacidad de procesamiento y precisión nos asombra y maravilla. Dada la complejidad de las sociedades actuales no parece posible gestionar la gran cantidad de información disponible ni la realización de ciertas tareas y rutinas sin algún tipo de sistema de decisión automatizada. En ello radica precisamente el poder de la IA, en la dependencia que crea y en la dificultad que conlleva advertir la condición artificial de su inteligencia. A pesar de que se nos presente con apariencia de neutralidad e independencia en su funcionamiento, no es autónoma, ni racional, ni capaz de pensar por sí misma. Es un producto social y está al servicio de los objetivos que le son fijados. Su funcionamiento depende de los datos con que opera, de la lógica del ecosistema dentro del que ha sido programada y, funda-

(1) Éric Sadin (2020) utiliza dicha expresión para dar cuenta de la potencia que tiene la inteligencia artificial en tanto *emergencia de una tecnología de lo integral*, que todo la abarca (pg. 20).

mentalmente, de los usos que se hagan de ella y de las regulaciones que se le impongan. Por otra parte, los sistemas de IA se desarrollan en el marco de relaciones sociales que involucran dimensiones de poder y económicas, inscriptas en un mercado y su alcance no puede prescindir de dicha lógica.

Estamos subrayando el *carácter instrumental de los algoritmos*, los modos de producción social de la IA y la ineludible responsabilidad de los Estados en asegurar que sus objetivos sean compatibles con los derechos humanos y la democracia (Innerarity, 2024). Es evidente que, cada vez más, las sociedades requieren de una gran movilización de conocimiento, instrumentos de obtención de información, análisis de datos más certeros, mejores tecnologías y resolución de problemas complejos. Pero la definición acerca del alcance y el sentido de la automatización de la vida no puede quedar en manos de los algoritmos ni de las empresas que los desarrollan.

Las niños, niños y adolescentes representan más de una cuarta parte de la población mundial, constituyen el grupo más numeroso entre quienes acceden a internet y son quienes experimentan un nivel cada vez mayor de exposición a los sistemas de IA. La forma en la que acceden y reciben información sobre el mundo, así como la forma en que se vinculan socialmente está cambiando. Necesitamos analizar los modos en que la IA impacta y transforma sus vidas, entender cómo influye en sus percepciones y en su desarrollo. Pero también necesitamos conocer cómo son diseñados los sistemas de IA y sus algoritmos, qué datos y fuentes de información utilizan, cómo jerarquizan y priorizan los contenidos, con qué sesgos se procesa la información, de qué modo se protege la privacidad de los datos personales de quienes interactúan con ellos. Se requiere, en definitiva, ejercer una verdadera gobernanza acerca del uso y el desarrollo de la IA, asegurando la necesaria protección contra su utilización indebida y adecuándola a los marcos normativos de los derechos humanos. Máxime cuando lo que está en juego es respetar, proteger y hacer efectivos los derechos de la población infantil en el entorno digital.

3. Ciudadanía infantil e inteligencia artificial: una brújula para encontrar nuevas coordenadas

> "En el camino hacia lo nuevo hay que avanzar la mayoría de las veces, aunque no siempre, paso a paso. No todo es posible y realizable en cualquier momento; la falta de condiciones no solo retarda, sino que cierra el camino. Todo es posible allí donde las condiciones existen parcialmente suficientes, pero también es verdad, que todo es fácticamente imposible allí donde las condiciones no existen en absoluto".
>
> Ernst Bloch, 1977.

La Convención sobre los Derechos del Niño (en adelante, "la Convención") constituye el instrumento de derechos humanos específico para la infancia, una carta magna de reconocimiento y de respeto de su ciudadanía

que establece el marco en el que deben inscribirse las prácticas y las políticas de los Estados que la han ratificado [2]. La adopción de sus principios supone su aplicación en complementariedad con otros pactos y tratados internacionales cuyos principios fundamentales son comunes, interdependientes, se interrelacionan y se refuerzan mutuamente [3]. En términos conceptuales la Convención universaliza la consideración de las niñas, niños y adolescentes como sujetos con derechos propios y con la garantía de protección especial por parte de los Estados cuya obligación es respetar, promover, proteger y asegurar su condición ciudadana (Van Hoof, en Abramovich, 2006). De este modo, en materia de políticas públicas destinadas a la niñez, la adopción de la Convención, supone diseñar programas, dispositivos y recursos con enfoque de derechos, es decir, establecer una correspondencia entre políticas y derechos, de acuerdo con los nuevos estándares fijados por los instrumentos internacionales. Los niños y niñas destinatarios de las políticas son titulares de derechos que generan obligaciones al Estado y requieren mecanismos para hacerlos exigibles y asegurar su cumplimiento. Se establece de este modo una vinculación estrecha entre los derechos reconocidos en los tratados y pactos internacionales, las medidas positivas que se requieren para que su vigencia sea efectiva (conjunto de políticas públicas a cargo del Poder Ejecutivo o del Legislativo) y la exigibilidad legal para asegurar su cumplimiento.

Esto significa que, en materia de políticas, entornos digitales y sistemas de IA los Estados deben aplicar los principios establecidos por la Convención y adoptar medidas especiales para respetar, proteger y hacer efectivos los derechos de la niñez. Tomando como referencia la Observación General 25 relativa a los derechos de los niños en relación con el entorno digital (CRC/C/GC/25, año 2021) [4], los cuatro principios descritos a continuación proporcionan una guía para garantizar el pleno ejercicio de la ciudadanía infantil en relación con las tecnologías en general y la IA en particular.

---

(2) Firmada el 20 de noviembre de 1989 en la sede de la Organización de las Naciones Unidas, la Convención sobre los Derechos del Niño se destaca por ser el instrumento de derechos humanos más aceptado a nivel mundial, 196 países la han ratificado. Entre ellos, todas las naciones de América Latina y el Caribe.

(3) Interdependencia con los tratados de Derechos Humanos tales como: La Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (CEDAW), el Pacto Internacional de Derechos Económicos, Sociales y Culturales (PIDESC), La Convención sobre los Derechos de las Personas con Discapacidad (CDPD), el Pacto Internacional de Derechos Civiles y Políticos (PIDCP), el Convenio sobre la prohibición de las peores formas de trabajo infantil, el Convenio 169 de la Organización Internacional del Trabajo (OIT). A nivel regional el *Protocolo de San Salvador,* se suma a la Convención Americana (Pacto de San José), la Convención Interamericana para prevenir, sancionar y erradicar la violencia contra la mujer, -Convención de Belem Do Para- junto a la Convención sobre los Derechos del Niño (CDN) conformando un marco regional de protección de la niñez en el continente americano.

(4) Según dicha Observación el entorno digital abarca el conjunto de tecnologías de la información y las comunicaciones (TICs) incluidas: las redes, los contenidos, los servicios y aplicaciones digitales, los dispositivos, la realidad virtual y aumentada, la robótica, la inteligencia artificial, los algoritmos de análisis de datos, la biometría, entre otros.

A. *Principio de igualdad y no discriminación:* implica que los Estados aseguren el acceso equitativo y efectivo de todas las niñas, niños y adolescentes al entorno digital, evitando exclusiones y adoptando las medidas necesarias para su utilización informada. Los Estados deben implementar medidas específicas para prevenir discriminación por motivos de sexo, discapacidad, situación socioeconómica, origen étnico o nacional, idioma o cualquier otro motivo o condición. Así como también cerrar las brechas digitales en sus diferentes formas: de clase, de género, territoriales, de inclusión y accesibilidad, de lenguas, etc.

B. *Interés superior:* los Estados deben asegurar que, en todas las actuaciones relativas al suministro, la regulación, el diseño, la gestión y la utilización del entorno digital, el interés superior de todas las niñas y niños sea una consideración primordial. Asimismo deben tener en cuenta todos los derechos de los niños, incluidos su derecho a buscar, recibir y difundir información, a recibir protección contra todo daño y a que sus opiniones se tengan debidamente en cuenta. Deben asimismo garantizar la transparencia e implicación en lo referido a la evaluación del interés superior del niño como principio transversal y estructurante y de los criterios aplicados al respecto.

C. *Derecho a la vida, a la supervivencia y al desarrollo:* los Estados deben adoptar todas las medidas para proteger a las y los niños frente a todo lo que constituya una amenaza para su derecho a la vida, la supervivencia y el desarrollo. Los riesgos relacionados con los contenidos, los contactos, las conductas y los contratos en ese ámbito abarcan, entre otras cosas, los contenidos violentos y sexuales, la ciber-agresión y el acoso, los juegos de azar, la explotación y el maltrato, incluidos la explotación y los abusos sexuales, la violencia de género y la promoción del suicidio o de actividades que pongan en peligro la vida o la incitación a estos por parte, entre otros, de delincuentes o grupos armados identificados como terroristas o extremistas violentos.

D. Respeto de las opiniones: se alienta a los Estados a que utilicen el entorno digital para consultar a los niños, niñas y adolescentes sobre medidas legislativas, administrativas y de otra índole pertinentes y velen por que se tengan en cuenta significativamente sus opiniones y su participación. Los Estados deben promover la concienciación sobre los medios digitales, su potencial y sus riesgos, así como ofrecer capacitación y apoyo a fin de que estos participen en igualdad de condiciones, garantizando procesos de consulta sin exclusiones de aquellos niños, niñas y adolescentes que no tienen acceso a la tecnología o que carecen de las aptitudes necesarias para utilizarla.

A partir de estos principios, queda claro que los Estados deben tomar medidas generales para asegurar su aplicación en pos de garantizar el efectivo ejercicio de derechos de la niñez en relación con los entornos digitales y los sistemas de IA. Entre dichas medidas señalamos las más urgentes:

- Adecuar la legislación vigente para garantizar entornos digitales y sistemas de IA compatibles con los derechos previstos en la Convención. Seguimiento y monitoreo de las innovaciones tecnológicas y de su posible afectación o daños a la infancia y adolescencia.

- Asegurar la creación de contenidos digitales diseñados y adaptados a la edad de los niños y destinados a potenciar su desarrollo integral y lograr que estos tengan acceso a una amplia diversidad de información, incluida la de los propios organismos públicos, sobre cultura, deportes, artes, salud, asuntos civiles y políticos y derechos de ciudadanía infantil.

- Velar por el acceso de todos las niñas, niños y adolescentes a información en línea de calidad, accesible y diversa.

- Incorporar políticas específicas de protección de la infancia en entornos digitales y sistemas de IA, como la ciber-agresión y la explotación y los abusos sexuales en línea contra niños facilitados por la tecnología digital, asegurarse de que se investiguen esos delitos en la justicia y ofrecer reparación y apoyo a quienes sean víctimas de esos actos.

- Asignar los recursos públicos necesarios para la aplicación de leyes, políticas y programas que aseguren el pleno ejercicio de derechos a las niñas, niños y adolescentes en entornos digitales y frente a sistemas de IA, promoviendo la igualdad de acceso, la conectividad y su asequibilidad por medio de una adecuada infraestructura pública.

- Garantizar la producción de datos e investigación. Realizar diagnósticos, evaluar efectos y medir el impacto de las políticas.

- Difundir información, realizar campañas de concientización y promover programas educativos destinados a la formación de competencias y de la ciudadanía digital infantil, dirigidos a la población de niñas, niños adolescentes, a sus entornos familiares, a las personas adultas cuidadoras, y a la comunidad en general.

- Asegurar la participación de la sociedad civil y de las propios niñas y niños en la promoción y protección de los derechos de los niños en relación con los entornos digitales y los sistemas de IA.

- Adoptar medidas para asegurar el respeto de los derechos de la niñez relacionados con IA por parte del sector empresarial. Los Estados deben cerciorarse de que las empresas cumplan sus obligaciones, construyan diseños adecuados, impidan que sus redes o servicios de IA vulneren derechos de niñas, niños y adolescentes a la privacidad, a la protección y a la integridad. Además, deben alentar a las empresas a proporcionar información pública y asesoramiento accesible y oportuno para apoyar la participación de los niños en actividades digitales y sistemas de IA seguros.

- Asegurar el acceso a la justicia y reparación de derechos de niñas, niños y adolescentes víctimas de delitos vinculados con los entornos digitales y la IA. La reparación puede incluir la restitución, la compensación y la satisfacción, puede requerir una disculpa, una corrección, la eliminación de contenidos ilícitos y el acceso a servicios de recuperación psicológica, entre otras medidas, ya que el efecto de las tecnologías impacta de forma desproporcionada en algunos grupos y de modo amplificado sobre la vida de niñas niños.

- Atender la relevancia del ámbito digital configurando un subsistema de protección integral para el desarrollo de las niñas niños y adolescentes, garantizando el ejercicio de sus derechos digitales, el derecho al olvido, a la salud mental, a la educación y al respeto de los valores propios de la democracia sin discursos de odio.

4. DERECHOS DE LA INFANCIA: USOS Y DESUSOS DE LOS SISTEMAS DE IA

> "Adam empezó a hablar con el chatbot, que funciona con inteligencia artificial, a finales de noviembre, sobre sentirse emocionalmente apagado y sin verle sentido a la vida.
> ChatGPT le respondió con palabras de empatía, apoyo y esperanza, y lo animó a pensar en las cosas que sí fueran significativas para él.
> Pero en enero, cuando Adam pidió información sobre métodos específicos de suicidio,
> ChatGPT se la proporcionó. Al poco tiempo Adam se quitó la vida".
>
> The New York Times, Agosto de 2025.

La interacción de niñas, niños y adolescentes con los sistemas de inteligencia artificial es compleja y requiere de consideraciones específicas dados los efectos o repercusiones que estos pueden tener en el desarrollo de sus capacidades físicas, cognitivas, emocionales y psicológicas, en su integridad, en las violaciones a su privacidad, en la vulneración de sus derechos. El potencial que los sistemas de IA tienen para la automatización de algunas tareas, así como para la producción, el procesamiento y la disponibilidad de la información de manera veloz y cada vez más amigable para las personas en general y para niñas, niños, niñas y adolescentes en particular, representa una oportunidad en materia de implementación de políticas con enfoque de derechos para la infancia y adolescencia.

Entre las oportunidades que la IA puede ofrecer para el ejercicio y ampliación de derechos de la ciudadanía infantil, se destacan algunas soluciones innovadoras en materia de educación, salud, ecología y urbanismo. En el ámbito del desarrollo educativo, diversas experiencias con utilización de IA ofrecen herramientas colaborativas de aprendizaje personalizado, como tutorías inteligentes, plataformas interactivas de resolución de problemas,

entornos de aprendizaje exploratorio y dispositivos de aprendizaje automatizado, entre otras, que contribuyen a reforzar los procesos de enseñanza-aprendizaje de todos los niños, niñas y adolescentes y favorecen que cada estudiante siga su propio ritmo. También resulta de interés la aplicación de IA para la capacitación docente y para la gestión educativa. En relación con esta última, se destacan los sistemas de información de acceso a la educación y la utilización, por ejemplo, de tableros visuales. Tal es el caso de Argentina, que desde el año 2022, en los establecimientos educativos de la provincia de Mendoza recurren a la IA para detectar estudiantes que se encuentran en situación de abandono prematuro de la escuela [5].

En el ámbito de la salud, los sistemas de IA pueden mejorar la precisión en el diagnóstico y tratamiento de enfermedades, la detección de brotes epidemiológicos, el desarrollo de dispositivos específicos para niñas, niños y adolescentes con discapacidades, así como también para la gestión de la información sobre datos de pacientes y mejoras en los sistemas de acceso a la salud. Otro uso potencial de la IA por parte de la niñez es como herramienta de acceso en línea a la información, a la comunicación, a la libertad de expresión. Esto incluye la libertad de buscar, recibir y difundir información, expresar ideas, opiniones y puntos de vista y reclamar derechos.

Pero, a la vez, la falta de aplicación o la inobservancia de los derechos de los niños, niñas y adolescentes en los usos de los entornos digitales y los sistemas de IA, constituyen un verdadero riesgo para el desarrollo, integridad y seguridad de la población infantil que exigen ser abordados de manera urgente por los Estados, las empresas y por la comunidad en general.

Nos estamos refiriendo a una diversidad de riesgos de vulneración de sus derechos de distinto tipo. Entre otros mencionamos:

- La creciente brecha digital de acceso, uso y aprovechamiento de la IA por parte de niñas, niños y adolescentes, especialmente en contextos de desigualdad social, aumento de la pobreza y de la vulnerabilidad.

- La falta de enfoque de derechos y perspectiva de género y diversidad en la programación, el desarrollo y la utilización de los sistemas de IA abarca tanto el diseño de los algoritmos como todo el funcionamiento de su ecosistema.

- La falta de regulación normativa, vigilancia y gobernanza de los sistemas de IA por parte de los Estados.

---

(5) El sistema, concebido por el Laboratorio de Inteligencia Artificial Aplicada de la Universidad de Buenos Aires, envía una alerta si se detecta una deserción y busca implementar una acción para tratar de mitigarlo. En Mendoza la prueba es universal, es decir, afecta a todos los colegios secundarios de la provincia. Ver más en: El Correo de la UNESCO, octubre-diciembre 2023.

- La falta de prevención y protección contra contenidos nocivos y perjudiciales para niñas y niños que promueven estereotipos de género, violencia de género digital, discriminación, racismo, violencia y maltrato, acoso y hostigamiento, pornografía y explotación o la circulación de relatos falsos, información errónea o desinformación.

- La presencia de importantes sesgos en los algoritmos, irracionalidades en los procesamientos de la información, inconsistencias y especulaciones en las respuestas de los sistemas de IA.

- La falta de controles a los proveedores de servicios de IA utilizados por niñas, niños y adolescentes sobre la aplicación de un correcto etiquetado de contenidos, que cumpla con las normas de filtrado y moderado escolares pertinentes. Así como también de las advertencias sobre posibles usos indebidos.

- La falta de protección a la libertad de expresión y contra ciber-agresiones, amenazas, censuras, filtraciones de datos y vigilancia digital sobre niñas y niños.

- La amenaza a la privacidad, dignidad, seguridad e integridad de niñas, niños y adolescentes en los sistemas de IA. La falta de cuidados en la utilización, el procesamiento y la circulación de sus datos privados. Las injerencias arbitrarias o ilegales en su derecho a la intimidad y en su propio consentimiento informado (acorde a su desarrollo madurativo).

- Mayor exposición a las diferentes formas de violencia, explotación y abusos sexuales o a ser influidos para hacerse daño a sí mismos o a otros. Insuficientes medidas de prevención, seguridad y protección de derechos especialmente de la población infantil.

- La relevancia de la falta de acompañamiento, orientación y apoyo a las familias y personas cuidadoras sobre sus responsabilidades parentales en el entorno digital y el cuidado requerido para el ejercicio de los derechos digitales de niños, niñas y adolescentes.

- Falta de medidas especiales de protección en los entornos digitales y sistemas de IA a grupos específicos de niños, niñas y adolescentes que puedan estar más expuestos a peligros. Entre ellos mencionamos, niños con discapacidad, con padecimientos de salud mental, niñas y adolescentes mujeres, diversidades sexuales, minorías étnicas, raciales, religiosas o políticas.

- Necesidad de mayor regulación y aplicación de filtros de acceso de niñas, niños y adolescentes a juegos y apuestas en línea oficiales, mayores y mejores mecanismos de persecución de juegos y de apuestas ilegales en línea a las que acceden niñas y niños.

- Falta de la adecuada atención a los derechos de niñas niños y adolescentes en el diseño de programas y algoritmos.

5. Zona de promesas y retos: hacia una IA con enfoque de derechos, perspectiva de género y diversidad

Los entornos digitales y la IA están transformando las vidas de las infancias. Su uso seguro y protegido constituye un requisito de acceso a la ciudadanía de niños, niñas y adolescentes y una obligación que los Estados deben garantizar. Es urgente y prioritario establecer mecanismos de cuidado y prevención de la vulneración de derechos de la población infantil en los espacios en los que ellos habitan. En contextos de aumento de las desigualdades sociales y de fragmentación de la niñez, las viejas formas de violencia se entrelazan con nuevos tipos de producción de violencia en entornos digitales y sistemas de IA [6]. Los modos de vinculación de niñas, niños y adolescentes en los ámbitos digitales, sus efectos, riesgos, características y alcances deben ser analizados y abordados en su transversalidad e interseccionalidad [7] con el conjunto de derechos de la población infantil y adolescente.

Para que las infancias y adolescencias puedan aprovechar al máximo la potencia y las ventajas que ofrecen los entornos digitales y la IA para su desarrollo integral y su bienestar, es preciso construir ciudadanías digitales con enfoque de derechos de niñas, niños y adolescentes, con perspectiva de género, diversidad e inclusión en clave interseccional.

Los Estados tienen una responsabilidad indelegable en dicha tarea, así como en garantizar condiciones materiales para el ejercicio de esas ciudadanías, así como las empresas deben considerar en sus desarrollos los riesgos a que pueden estar expuestas las infancias y adolescencias y desarrollar mecanismos y alertas de cuidado y prevención. El sistema de protección integral tiene que constituirse en dos subsistemas, el primer subsistema desde lo territorial físico y material y otro subsistema especializado para el entorno digital.

Es urgente promover colectivamente en este entorno digital y de IA el fortalecimiento de los lazos sociales, la preservación de lo común y la pre-

---

(6) Nos referimos a los tipos de violencia según la clasificación de UNICEF (2023) que incluye las formas derivadas del adultocentrismo (crianza, simbólica, entre pares, adultista), el patriarcado (género, sexual, uniones tempranas, noviazgos violentos, machismo), pobreza infantil (económica, monetaria y no monetaria) y entornos (urbana, digital, ambiental).

(7) Interseccionalidad: este concepto fue acuñado en 1989 por Kimberlé Williams Crenshaw, académica y profesora estadounidense especializada en el campo de la teoría crítica de la raza. Lo define como "el fenómeno por el cual cada individuo sufre opresión u ostenta privilegio en base a su pertenencia a múltiples categorías sociales". La interseccionalidad pone de manifiesto cómo las diferentes categorías sociales generan opresiones y privilegios muy dispares al entrecruzarse entre ellas, es decir, no se trata de capas de discriminación que se superponen o suman (género + raza + ruralidad etc.), sino que su entrecruzamiento genera dinámicas diversas que multiplican la discriminación. Se trata así de un concepto fundamental para el análisis tanto de la situación de las mujeres y la desigualdad de género en el mundo, como de cualquier análisis relacionado con los derechos humanos. Fuente "Derechos de las niñas para un futuro en Igualdad: Renovando compromisos en América Latina y el Caribe" . INFORME UNICEF ALyC. Copyright © 2021 UNICEF

valencia de los derechos humanos. Las niñas niños y adolescentes tienen derecho al disfrute y ejercicio del presente como sujetos sociales que gozan de una protección especial, y sostener aún en contextos de incertidumbre y de vertiginosos avances tecnológicos, las políticas públicas, que procuren asegurar condiciones y capacidades del derecho al futuro.

Bibliografía

Abramovich, Víctor (2006). "Una aproximación al enfoque de derechos en las estrategias y políticas de desarrollo", en Revista de la CEPAL, Nº 88, pp. 35-50.

Bloch, Ernst (1977). *El Principio Esperanza*. Biblioteca Filosófica Aguilar: Madrid.

Calvino, Italo (1972). *Las ciudades invisibles*. Editorial Minotauro. Argentina.

Comité de los Derechos del Niño. Observación General 25 (2021). Relativa a los derechos de los niños en relación con el entorno digital. CRC/C/GC/25, 2 de marzo de 2021.

Comité de Inteligencia Artificial -CAI- (2024). Convenio Marco sobre inteligencia artificial, derechos humanos, democracia y estado de derecho. Consejo de Europa.

El Correo de la UNESCO (octubre-diciembre 2023). En Argentina un algoritmo contra la deserción escolar. Natalia Páez. Revista. UNESCO, 4 de octubre de 2023.

Innerarity, Daniel (2024). *Inteligencia Artificial y Democracia*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), Montevideo, Uruguay. Disponible en:

https://unesdoc.unesco.org/ark:/48223/pf0000389736_spa

Sadin, Éric (2020). *La inteligencia artificial o el desafío del siglo. Anatomía de un antihumanismo radical*. 1a ed. - Ciudad Autónoma de Buenos Aires: Caja Negra, 2020 328 p.; 20 x 13 cm. - (Futuros próximos; 31). Traducción de Margarita Martínez ISBN 978-987-1622-86-3. Disponible en:

https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-43602021000100318

The New York Times (2025). Un adolescente tenía tendencias suicidas. ChatGPT fue el amigo en quien confió. Kashmir Hill, 27 de agosto de 2025.

UNESCO (2021). *Inteligencia artificial y educación. Guía para las personas a cargo de formular políticas*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, Francia. Disponible en:

https://unesdoc.unesco.org/ark:/48223/pf0000379376

UNESCO (2024). *De qué hablamos cuando hablamos de Inteligencia Artificial*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cul-

tura, Francia. Oficina Regional Montevideo. Consejo Latinoamericano de Ciencias Sociales (CLACSO).

UNICEF (2021). *Orientación de políticas sobre el uso de la inteligencia artificial en favor de la infancia*. Fondo de las Naciones Unidas para la Infancia (UNICEF) Finlandia, noviembre de 2021. Disponible en:

https://www.unicef.org/innocenti/media/1351/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021_ES.pdf

UNICEF (2023). *Qué decimos cuando hablamos de violencia contra las niñas, niños y adolescente*. Fondo de las Naciones Unidas para la Infancia (UNICEF). Primera edición abril de 2023. Disponible en:

https://www.unicef.org/argentina/media/18461/file/Qu%C3%A9%20decimos%20cuando%20hablamos%20de%20violencia%20contra%20las%20ni%C3%B1as,%20ni%C3%B1os%20y%20adolescentes..pdf

Las decisiones de consumo desafiadas (y potenciadas) por la inteligencia artificial y el contexto digital

Por Demetrio Alejandro Chamatropulos (*)

Resumen: El trabajo examina la interacción entre la inteligencia artificial (IA) y las decisiones de consumo en el contexto digital, destacando tanto los riesgos como las oportunidades que esta relación implica para el derecho del consumidor. En primer lugar, se presenta un recorrido histórico que rescata hitos poco visibilizados de la IA, desde los aportes de Pitágoras, Téano y Ramón Llull, hasta figuras como Ada Lovelace, Alan Turing y Grace Hopper, para mostrar que los actuales debates sobre algoritmos, automatización y razonamiento artificial tienen raíces antiguas y multidisciplinarias. En segundo lugar, el artículo analiza críticamente cómo la IA puede "potenciar" o "empeorar" las decisiones de los consumidores en entornos digitales. Se destacan dos fenómenos problemáticos: el "filtro burbuja", que personaliza y restringe la información disponible para el consumidor afectando su autonomía y el deber de información, y la "paradoja de la elección", donde la sobreabundancia de opciones puede conducir a la parálisis decisoria y disminuir la satisfacción del usuario. El autor propone reflexionar sobre cómo el diseño algorítmico, la regulación y el deber de información deben adaptarse para garantizar que la IA sirva al interés del consumidor y no lo distorsione.

Palabras clave: Inteligencia artificial; derecho del consumidor; filtro burbuja; paradoja de la elección; algoritmo.

(*) Doctor en Derecho (UBA) —calificación sobresaliente, con tesis sobre Neurociencias, Ciencias del Comportamiento, Filosofía, Cronobiología, Inteligencia Artificial y Derecho—. Posdoctor en Derecho (UNT). Magíster en Derecho del Trabajo y Relaciones Laborales Internacionales (UNTREF) —tesis aprobada con recomendación de publicación—. Magíster en Derecho Empresario (ESEADE) —tesis calificada como sobresaliente cum laude con recomendación de publicación—. Abogado (UNT). Profesor encargado de la materia de grado "Neurociencias, Inteligencia Artificial y Derecho" (UNT). Director del Suplemento periódico "Innovación & Derecho" del Diario La Ley. Profesor de posgrado en distintas universidades argentinas y del exterior. Director de distintos posgrados en diferentes universidades argentinas. Autor de varios libros (entre ellos *Estatuto del Consumidor Comentado*, tres ediciones, en tres tomos, *Los Daños Punitivos en la Argentina* y *El Deber de Información frente a las decisiones no racionales de los consumidores*). Miembro de la Academia de Ciencias Morales, Políticas y Jurídicas de Tucumán. Actual relator de la Corte Suprema de Justicia de la Provincia de Tucumán.

Abstract: This paper examines the interaction between artificial intelligence (AI) and consumer decision-making in digital environments, highlighting both the risks and opportunities that this relationship poses for consumer law. First, it offers a historical overview that recovers often-overlooked milestones in the development of AI—from the contributions of Pythagoras, Theano, and Ramon Llull to figures such as Ada Lovelace, Alan Turing, and Grace Hopper—showing that today's debates on algorithms, automation, and artificial reasoning have deep and multidisciplinary roots. Second, the article critically analyzes how AI can either enhance or hinder consumer decision-making online. Two problematic phenomena are emphasized: the "filter bubble," which personalizes and restricts the information available to consumers, affecting both their autonomy and the duty to inform, and the "paradox of choice", where the overabundance of options can lead to decision paralysis and reduce user satisfaction. The author argues that algorithmic design, regulation, and the duty to inform must be adapted to ensure that AI serves consumer interests rather than distorting them.

Keywords: Artificial Intelligence; consumer law; filter bubble; paradox of choice; algorithm.

1. Introducción

Este breve trabajo se propone, en primer lugar, visibilizar y poner en valor algunos nombres y momentos de la historia de la inteligencia artificial (IA) que, ante el incesante bombardeo de "noticias de último momento" sobre el tema que se experimenta en la actualidad, suelen ser pasados de largo incluso por muchos tecnólogos y personas marcadamente especialistas en el tema.

El mundo de lo legal en particular no resulta ajeno a lo descripto, por lo que resulta vital que los operadores jurídicos tomen nota de lo reseñado pues pueden extraer de allí aportes de valor que hoy en día quizás desconozcan. Se aclara que dicha síntesis ha llegado hasta mediados del siglo XX pues la historia desde allí en adelante es más que conocida. Por otro lado, luego de ello, se dejarán esbozadas algunas ideas "en fresco" y "no del todo masticadas" que se disparan como consecuencia de esta interacción entre IA y actuación de los consumidores en los entornos digitales.

2. Antecedentes e hitos (no del todo visibles) en la historia de la inteligencia artificial

Así como las personas siempre buscaron crear máquinas que reemplacen su fuerza física, también desde la antigüedad se intentó idear dispositivos que piensen por ellas o que, por lo menos, colaboren en dicha tarea.

El ábaco en Babilonia, inventado entre los siglos IV a VI a.C., constituye uno de los primeros instrumentos que se conocen para realizar operaciones aritméticas simples. Dispositivos similares fueron hallados en el mundo oriental como el soroban japonés. Estas herramientas mantuvieron vigencia durante siglos a causa de su utilidad práctica [1].

Ya en los tiempos presocráticos, Pitágoras (569 a.C.-475 a.C.) y sus seguidores pusieron de resalto algo que aún hoy puede parecer contraintuitivo: la ausencia de diferencias entre números y cosas (principio de semejanza formal). Dicho en otras palabras, las cosas se pueden reducir a números. Todo tiene una regularidad matemática. Este (en apariencia) simple razonamiento (que tiene en parte carácter simbólico) ha tenido un impacto enorme en la innovación científica y, lógicamente, estimula a que los desarrolladores de IA busquen objetivos de lo más diversos sin pensar en límites.

Cabe remarcar no obstante que la idea expuesta ha sido matizada por una de las primeras filósofas de la historia de las que se tenga registro: Téano de Crotona, que vivió en el siglo VI y fue precisamente esposa de Pitágoras. Esta pensadora razonó que no es que todas las cosas se reducían al número, sino que en el número residía el orden esencial, por lo que todo debía corresponderse con él.

Para Platón (427 a.C.-347 a.C.) las matemáticas eran tan importantes en la construcción del saber que siempre se ha recordado que en la puerta de su célebre academia se leía la siguiente leyenda: "Está prohibida la entrada a las personas que no sepan geometría".

La lógica aristotélica con sus desarrollos secuenciales a partir de premisas también fueron obviamente un punto de apoyo para que a mediados del siglo XX los ordenadores comiencen a "razonar". Su aporte fue una señal de largada para comenzar a "codificar".

Un discípulo de Arquímedes, Ctesibio (285 a.C.-222 a.C.), inventó lo que se conoce como la primera máquina de la historia que se autocontrolaba. Se trataba de un reloj de agua (la clepsidra) utilizado para regular y mantener a niveles constantes un flujo de agua.

En 1900 fue hallado el llamado "Mecanismo de Anticitera" en Grecia. Se cree que fue construido entre los siglos II y III a.C. Su calidad y complejidad permitiría razonar que no fue un invento disruptivo, sino que se desprendería de la evolución de la tecnología griega que aún no se llegó a descubrir hasta el momento. Se trata de una especie de "computadora embrionaria"

---

(1) Desde lo anecdótico puede recordarse que en el año 1946 un oficinista de dicha nacionalidad le ganó una competición a un estadounidense experto en el uso del mejor dispositivo eléctrico de cálculo existente en dicho momento. WILLIAMS, Trevor, *Historia de la tecnología*, 9ª edición, Siglo XXI editores, Buenos Aires, 2006, tomo 5, p. 498.

que fue útil para realizar predicciones en materia astronómica, entre otras funciones.

Carnéades (214 a.C.-129 a.C.) destacó una noción fundamental para la posterior dinámica de la IA: su lógica de funcionamiento a través de criterios de probabilidad. Señalaba que la imposibilidad de acceder a lo verdadero llevaba a que haya que contentarse simplemente con criterios de aproximación o probabilidad ("probablemente verdadero"). Esto se relaciona con la lógica bayesiana que luego se verá. De este modo, mientras más elementos concordantes existan esa probabilidad aumenta, terminándose de potenciar cuando todo ello se evalúa con un método analítico.

Al matemático Al-Juarismi (780-850) se le debe el comienzo del uso de la palabra clave en materia de IA: el "algoritmo". Ello en tanto la versión en latín del nombre de este autor es Algorithmi pues una de sus obras se llamó "Algoritmi de numero Indorum".

Más adelante en el tiempo, la lógica experimentó nuevos aportes y precisiones. Sobre todo, durante la Edad Media. Puede situarse en ese momento el caso de Ramón Llull (1232-1316) con su intento de construir una "Máquina lógica" o "Máquina pensante" (llamada Ars Magna). Se trataba de un dispositivo mecánico muy sofisticado mediante el cual Llull se proponía demostrar en cada caso la verdad o no de un postulado.

En el siglo XVII René Descartes (1596-1650) estructuró el razonamiento en clave matemática. También dejó sus célebres coordenadas como legado. Esto también debe ser tomado como una de las bases del edificio a construir. Cabe recordar además la concepción de este filósofo que consideraba a los animales y al cuerpo humano mismo como máquinas, aunque recalcó la diferenciación entre lo físico y la mente, asumiendo así una posición dualista.

No se puede olvidar tampoco que Thomas Hobbes (1588-1679) consideró a las personas como "máquinas de carne y hueso que suman y restan de manera silenciosa en su pensamiento", ni a Julien Offray de La Mettrie (1709-1751) quien llegó a titular a una famosa obra suya El hombre máquina.

La creación de los logaritmos por parte de John Napier (1550-1617) fue fundamental para desarrollos posteriores. Lo señalado sentó las bases para la invención de la primera calculadora mecánica en 1623 por Wilhelm Schickard (1592-1635) y de la primera regla deslizante, llevada a cabo en 1624 por William Oughtred (1574-1660). No puede dejar de mencionarse que antes de eso Leonardo da Vinci (1452-1519), entre sus múltiples creaciones, diseñó (no se sabe en qué año) una "máquina de sumar", que fue el primer intento (inconcluso) que se conoció de un dispositivo para ejecutar dicha tarea.

Cabe aludir aquí a la "Pascalina", calculadora mecánica ideada en 1645 por Blaise Pascal (1623-1662) y que funcionaba sobre engranajes y ruedas. Constituye para muchos el primer antecedente de la computadora. Claramente permitió que un trabajo mental se pueda reemplazar por uno de carácter ciertamente automatizado.

Puede recordarse además que Isaac Newton (1642-1727) con sus espectaculares contribuciones en materia de cálculo, abrió un campo de desarrollo prácticamente infinito en tiempos actuales que no hubiese sido posible sin sus trabajos fundacionales.

Claro está que la obtención primaria de ese logro fue siempre discutida con Gottfried Leibniz (1646-1716), siendo este último además para muchos el creador del sistema binario (aunque existen antecedentes previos a su nacimiento). Leibniz fue asimismo el inventor de la primera calculadora de propósitos generales.

Los aportes realizados por distintos pensadores en materia de probabilidad tienen un impacto singular en los desarrollos de IA contemporáneos. Sin perjuicio de lo advertido por Carnéades en la Antigüedad y que se mencionó más arriba, se destaca especialmente aquí la contribución de Thomas Bayes (1702-1761), quien puso el foco en observar cómo se iban determinando las probabilidades a partir de resultados concretos que se producían momento a momento.

El famoso teorema bayesiano se enfocó en el carácter condicionado o provisorio de una probabilidad, la cual requería actualizarse cuando surgían nuevas evidencias que obligasen a realizar la corrección. El monitoreo permanente de efectos permitiría así una mejor aproximación o precisión de la medición de las probabilidades de las causas que los generaban.

Thomas Reid (1710-1796) se interesó mucho por reconstruir la historia de una mente. Este aspiracional integra los objetivos de muchos desarrollos de "cerebros artificiales", en cuanto a la trazabilidad de su proceso de formación, las asociaciones que pueden llevar adelante sus algoritmos, etc. que podrían dar lugar eventualmente a esa anhelada reconstrucción.

Un hito que pasa a veces inadvertido es la creación en 1777 del denominado "Demostrador lógico" de Charles Stanhope (1753-1816) que llegó a permitir la realización de cálculos de probabilidades (aunque bastante básicos), desarrollo que es hoy uno de los grandes contenidos del mundo de la IA y sus sistemas predictivos.

Joseph Marie Jacquard (1752-1834) es citado pues en 1801 logró programar mediante un mecanismo de tarjetas perforadas un telar que permitía tejer seda de manera automática con programación previa. Ello facilitaba acelerar enormemente el ritmo de trabajo en la fabricación de indumentaria

y, a su vez, habilitó que la labor pueda ser llevada adelante por personas sin experiencia en el oficio.

Charles Babbage (1791-1871), influenciado por el sistema de tarjetas del telar de Jacquard, desarrolló su célebre "Máquina analítica". Se trataba de un dispositivo que tenía como norte ser programado para realizar los más diversos tipos de cálculo. Esta tarea le tomó alrededor de diez años (desde 1833 hasta 1842) aunque quedó inconclusa. Por sus peculiares características muchos sostienen que la referida máquina fue la primera computadora de la historia. Una de sus grandes virtudes fue la multitarea, en tanto podía pasar de una a otra en poco tiempo.

Ada Lovelace (1815-1852) es una gran figura y pionera en el estudio de los algoritmos y un verdadero hito en materia de programación. Ello surge a partir de sus notas sobre la máquina analítica de Charles Babbage (con quien colaboró). Allí ideó diversos programas que podían potenciar sus capacidades. Más precisamente, su nota "G" constituyó el algoritmo que le valió el título de primera persona programadora de la historia.

Sus inquietudes fueron de vanguardia al iniciar con una frase suya una polémica que aún hoy se sostiene. Así, dijo ya en aquellos años que "las máquinas no podían pensar por sí mismas". También puso el acento en la multiplicidad de tareas que podía realizar un dispositivo, superando largamente la mera realización de cálculos matemáticos. Llegó a razonar que las máquinas podían jugar al ajedrez o componer música, dando a entender que cualquier dato fáctico (como un sonido) podía pasar al mundo digital bajo la guía de lo que hoy se conoce como un ordenador.

George Boole (1815-1864) ideó un sistema algebraico que habilitó el razonamiento lógico de las máquinas, en tanto creó reglas que permitían la reducción y simplificación de situaciones problemáticas a dos estados: verdadero o falso (sistema binario de "ceros" y "unos"). El "Piano lógico" de 1869, creado por William Stanley Jevons (1835-1882), constituyó la primera máquina que tomó en cuenta el modelo de Boole.

Ernst Mach (1838-1916) con su noción de "Economía del pensamiento" expuso una forma de adquirir conocimiento que experimentó algunas similitudes con ciertos mecanismos de aprendizaje de la IA, al explicar que los problemas aparecen cuando el conocimiento disponible no permite hacer frente a novedades que trae el entorno y genera, como primera respuesta, la formulación de distintas hipótesis que serán sometidas a prueba hasta que se dé con la correcta, permitiendo esa nueva adaptación e incremento de la experiencia.

Gottlob Frege (1848-1925), precursor de la Lógica Matemática moderna, sentó las bases para los desarrollos en materia de programación ocurridos a partir del siglo XX. Asimismo, realizaron también contribuciones de valor

Georg Cantor (1845-1918), Giuseppe Peano (1858-1932), Alfred Whitehead (1861-1947), David Hilbert (1862-1943), Bertrand Russell (1872-1970), como, asimismo, otros matemáticos con fuerte influencia en lo que luego serían las Ciencias de la Computación como Emil León Post (1897-1954), Haskell Curry (1900-1982), Alfred Tarski (1901-1983), Alonzo Churh (1903-1995), Kurt Gödel (1906-1978) y Stephen Kleene (1909-1994).

Leonardo Torres Quevedo (1852-1936) fue uno de los primeros que examinó la posibilidad de que las máquinas piensen. No suele ser tan citado en las reseñas referidas a la IA, pero su célebre escrito del año 1913 titulado "Ensayos sobre la automática" (2) profundizó sobre el tema y constituye un trabajo de recuerdo obligado.

Hay quienes han puesto de resalto que Sigmund Freud (1856-1939) ha ejercido también una influencia decisiva para lo que luego de conoció como IA, ya que concibió a la actividad mental a partir de distintos componentes. De este modo, el conjunto conformado por el "ello", el "yo" y el "superyó", al actuar en paralelo e influenciarse recíprocamente, está muy presente en los aportes teóricos realizados muchos años después por Marvin Minsky (1927-2016), quien señala a Freud como "el primer buen teórico de la IA", siendo ese incluso el título de un capítulo suyo en una obra colectiva (3).

En materia de procesamiento de datos se destaca la figura de Herman Hollerith (1860-1929) que, al crear una máquina de tarjetas perforadas de tabulación, dio el puntapié inicial al tratamiento automatizado de volúmenes muy grandes de información.

Warren McCulloch (1898-1969) y Walter Pitts (1923-1969) escribieron en 1943 un famoso *paper* (4) en donde se puso sobre la mesa por primera vez la posibilidad de crear neuronas artificiales, siendo esta una idea que fue tomada luego por los distintos desarrollos de IA hasta la actualidad. Hoy en día su creación se conoce como la "Neurona de McCulloch-Pitts", un modelo-base que busca simplificar y copiar de algún modo el comportamiento de una neurona humana para luego generar redes de comunicación entre ellas.

Vannevar Bush (1890-1974), al idear el "Memex" fue de los primeros en buscar conexiones estrechas entre las personas y las máquinas, tendencia que luego profundizaría especialmente Joseph Licklider (1915-1990). En un

---

(2) TORRES QUEVEDO, Leonardo, "Ensayos sobre Automática. Su definición, extensión teórica de sus aplicaciones", Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales de Madrid, vol. 12, 1914, pp. 391-419.

(3) MINSKY, Marvin, "Why Freud Was the First Good AI Theorist", en MORE, Max – VITA-MORE, Natasha (ed.), *The transhumanist reader: classical and contemporary essays on the science, technology, and philosophy of the human future*, Wiley Blackwell, New Jersey, 2013, pp. 167 y ss.

(4) McCULLOCH, Warren S. - PITTS, Walter, "A logical calculus of the ideas immanent in nervous activity", Bulletin of Mathematical Biophysics, vol. 5, 1943, pp. 115-133.

famoso trabajo de 1945 consideraba al dispositivo a inventar como un íntimo y ampliado complemento de la memoria humana y como un gran organizador del conocimiento [5].

Norbert Wiener (1894-1964) publicó en 1948 un libro que fue muy leído en su época: La cibernética. Este matemático fue el iniciador de lo que se denomina "Teoría del control". La cibernética de Wiener buscaba aportar elementos que evidencien que los mecanismos de funcionamiento de los seres vivos y de las máquinas no diferían demasiado. La noción misma de ciborg se inspira en él, en tanto se estaría ante individuos "mejorados" a partir de incorporaciones tecnológicas.

Además de programar una máquina para jugar al ajedrez, Claude Shannon (1916-2001), junto a su esposa Betty Shannon (1922-2017) idearon en 1950 un experimento en el cual un ratón magnético llamado "Teseo" (controlado por un dispositivo electromecánico) debía encontrar la salida de un laberinto, para lo cual recurría a la experiencia previa o "aprendizaje" que iba adquiriendo por el método del ensayo y error. El ratón iba registrando lo que sucedía con cada movimiento y si ello era o no exitoso para encontrar la salida. Se trató del primer experimento de gran impacto sobre inteligencia artificial.

El desarrollo espectacular de la IA ha puesto a la figura de Alan Turing (1912-1954) en el centro indiscutido de la materia. En 1936 sentó las bases teóricas de lo que sería la computadora universal, capaz de llevar adelante funciones de lo más diversas. Ese aporte fue conocido como "la Máquina de Turing" [6]. Sin embargo, su aporte más popular sobre IA es el famoso "Test de Turing", realizado en 1950 y mediante el cual intentó fijar las bases o lo que se requería para concluir que una máquina era inteligente (usando siempre como referencia o elemento de confronte a la inteligencia humana). En dicho trabajo buscó especialmente neutralizar la llamada "objeción de Ada Lovelace" [7].

Grace Murray Hopper (1906-1992) se destacó porque desarrolló el primer compilador informático en el año 1952. A su vez, propició la autonomía de los lenguajes de programación con respecto a las máquinas, lo que permitió la creación del todavía hoy útil lenguaje "COBOL" (Common

---

(5) BUSH, Vannevar, "As We May Think", The Atlantic Monthly, 1945, pp. 108 y ss.

(6) TURING, Alan M., "On Computable Numbers, with an Application to the Entscheidungsproblem", Proceedings of the London Mathematical Society, vol. 2, n° 42, 1936, pp. 230-265.

(7) TURING, Alan M., "Computing machinery and intelligence", Mind, vol. LIX, n° 236, 1950, p. 433-460.

Hasta la fecha distintos autores y experimentos han intentado rebatir las ideas de Turing y su test sobre las máquinas pensantes. Quizás el intento más representativo es el de la denominada "Habitación china" ideado por John Searle (1932-al presente), quien buscó demostrar que el pensamiento de la persona no puede ser reducido a la lógica computacional. En su planteo la máquina puede hablar chino sin entender cómo y de qué manera lo hace.

Business Oriented Language), de 1959, mediante el cual las órdenes de los programadores se emitían directamente en inglés y ellas eran "traducidas" directamente a símbolos por la máquina. Se buscó dar origen a un lenguaje universal que sea fácilmente conocido y utilizable fundamentalmente en el mundo empresarial.

El término "inteligencia artificial" se utilizó por primera vez en la Conferencia realizada en Darthmouth College, institución universitaria situada en New Hampshire (Estados Unidos). Ello ocurrió en 1956 y sus principales impulsores fueron John McCarthy (1927-2011), Marvin Minsky, Claude Shannon y Nathaniel Rochester (1919-2001), a la postre, todas figuras de la IA. A dicha Conferencia (de alcance muy pequeño y limitado en cantidad de personas) asistieron especialistas de alto nivel, como Ray Solomonoff (1926-2009), Arthur Samuel (1901-1990), Herbert Alexander Simon (1917-2001), Allen Newell (1927-1992), Oliver Selfridge (1926-2008) y Trenchard More (1930-2019). La propuesta del 31 de agosto de 1955 que dio nacimiento al proyecto de investigación delineó diversos ejes y tiene pasajes de mucho interés para recordar [8].

Marvin Minsky y John McCarthy proponían que las máquinas podían actuar como el cerebro humano, o, mejor dicho, obtener resultados similares a los suyos, requiriéndose solamente una capacidad de procesamiento suficiente que permita replicar la dinámica neuronal exhibida por el cerebro humano. Minsky incluso sostenía que la máquina podría llegar a superar a la inteligencia humana. A partir de la figura del "perceptrón" , escribió junto a Seymour Papert (1928-2016) una obra de especial impacto y de pronóstico pesimista sobre el desarrollo de la IA, a la que muchos señalan como la que produjo lo que se denominó "el invierno" de la materia durante las décadas de los setenta y ochenta del siglo XX [9].

John McCarthy, por su parte, había sido el principal impulsor de la Conferencia de 1956 sobre Inteligencia Artificial y quien propuso dicha denominación que se mantiene hasta la actualidad. Junto a Minsky fundaron el conocido Laboratorio de Inteligencia Artificial del MIT (Massachusetts Institute of Technology) en el año 1957.

Herbert Simon y Allen Newell crearon en 1956 junto a Cliff Shaw (1922-1991) la máquina de "Lógica Teórica" (Logic Theorist) que fue una de las primeras aplicaciones prácticas sobre IA que podía imitar la actuación humana para realizar una tarea determinada (en este caso concreto dar respuesta a

(8) El texto completo de la propuesta puede consultarse en su idioma original en http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (fecha de consulta: 06/10/2025).

(9) MINSKY, Marvin – PAPERT, Seymour A., *Perceptrons. As Introduction to Computational Geometry*, MIT Press, Cambridge, 1969.

problemas matemáticos). La máquina llegó incluso a demostrar muchos teoremas que surgían de la obra de Bertrand Russell Principia Mathematica.

John von Neumann (1903-1957), además de estructurar su célebre arquitectura (tan útil para la materia), fue uno de los primeros en advertir que el sistema binario sobre el cual trabajaban los lenguajes de programación constituía un serio límite para lograr el objetivo de que la máquina se comporte igual que el cerebro (o mejor), ya que este último convivía diariamente con "zonas grises" y matices que, en gran parte, no se podían reducir a un simple juego de "blanco" o "negro". Esto lo hizo en una obra inconclusa de su autoría que se publicó en el año 1958 [(10)].

Noam Chomsky (1928-al presente), con sus aportes lingüísticos ha colaborado en el desarrollo de la IA mediante el desarrollo de lo que se conoce como "Jerarquía de Chomsky" que, a partir de diferentes tipos de gramáticas formales produce lenguajes formales (basales en las ciencias computacionales).

Ray Solomonoff (1926-2009) realizó la adaptación del pensamiento de Bayes para lograr que las máquinas puedan predecir a partir de cálculos de probabilidades en un proceso de aprendizaje automático. Gran parte de los sistemas expertos que se crearon con posterioridad usaron esa lógica para predecir mediante probabilidades lo desconocido.

El informático Joseph Licklider, además de ser uno de los principales impulsores de lo que luego fue Internet, fue de los primeros desde su área de trabajo que puso especial atención en la relación cerebro-máquina, al igual que lo hizo Wiener. El enfoque de estos especialistas era explotar las potencialidades de esa vinculación en vez de reemplazar al cerebro por máquinas. Su idea era esencialmente pragmática pues la interfaz en cuestión permitiría sacar lo mejor de cada uno: la potencia de procesamiento de información de la máquina y la creatividad y aportes originales del cerebro. En síntesis, concretar un trabajo cooperativo a partir de ambos elementos.

Licklider profundizó este esquema e ideó en un célebre paper de 1960 [(11)] el término "simbiosis hombre-computadora" (man-computer simbyosis). Esta idea cooperativa de Wiener y Licklider confrontaba de algún modo con el otro modelo estratégico y de desarrollo que era impulsado por McCarthy y Minsky, que apuntaba a que las máquinas aprendan por sí mismas reproduciendo lógicas cognitivas que eran originalmente del cerebro. En otras palabras, mientras que la consecuencia de la última postura era el reemplazo del cerebro, la idea de los primeros se basaba en que había aspectos insustituibles.

---

(10) NEUMANN, John Von, *The Computer & The Brain*, Yale University Press, New Haven, 1958.

(11) LICKLIDER, Joseph C. R., "Man-Computer symbiosis", IRE Transactions on Human Factors in Electronics, vol. 1,1960, pp. 4-11.

En línea con quienes no proponían un reemplazo de la inteligencia humana por la artificial sino una íntima complementación entre las capacidades de personas y máquinas, Douglas Engelbart (1925-2013) guió su actuación buscando como norte incrementar la capacidad intelectual de las personas. Ello lo escribió en un famoso artículo del año 1962 (12) en donde expresaba que la intención era que el complemento no solo funcione para las matemáticas sino para cualquier tipo de tarea en cuanto ella pueda ser simbolizada, todo ello al trabajar con las habilidades privativas de los humanos en tiempo real para mejorar la performance de estos.

En 1965 Lofti Zadeh (1921-2017) ideó el concepto de "lógica difusa" para tomar decisiones por aproximación, lo cual es un contenido de mucha utilidad para la IA en la actualidad.

Se trató de una novedosa manera de encontrar alguna alternativa al clásico esquema binario del "si" o "no" (o "ceros" y "unos"), dando lugar a múltiples matices útiles y necesarios en la programación algorítmica ya que gran parte de la información que utiliza una IA para "tomar una decisión" requiere de esa suerte de "suavización" o "flexibilidad" previa que permite ganar mucho en precisión.

Por supuesto que el mundo de la ficción también colaboró con lo suyo, pudiendo recordarse en tal sentido, y al solo título ilustrativo, las llamadas "Tres Leyes de la Robótica" de Isaac Asimov (1919-1992) (13).

Como fácilmente se observa a través de esta sintética, incompleta y algo desordenada reseña, gran parte de los problemas, interrogantes, debates y desafíos que genera la IA y la automatización ya vienen siendo objeto de tratamiento desde la etapa embrionaria de su desarrollo (y mucho más atrás en el tiempo aun), la cual como se ha visto, no puede ser reducida solamente a las contribuciones y planteos del genial Alan Turing como habitualmente se expone en algunos esquemas reduccionistas. En tal sentido, la mención de nombres que van desde Pitágoras a Lofti Zadeh, Joseph Licklider o Grace Murray Hopper, pasando por Thomas Bayes, Ada Lovelace o Leonardo Torres Quevedo no ha sido inocente al respecto.

---

(12) ENGELBART, Douglas C., *Augmenting Human Intellect: A Conceptual Framework*, Stanford Research Institute, Menlo Park, 1962.

(13) Isaac Asimov fue un escritor que creó obras de ciencia ficción muy conocidas y que elaboró lo que se conoce como las "Tres Leyes de la Robótica". Aparecieron por primera vez en su cuento "Círculo vicioso" y su contenido (de asombrosa vigencia) es el siguiente: 1) Un robot no puede dañar a un ser humano o, por omisión, dejar que este experimente daño; 2) Un robot debe obedecer las órdenes que les den los seres humanos, excepto cuando ellas se opongan a la primera ley; y 3) Un robot debe proteger su propia existencia en la medida que ello no entre en conflicto con las dos leyes anteriores.

3. LA INTELIGENCIA ARTIFICIAL COMO HERRAMIENTA PARA "EMPEORAR" (O "MEJORAR") LAS DECISIONES DEL CONSUMIDOR

Realizada la cronología expuesta en el apartado anterior y la mención reivindicante de distintas figuras que han sido importantes para la historia y el desarrollo de la IA toca ahora exponer algunas "ideas en crudo" que buscan servir fundamentalmente como disparador de futuras discusiones entre los operadores del derecho privado y del consumidor.

3.1. El "filtro burbuja" y la información recibida por el consumidor

El perfilamiento digital de los consumidores lleva a que gran parte de la información que reciban, por ejemplo, sobre bienes y servicios se encuentre particularizada y, posiblemente, haya tomado nota de "sus zonas débiles", documentadas con el historial de navegación, sus compras, sus clicks, etc.

Mirado el fenómeno desde el cumplimiento del deber de información, se advierte una indisimulable tensión respecto a contenidos que podrían ser omitidos para dicho sujeto pero que indudablemente harían a la esencia de su acto de consumo. Se estaría ante la omisión de algo relevante que, por lo tanto, no podría o, mejor dicho, no debería ser pasado por alto.

Esa "particularización" o carácter "teledirigido" de la información es una circunstancia relevante que tiene que ser informada también al consumidor. Es decir, el individuo debiera saber que la oferta, el momento del día en el cual la recibió, e incluso gran parte de sus características se explican bajo el esquema descripto.

El recorte en cuestión funciona entonces como un dique de contención que impide que toda la información útil y disponible llegue efectivamente a la persona. De esta manera, se limitan las posibilidades de que el consumidor tome decisiones que estén alineadas con la salvaguarda de sus intereses.

Sobre este tema es de pertinente cita un libro de Eli Pariser [14] en el cual el autor pone de resalto que la decisión de Google de personalizar los resultados de las búsquedas en 2009 llevó a que las personas comiencen a vivir en una especie de "burbuja personalizada" en la cual solo se recepciona aquello que coincide o tiene afinidad con su perfil digital, limitando la obtención de información sobre lo diferente. Advierte, en otras palabras, que

(14) PARISER, Eli, *El filtro burbuja. Cómo la red decide lo que leemos y lo que pensamos*, Taurus, Barcelona, 2017.

estas tecnologías obstaculizarían el acceso a las ideas y opiniones ajenas, afectando, en última instancia, la diversidad que constituye uno de los elementos esenciales de la vida en comunidad.

3.2. El menú interminable de opciones y la "paralización" de la decisión del consumidor

La llamada "Paradoja de la elección" [15] conduce a que una cantidad exagerada de opciones pueda traer más problemas que soluciones al consumidor. Este hallazgo no solo es contraintuitivo, sino que tensiona contra las bases mismas de la teoría del acto jurídico, el derecho del consumidor y la defensa de la competencia que buscan asegurar que un individuo tenga la mayor cantidad de alternativas disponibles para elegir.

Se observa entonces que una multiplicación y atomización excesiva de la oferta de bienes y servicios puede terminar produciendo algo parecido a la sobreinformación.

A partir de lo manifestado, no resulta difícil advertir que el deber de información no podrá omitir esta cuestión. Así, si la oferta del bien o servicio en un entorno digital incluye la propuesta de una lista interminable de alternativas que, en teoría, el consumidor tiene que evaluar previo a elegir, el deber en cuestión puede verse afectado pues su forma de exposición estaría llevando a una especie de "paralización" de la decisión de consumo. Esto puede observarse verbigracia en la inmensa cantidad de contenidos a disposición que tiene el suscriptor de una plataforma de streaming.

Un ilustrativo estudio [16] sobre el tema resaltó que la conducta de los consumidores se paraliza o ralentiza cuando se le entregan demasiadas opciones para elegir. Así, en el caso examinado, se comparó la reacción de los adquirentes frente a seis alternativas de mermeladas o chocolates gourmet, en el primer caso, y de 24 o 30 alternativas, en el segundo. Entre los sorprendentes resultados surgieron dos particularmente llamativos.

Se comprobó, en primer término, que la ampliación del menú generaba el efecto de que los consumidores se detengan más tiempo a comparar las opciones. En el caso de 24 opciones se detuvo el 60% de las personas mientras que en el de seis el 40%.

Sin embargo, en el caso de alternativas reducidas el 30% de los individuos terminaron comprando mientras que en el de muchas más opciones

---

(15) Ver: SCHWARTZ, Barry, *The Paradox of Choice: Why More is Less*, ed. revisada, Harper Collins, 2009; SCHWARTZ, Barry - WARD, Andrew, "Doing Better but Feeling Worse: The Paradox of Choice", en LINLEY, P. Alex – JOSEPH, Stephen (eds.), *Positive Psychology in Practice*, John Wiley / Sons, Nueva Jersey, 2012, pp. 86-104.

(16) IYENGAR, Sheena - LEPPER, Mark, "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?", Journal of Personality and Social Psychology, vol. 79, n° 6, 2001, pp. 995-1006.

ese índice bajó hasta el 3%. Incluso el porcentaje de satisfacción de las personas fue superior cuando eligieron en base a un menú de opciones reducido.

Para sortear estas dificultades se debe trabajar la información de manera tal que el consumidor, por ejemplo, pueda filtrar fácilmente lo que a priori le es irrelevante. Asimismo, las clasificaciones o agrupaciones en géneros y categorías para achicar las opciones van en el mismo sentido. Un adecuado cumplimiento del deber de consejo por parte de los proveedores, facilitando la comparación entre productos y servicios, también puede ayudar. De lo que se trata es, en definitiva, de acercar la información al cerebro de una manera amigable que, al mismo tiempo, no desatienda incentivar la reflexión cuando ello resulte necesario.

3.3. La información provista de manera no tradicional por las tecnologías emergentes

Los consumidores deberán acostumbrarse a que en la era digital la información pueda provenir de fuentes de lo más diversas.

Así, los bienes comercializados comienzan a traer consigo códigos QR (Quick Response) que, una vez escaneados, permiten acceder a determinada información sobre el producto. Otro tanto sucede con el fenómeno de la Realidad Aumentada (RA).

Ello trae desafíos de relevancia pues podría incrementar los riesgos derivados de una sobreabundancia de información frente a la cual el consumidor no se encuentre en condiciones de lidiar cognitivamente. En otras palabras, el gran volumen de información suplementaria que se suministraría mediante estos mecanismos podría superar largamente la capacidad del cerebro humano para procesarla.

La apuntado podría ser mitigado, por lo menos parcialmente, si la información en cuestión es entregada de manera inteligente, adaptándola a los formatos y de presentación que sean más compatibles con las características cognitivas de los individuos.

3.4. Los rankings, las recomendaciones y la memoria episódica

Una fuente cada vez más importante que inclina las decisiones de las personas en los entornos digitales es el innegable valor que surge de los rankings de plataformas abiertas y colaborativas como, asimismo, de las puntuaciones, recomendaciones y comentarios allí vertidos.

Si bien se trata de una herramienta válida que indudablemente pueda ayudar a contar con información más fidedigna sobre productos y servicios, es factible que aparezcan tensiones jurídicas cuando esas opiniones “supues-

tamente neutras" en realidad han sido suministradas por el proveedor por sí o por terceros, sea bajo la figura de influencers, o bien de "personas anónimas" que reciben un pago o beneficio a cambio de una valoración positiva.

Cuando se altera la pureza de la fórmula en cuestión, se está ante un incumplimiento notorio en lo referido a la veracidad de la información y a la omisión de circunstancias relevantes como, por ejemplo, el carácter interesado del comentario, posteo o recomendación.

Lo dicho debe ser supervisado con especial atención pues estos datos suelen tener particular incidencia en la formación de la decisión en tanto el sujeto no confía en lo entregado por el vendedor sino, por el contrario, en las experiencias y vivencias de sus pares mediante la generación de un fenómeno identificatorio.

Es decir, el mecanismo trabaja sobre la base de la memoria episódica mediante experiencias que, si bien no son propias, se las percibe muy cercanas a ellas.

3.5. La inteligencia artificial como "sistema de apoyo" de las decisiones del consumidor

La IA no solo debe ser vista como una amenaza para los consumidores [17]. También podría ser parte de la solución de sus problemas [18].

Se ha resaltado que, si bien una IA podría "comprender" el contenido de un documento, "una cosa es saber lo que dice literalmente", pero "otra comprender su significado" pues "para ello es preciso el ser humano" [19]. No obstante, las predicciones o sugerencias que entregan los algoritmos, aun sin entender "por qué hacen lo que hacen", son de alta significancia.

---

(17) Dentro de Argentina ya se han escrito distintos trabajos que abordan el fenómeno de la Inteligencia Artificial desde una perspectiva de tutela y protección de los consumidores y usuarios. Se puede ver, por ejemplo, ÁLVAREZ LARRONDO, Federico M., "Desafíos para el Derecho en general, y del consumo en particular. La construcción del nuevo 'Derecho artificial' y la excusa para cambiar el rumbo de la Argentina", RDCO 292, 119; CHAMATROPULOS, Demetrio Alejandro, "Inteligencia Artificial, prevención de daños y acceso al consumo sustentable", LA LEY, 2017-E, 1044; FRUSTAGLI, Sandra A., "Derecho del consumidor frente a los desafíos del uso de la Inteligencia Artificial (IA) en las relaciones de consumo", SJA 14/07/2021, 19; MULER, Germán E., "Los principios del sistema de protección del consumidor y su proyección en los entornos digitales", LA LEY 2021-A ,99 y VALDEZ, Fernando-IRIARTE YANICELLI, Adolfo – NACUL, María Sofía – SAHIAN, José Humberto – STEIMBERG, Bárbara, "Los nuevos derechos de los consumidores en la era de la Inteligencia Artificial", en CORVALÁN, Juan G. (dir.), *op. cit.*, pp. 371 y ss.

(18) Ello no implica desconocer que "el 'robot' como un 'otro' ¨(...) ya ha comenzado a generar preocupaciones concretas para la teoría jurídica, particularmente, a partir del hecho de que la robótica se ha desplazado de un paradigma 'mecánico' (que tenía por objetivo la aceleración de movimientos precisos y complejos) hacia un paradigma 'cognitivo' (que tiene por fin desarrollar capacidades artificiales autónomas)". DE LORENZO, Miguel Federico, "Repensar al 'otro': reflexiones sobre el derecho civil", SJA 17/04/2019, 3.

(19) NIEVA FENOLL, Jordi, *Inteligencia Artificial y proceso judicial*, Marcial Pons, Madrid, 2018, p. 93.

La normativa que regula la salud mental en Argentina establece un "sistema de apoyos" (art. 43, Cód. Civ. y Com.) para respaldar la actuación de los sujetos con padecimientos mentales.

Sería interesante proponer, ya de manera amplia y general, que una IA pueda cumplir ese rol de "asistente" de los consumidores en sus decisiones diarias, aconsejando sobre la conveniencia de una contratación, sugiriendo cursos de acción o señalando cómo interpretar determinada información recibida.

Desde ya que el consumidor puede o no tomar el "consejo artificial". Sin embargo, lo señalado podría ser de mucha utilidad para mejorar la performance decisional de los individuos en cuestión, neutralizando, aunque sea parcialmente, la actuación en contra de sus propios intereses.

En caso que la alternativa sea viable y entregue resultados exitosos se podría dar nacimiento a un "sistema de apoyos" de naturaleza algorítmica que no esté destinado solo a las personas con padecimientos mentales sino a todos los individuos que lleven adelante actos de consumo.

Herramientas como las descriptas no encontrarían ningún impedimento legal actual ya que se estaría ante mecanismos que podrían funcionar como "mejoradores de las decisiones".

3.6. ¿La inteligencia artificial puede detectar que las decisiones de consumo han sido irreflexivas?

Íntimamente relacionado con lo anterior, la IA y la predicción algorítmica podrían ser muy útiles también para "detectar" ex post comportamientos o decisiones "irracionales" de los consumidores.

Así, verbigracia, se podría elaborar una especie de "semáforo" mediante el cual la IA realice un trabajo de valoración a posteriori sobre datos referidos a los citados comportamientos y luego entregue un estado "rojo", "verde" o "amarillo" al consumidor para que este pueda registrar qué características tiene su decisión o incluso el historial de estas, pudiendo incluso vislumbrarse tendencias que le faciliten la propia evaluación de su comportamiento.

Estos datos se podrían cargar manualmente si se trata de actuaciones en el "mundo físico" o ser recogidos directamente por el software si se actúa en el entorno digital.

Por supuesto que sería muy útil que este semáforo se despliegue como paso previo a una contratación electrónica, pero para esto ya se dependería de los proveedores siendo difícil que, por iniciativa propia, ellos pongan herramientas como la descripta a disposición del consumidor en sus plataformas digitales. No obstante, bien podría suceder que un tercer proveedor actúe como soporte en auxilio del consumidor en dichos contextos (a cambio

de un precio, por ejemplo) proveyendo el sistema de evaluación en cuestión y colaborando con la protección de su situación contractual.

La misma lógica podría estar presente en un dispositivo que, en base a una serie de preguntas previas que se le realicen al consumidor sobre las características e implicancias de un contrato que está por celebrar, puede predecir cuan reflexivo ha sido su consentimiento al respecto.

3.7. La detección y las advertencias del cansancio mediante algoritmos

El cumplimiento del deber de advertencia tiene mucho que hacer cuando se está ante supuestos en donde se incrementa el riesgo y el peligro por el cansancio experimentado por el consumidor.

El comportamiento de los conductores al volante es quizás un ejemplo muy claro al respecto.

En este punto la IA tiene mucho que aportar. Por ejemplo, diseñar algoritmos que, en base a determinados parámetros, puedan predecir el nivel de sueño o cansancio del conductor. Ello se puede derivar del horario y de la zona por donde se está conduciendo, de la realización o no de maniobras bruscas o sorpresivas, de la marcha zigzagueante del vehículo y de muchos otros factores. En la actualidad, determinados automóviles de calle ya cuentan con ciertas tecnologías que apuntan a esta finalidad, sugiriendo detenerse a descansar o a tomar un café mediante alertas que se disparan a partir de indicadores como los recién mencionados.

Realizada dicha prueba, y comprobada una alta probabilidad de cansancio de la persona, además de activarse respuestas preventivas como la del aviso expreso de dicha circunstancia al individuo, se podría ir más lejos e informarle en ese mensaje que el vehículo se detendrá en los próximos quince minutos, por ejemplo, y que solo podrá ser encendido nuevamente luego de un tiempo determinado si es que se supera un test que demuestre que ese cansancio disminuyó y que es posible seguir manejando.

3.8. Delegación de decisiones en máquinas y "derecho a equivocarse"

Como salvedad a lo expuesto recién respecto a los "consejos artificiales" hay que precisar que, en determinadas situaciones, no pasará mucho tiempo para que algunas decisiones de los consumidores deban delegarse obligatoriamente en algoritmos.

Es muy posible que ello suceda con la conducción de vehículos automotores cuando se llegue a un punto tal en el que una máquina procese información contextual relevante de manera mucho más satisfactoria que un ser

humano y, en consecuencia, tome decisiones manifiestamente mejores que él, pulverizando los riesgos que hoy surgen de la conducta del individuo.

Dadas las externalidades negativas que produciría el comportamiento de la persona al volante (accidentes, muerte o lesiones de individuos, daños patrimoniales, etc.) pareciera forzoso concluir que en no muchos años los sujetos tendrán prohibido conducir por sí y que dicha tarea sea llevada adelante por una IA.

Sin embargo, existe un universo extremadamente amplio de actuación del individuo en el que esos daños no se generan o situaciones en las cuales, en todo caso, el único afectado es el consumidor mismo. Sería la situación de todo lo que entra en el principio de reserva de la segunda parte del art. 19 de la Constitución argentina [20].

Así, verbigracia, si el día de mañana se demuestra que un algoritmo evidencia "índices muy exitosos" para elegir pareja, la persona podría "renunciar" a ese consejo, aun a riesgo de "equivocarse" pues se estaría ante una clarísima manifestación de ejercicio de la dignidad humana. Al no haber terceros potenciales afectados ni tampoco una comunidad perjudicada, el consumidor tendría consagrado para sí una especie de "derecho a equivocarse", sobre el cual la doctrina debiera comenzar a trabajar urgente para determinar sus límites y zonas difusas.

Nótese que esta idea no es nueva y que incluso ha sido objeto de tratamiento en un célebre y clásico precedente de la Corte Suprema de Justicia de la Nación. La referencia concreta es al caso "Sejean vs. Zaks de Sejean" [21] de 1986 que en su momento habilitó el divorcio vincular en Argentina.

4. A MODO DE CIERRE

El sintético repaso histórico que se efectuó, yendo incluso muchos siglos hacia atrás, permite dejar expuestas distintas ideas y autores que han incidido de manera definitoria en la configuración actual de la inteligencia artificial, en sus problemas, debates y desafíos. Gran parte de lo reseñado

---

(20) Art. 19, Constitución Nacional: "Las acciones privadas de los hombres que de ningún modo ofendan al orden y a la moral pública, ni perjudiquen a un tercero, están solo reservadas a Dios, y exentas de la autoridad de los magistrados. Ningún habitante de la Nación será obligado a hacer lo que no manda la ley, ni privado de lo que ella no prohíbe".

(21) CS, 27/11/1986, "Sejean, J. B. c. Zaks de Sejean, A. M.", Fallos: 308:2268. El último párrafo del considerando 15 del voto del juez Enrique Petracchi dejó expuesta la idea del "derecho al fracaso". Concretamente expresó allí: "En cualquiera de los otros derechos constitucionales cada habitante de la Nación que goza de él y lo ejerce puede fracasar en obtener los resultados que buscaba al ejercerlo. Se puede fallar al ejercer el derecho a trabajar, o el de enseñar, o el de aprender, o el de ejercer una industria lícita, o el de peticionar a las autoridades, y así con los demás. En ninguno de los casos ese fracaso, que tratándose del derecho a casarse puede no obedecer en absoluto a causas controlables por la voluntad del titular del derecho, conlleva su pérdida definitiva".

no suele ser mencionado con frecuencia en los trabajos sobre la materia y es precisamente por esa razón por la cual fueron aquí incluidos.

Si bien se podría sostener que esas páginas iniciales no son propias de un artículo de doctrina jurídica, la decisión de dedicar espacio a esos contenidos fue realizado en el entendimiento de que los operadores jurídicos, para comprender el fenómeno, requieren tener acceso al panorama más completo que puedan para intentar entender la historia y evolución de la IA, con sus avances, retrocesos, complejidades y zigzagueos.

Desde otro lugar, este trabajo buscó también poner el foco en distintas situaciones y circunstancias en las cuales las decisiones de los consumidores en los entornos digitales experimentan dificultades a partir de desarrollos tecnológicos y estrategias empresariales diseñadas bajo el paradigma de la "algoritmización".

Complementariamente, y como la otra cara de la moneda, se intentó poner en valor también a la IA como herramienta "mejoradora" de las decisiones de los consumidores que puede ayudar a elevar la performance de estos últimos y así resguardar mejor sus intereses.

Por último, se indagó sobre un tema que viene adquiriendo relevancia y que necesita una atención plena y prioritaria por las innumerables tensiones que puede traer aparejadas. La referencia es ni más ni menos que al "derecho a equivocarse".

Inteligencia artificial y responsabilidad penal: algunas implicancias desde la teoría del delito [*]

Por Juan Ignacio Diaz [**], Graciela Elizabeth Gómez [***] y Adriana Belén Pujol [****]

Resumen: La inteligencia artificial ha transformado diversos ámbitos sociales, pero sus implicancias en el ámbito penal aún carecen de un análisis profundo. Ello plantea un interrogante central: ¿cómo debe configurarse la responsabilidad penal cuando un sistema inteligente genera un resultado lesivo? Este trabajo examina la cuestión desde tres ejes: los alcances conceptuales de la IA, el derecho penal y las implicancias de la teoría del delito, con el fin de determinar las posibles responsabilidades penales que podrían corresponder en los casos bajo estudio, así como los límites normativos que condicionan su tratamiento.

Palabras clave: inteligencia artificial; derecho penal; teoría del delito; sistemas inteligentes; responsabilidad.

Abstract: Artificial intelligence has transformed various social spheres, but its implications for criminal law remain to be thoroughly analyzed. This raises a central question: how should criminal liability be defined when an intelligent system generates a harmful result? This paper examines the issue from three perspectives: the conceptual scope of AI, criminal law, and the implications of criminal theory. This paper aims to determine the potential criminal liability that could apply in the cases under study, as well as the regulatory limits that govern their treatment.

Keywords: artificial intelligence; criminal law; crime theory; intelligent systems; liability.

(*) El presente trabajo surge en el marco del Proyecto de Investigación: "Un Análisis Constitucional de los Delitos Informáticos en el Código Penal Argentino" (UCP, res. 920/2023), Facultad de Ciencias Jurídicas y Políticas, Instituto de Investigaciones Científicas (IDIC), Universidad de la Cuenca del Plata, Corrientes, Argentina.

(**) Abogado, Esp. en Derecho Penal, Esp. en Docencia Universitaria; Prof. Tit. de "Derecho Penal Especial", Universidad de la Cuenca del Plata.

(***) Abogada, Esp. en Derecho Procesal Penal, Prof. Universitaria, Prof. Tit. de Práctica IV, Universidad de la Cuenca del Plata; y Prof. Adj. de Derecho Procesal Penal, Universidad Nacional del Nordeste.

(****) Esp. en Derecho Penal, Prof. Universitaria.

1. Introducción

Este trabajo estudia los retos que plantea la inteligencia artificial (en adelante, "IA") en el derecho penal, en particular la configuración de la responsabilidad cuando sistemas inteligentes producen resultados lesivos. La cuestión se examina desde tres ejes: los alcances conceptuales de la IA, su irrupción en el derecho penal y las implicancias derivadas de la teoría del delito. El objetivo es determinar las posibles responsabilidades penales que podrían corresponder en los casos bajo estudio, considerando tanto su impacto social como los riesgos que plantean para la protección de los derechos humanos. La sociedad se ha acostumbrado al vertiginoso, dinámico e imprevisible avance de la tecnología. En este sentido, ya no resulta discutible que los avances científicos al respecto han impactado de lleno en la vida de las personas, generando beneficios innegables. Sin embargo, los sistemas inteligentes trazan múltiples contradicciones en relación con la protección de los derechos humanos.

2. La inteligencia artificial: concepto, características e impacto en la sociedad inteligente

En la actualidad, se habla de un término encantador: "sociedad inteligente". Este concepto se centra en un modelo de organización social orientado a optimizar los recursos humanos y tecnológicos con el fin de lograr un mayor bienestar general. Esta denominación es utilizada desde distintas disciplinas para explicar la vinculación de la tecnología y la informática con la vida de las personas y su impacto en la sociedad.

- Nos recuerda con acierto Ana Dobratinich (2023) que fue en la década de 1950 cuando la IA alcanzó su mayor autonomía como disciplina emergente dentro de las ciencias informáticas. Lo que en sus inicios se concebía principalmente como un conjunto de interrogantes filosóficos, progresivamente se fue incorporando a distintos ámbitos de la sociedad y de las ciencias en particular, tales como la medicina, el derecho, entre otros.

- Los sistemas de IA (o también llamados "sistemas inteligentes") son principalmente elementos de *software* y, en algunos casos, también de hardware, creados por seres humanos, que actúan conforme a un objetivo determinado. Estos sistemas pueden operar de manera activa u omisiva, en ámbitos tanto físicos como digitales, distinguiendo su ambiente a través de la adquisición e interpretación de datos, ya sean estructurados o no estructurados. A partir de ello, razonan sobre el conocimiento o procesan la información obtenida de dichos datos, para finalmente indicar y adoptar las decisiones más adecuadas con el fin de alcanzar el objetivo anteriormente preestablecido.

Los sistemas de IA actúan únicamente dentro de los objetivos previamente fijados por los seres humanos, pues su diseño y programación siempre dependen de la intervención humana. En consecuencia, su aplicación en el mundo real está necesariamente limitada por esa finalidad y exige establecer medidas de protección que prevengan eventuales lesiones a bienes jurídicos.

3. La irrupción de la inteligencia artificial en el derecho penal

La problemática gira en torno a los sistemas inteligentes que, con mayor o menor grado de autonomía según los datos que procesen, pueden generar consecuencias imprevistas. Ahora bien, si de la actuación de un robot resulta un daño no previsto ni calculado por el programador o por el ingeniero garante de su funcionamiento, surge el interrogante acerca de quién debe responder penalmente por los resultados lesivos ocasionados por la máquina. En este ámbito, cuando los robots inteligentes causan una lesión a bienes jurídicos protegidos por el derecho penal, aparece la cuestión de la responsabilidad penal: ¿puede el robot ser considerado autor de un delito?, ¿la responsabilidad recae en el programador que incurrió en un error al implementar la información o los datos?, o, en su caso, ¿es responsable el ingeniero garante del funcionamiento del robot que no procesó correctamente la información necesaria para que este actuara conforme a lo ordenado?

La revolución tecnológica, y en particular el uso de la IA en aparatos, máquinas, robots y objetos análogos, han irrumpido en el ámbito del derecho penal. Los casos hipotéticos plantean el interrogante de estudiar la responsabilidad penal, formular algunas precisiones desde la teoría del delito y, asimismo, reflexionar sobre la autoría y/o participación criminal a partir de la dogmática penal. Incluso podría agregarse un aspecto más complejo: la posibilidad de que esta temática implique una reestructuración de la definición tradicional de delito.

4. Inteligencia artificial y responsabilidad penal: algunas implicancias desde la teoría del delito

La responsabilidad penal encuentra su límite de aplicación en el principio de legalidad penal (arts. 18 y 75, inc. 22, Constitución Nacional). De este modo, el juez no puede imponer una sanción basándose únicamente en su voluntad o en principios morales, ya que tiene expresamente prohibido crear tipos penales, construir ficciones de autoría o llenar vacíos legales.

El interrogante que surge en relación con la problemática abordada es si los sistemas inteligentes pueden ser concebidos como personas naturales (físicas) o personas jurídicas (ideal). Para responder a ello, resulta relevante

remitirse a los artículos 19 a 21 del Código Civil y Comercial de la Nación (CCCN), que prescribe que persona física es aquel ente humano que, concebido o implantado en la mujer, nace con vida, cumpliendo los requisitos exigidos para adquirir derechos y contraer obligaciones conforme al ordenamiento jurídico argentino.

El concepto de responsabilidad jurídico-penal se desarrolla exigiendo la existencia de un sujeto imputable. Por ello, se ha afirmado que sin imputabilidad no hay responsabilidad penal. En consecuencia, el derecho penal requiere que el sujeto posea la capacidad necesaria para responder conforme al sistema penal y a las normas que demandan su comprensión dentro del orden legal.

De lo expuesto se puede concluir que los sistemas inteligentes no pueden ser imputables de acuerdo con la legislación vigente en Argentina, ya que no comprenden la criminalidad de los actos que realizan y carecen de autodeterminación, dado que no poseen la capacidad de ser sujetos imputables —seres humanos— en el ámbito penal. [(1)] De esta manera, los sistemas inteligentes pueden emular al ser humano. Sin embargo, normativamente no cumplen los requisitos exigidos para ser imputables de un delito, puesto que, toda la regulación vigente hace referencia exclusivamente a seres humanos.

Al respecto, se ha propuesto como alternativa considerar y equiparar a los sistemas inteligentes con las personas jurídicas. En este sentido, se sabe que las personas de existencia ideal carecen de existencia individual (física o natural) aunque poseen capacidad legal para ser sujetos de derechos y obligaciones. En principio, cabe aclarar que la doctrina penal mayoritaria niega esta posibilidad de reconfigurar el concepto de culpabilidad y de dotar de imputabilidad a las personas jurídicas. [(2)]

En Argentina, el Código Penal de la Nación, en su art. 34 inc. 1, establece que no son punibles quienes en el momento del hecho no hubieran podido "comprender la criminalidad del acto o dirigir sus acciones". Ello permite concluir que, desde el derecho positivo y en cumplimiento del principio de legalidad penal (art. 18, CN), resulta imposible adjudicar responsabilidad penal a las personas jurídicas. Se requiere un ser humano, con cuerpo (físico) y psiquis (salud mental sana), que lo haga imputable de acuerdo con la normativa penal.

---

(1) Véase, arts. 1° y 2° de la del Régimen Penal de la Minoridad (ley 22.278 de 1980)

(2) Entre las posturas más destacadas se encuentra la del Prof. Claus Roxin, quien sostiene que las personas jurídicas no pueden ser culpables en sentido jurídico-penal, ya que carecen de sustrato anímico y espiritual, atributos que sí poseen los seres humanos. Por lo tanto, ni siquiera realizan una "acción" en el sentido estricto de la teoría del delito (Roxin, 1997). En la misma línea, el Prof. Günther Jakobs niega que las personas jurídicas tengan capacidad de culpabilidad. Según su desarrollo, la culpabilidad exige que el sujeto sea autoconsciente y comunicativamente competente, condiciones que solo cumplen las personas físicas y no las de existencia ideal (Jakobs, 1997).

Por otro lado, en la doctrina disidente se encuentra el Prof. González (2018), quien afirma que las personas jurídicas pueden ser sujetos imputables, es decir, se les podría atribuir responsabilidad penal sin quebrar el principio de culpabilidad, siempre que se elaboren criterios funcionales y normativos claros y precisos. No obstante, Camilo Momblanc (2024) descarta esta posibilidad de imaginar y equiparar a los sistemas inteligentes con las personas jurídicas, dado que las personas de existencia ideal presentan diferencias sustanciales que conllevan una valoración jurídica distinta.

En esta línea, se sostiene que las personas jurídicas no tienen cuerpo físico, sino que actúan por medio de sus representantes. Se argumenta que la IA podría actuar en determinados supuestos realizando maniobras u operaciones jurídicas por sí misma, sin necesidad de la participación de una persona física; por ejemplo, trabajando y ocasionando un daño de forma autónoma. En cambio, una persona jurídica no puede provocar una lesión por sí misma, sino únicamente a través de sus representantes. En efecto, la consumación del tipo penal exige la concurrencia de un comportamiento humano lesivo a los intereses de la sociedad (Camilo Momblanc, 2024; Almazán Salazar, 2021).

Consideramos que las personas jurídicas poseen personalidad jurídica fundada en la capacidad de obrar de las personas físicas que integran dicha organización ideal. De esta forma, la persona jurídica no tiene cuerpo físico: únicamente actúa por medio de sus representantes, quienes desde el punto de vista penal responderán por las conductas directivas (mandatos) que lesionen intereses individuales o colectivos.

De lo anteriormente desarrollado, corresponde preguntarnos: ¿cómo se vincula la IA aplicada a robots, máquinas u objetos análogos con la teoría del delito cuando su conducta (activa u omisiva) causa lesión a un bien jurídico individual o colectivo?

La problemática puede analizarse a partir de cuatro supuestos: (1) Lesiones a bienes jurídicos protegidos ocasionadas por personas que utilizan IA, ya sea programándola de forma específica para fines ilícitos, o bien alterando o modificando el sistema o *software* con determinada finalidad antijurídica; (2) Daños derivados de la responsabilidad del creador del objeto dotado de IA, vinculados a defectos de diseño, programación o implementación del sistema; (3) Lesiones que provienen de la conducta del consumidor o usuario como consecuencia de la inobservancia de los deberes de cuidado que debía adoptar respecto del aparato con sistema inteligente bajo su cargo y responsabilidad; (4) Situaciones de daños a bienes jurídicos individuales o colectivos ocasionados por robots, maquinas u objetos análogos con IA altamente avanzados, capaces de adoptar decisiones autónomas e imprevisibles sin intervención directa del ser humano.

En el primer supuesto, no existen mayores inconvenientes para determinar la responsabilidad penal. El sujeto que utiliza un robot, máquina u objeto análogo programando su actuación u ordenándole que realice alguna actividad dañosa contra una persona será responsable bajo la modalidad de autor directo. En este sentido, los delincuentes utilizarían la IA en sus diferentes modalidades y formas de aplicación para cometer delitos. Por lo tanto, consideramos que, en estos casos, conforme a la doctrina penal del dominio del hecho, será autor toda persona que realice una conducta ilícita valiéndose de sistemas inteligentes, sean estos utilizados como herramientas, instrumentos o medios para la ejecución del ilícito penal.

Para Quintero Olivares (2017), en estos casos será autor y responsable en términos penales quien disponga de la capacidad de programar la acción de un robot inteligente, ya sea un humanoide o una máquina sin forma humana, con la finalidad de delinquir y habiendo cometido un delito.

Con respecto al segundo supuesto, el sujeto creador del robot, máquina u objeto análogo al cual instauró determinada IA para que lleve adelante sus fines de creación será responsable de los daños derivados de su responsabilidad por defectos de diseño, programación o implementación del sistema de funcionamiento (*software*). En efecto, la responsabilidad penal será atribuida al individuo bajo la modalidad de un actuar culposo. En estos casos será determinante analizar si el resultado lesivo fue causado por la IA aplicada a los objetos a consecuencia de errores en los sistemas o programación, o bien porque el creador u operador ha realizado una conducta que encuadra en la infracción del deber objetivo de cuidado. De ahí que, si en el caso concreto el robot, máquina u objeto análogo ejecuta una tarea fallida o incorrecta, podrá quedar bajo la órbita de la responsabilidad penal del creador, operario o ingeniero, siempre y cuando se cumplan los presupuestos que exigen los delitos imprudentes.

De acuerdo con lo expresado por el Prof. Domínguez Henaín (s. f.), el tipo culposo se caracteriza porque contiene como materia de prohibición aquellos comportamientos en los que el autor no persigue ni quiere la afectación del bien jurídico penalmente tutelado. Tales conductas se consideran prohibidas no por su finalidad —que incluso puede ser irreprochable—, sino por la forma en que el sujeto pretende alcanzarla. En este marco, es la falta del cuidado debido en su obrar (conducta defectuosa) lo que provoca la lesión del objeto de la acción, es decir, el resultado típico.

Habrá imprudencia cuando el autor no se represente en absoluto el riesgo de lesión, pero, conforme a las circunstancias de su obrar y a los deberes objetivos de cuidado que le eran exigibles, debió representárselo (culpa sin representación o culpa con representación). Lo que caracteriza al delito imprudente es, entonces, una conducta que infringe deberes de cuidado y cuya infracción resulta determinante en la producción del resultado. Co-

rresponde al juez precisar, en cada caso, cuál es el deber de cuidado vulnerado por el autor (González, 2018). Para aclarar entonces, en estos casos en los que el sujeto creador del robot, máquina u objeto análogo, les instauró determinada IA para que lleve adelante sus fines de creación, aquel será responsable de los daños derivados de su responsabilidad por defectos de diseño, programación o implementación del sistema de funcionamiento (*software*). En efecto, la responsabilidad penal será atribuida al individuo bajo la modalidad de un actuar culposo.

La misma solución debe darse al tercer supuesto, cuando las lesiones provienen de la conducta del consumidor o usuario como consecuencia de la inobservancia de los deberes objetivos de cuidado que debía adoptar respecto del aparato con sistema inteligente bajo su cargo y responsabilidad. En este caso, si el propietario del robot, máquina u objeto análogo con IA ordena determinada actividad sin atender a las consecuencias lesivas que se podrían ocasionar a bienes jurídicos protegidos de las personas, su actuación obedecerá a una infracción a los deberes objetivos de cuidado, siempre que el hecho dañoso sea previsible y evitable.

La conducta previa que originó el accionar —activo u omisivo— del sistema inteligente deberá ser examinada de acuerdo con las pautas de la imprudencia y negligencia sobre las cuales se construyen los tipos culposos en el Código Penal argentino. Por su parte, los autores Valls Prieto (2022) y Camilo Momblanc (2024), al abordar esta problemática penal, advierten que existen riesgos en el desarrollo de los sistemas inteligentes, puesto que las indicaciones a los ciudadanos (consumidores o usuarios) pueden ser muy técnicas y complejas, lo cual generaría la atipicidad de la conducta del fabricante o creador por los daños ocurridos, así como también porque estos objetos que funcionan con IA pueden, por sí solos, causar daños no previsibles o inesperados debido a defectos del sistema de funcionamiento, lo que implicaría no responsabilizar penalmente a ninguna persona por los daños producidos en estas condiciones.

Para ello será necesario e indispensable determinar las normas de cuidado que deben respetar las personas que utilizan los sistemas inteligentes, así como aquellos que diseñan dichos robots, máquinas u objetos análogos con IA. Como política estatal se deberá regular cuáles son las normas que deben cumplir los ciudadanos que intervengan en la cadena de fabricación, programación y ejecución de un sistema de IA altamente peligroso para la sociedad.

El cuarto supuesto se relaciona con la atribución de la responsabilidad penal en situaciones de daños a bienes jurídicos individuales o colectivos ocasionados por robots, maquinas u objetos análogos con IA altamente avanzados (IA fuerte), capaces de adoptar decisiones autónomas e imprevisibles sin intervención directa del ser humano. Dicho supuesto indicado

resulta de mayor complejidad, puesto que se inscribe en el contexto actual del desarrollo del derecho penal frente al dinámico avance de la IA y su aplicación a diversos ámbitos con impacto directo e indirecto en los ciudadanos. La capacidad de ciertos sistemas de IA para actuar con un grado de autonomía y tomar decisiones en situaciones reales constituye un desafío indiscutible para las ciencias y, en particular, para el derecho.

No existe una clasificación única de los sistemas inteligentes. Sin embargo, resulta útil distinguir entre IA débil, diseñada para tareas específicas con autonomía limitada, e IA fuerte, caracterizada por una mayor capacidad para razonar, decidir y resolver problemas, llegando incluso a emular conductas semejantes a las humanas. Mientras la primera corresponde a los sistemas más comunes y de uso cotidiano, la segunda plantea mayores desafíos frente al derecho penal, ya que, a pesar de su nivel de autonomía, no puede ser considerada persona en los términos del ordenamiento jurídico, ni ser sujeta de responsabilidad penal. En consecuencia, los daños ocasionados por estos sistemas deben analizarse siempre desde la responsabilidad de los seres humanos involucrados en su creación, programación o uso.

Estos últimos son los más problemáticos frente a la cuestión de la atribución de responsabilidad penal. Como se ha explicado, aun cuando puedan desarrollar conductas semejantes a las humanas, no son personas en los términos definidos por el ordenamiento jurídico argentino. Por lo tanto, no pueden ser considerados responsables penalmente ni se les puede reprochar una conducta lesiva desde la perspectiva jurídico-penal.

Por muy avanzados que se encuentren los sistemas inteligentes (IA fuerte), siempre conservarán un grado de autonomía determinado por su creador o programador, o, en su caso, por el consumidor y/o usuario. En consecuencia, los daños ocasionados por esta clase de aparatos inteligentes deben ser analizados desde la estricta observancia del deber objetivo de cuidado que corresponde en las distintas fases de su creación, desarrollo, ejecución y funcionamiento (Camilo Momblanc, 2024).

En el caso de que se determine que la conducta llevada adelante no encuadra como culposa por parte del creador, programador o ingeniero (y, en su caso, del usuario) y que, por el contrario, estuvo guiada por conocimiento y voluntad de generar un daño a bienes jurídicos protegidos por el derecho penal, la conducta de estos sujetos que intervienen en alguna de las fases mencionadas se analizará bajo los presupuestos que exigen las figuras dolosas. Así, por ejemplo, debería estudiarse el caso en el que una persona física (usuario) programa un sistema inteligente con fines ilícitos y, como consecuencia, ordena a un robot humanoide cometer un homicidio.

Al respecto, agrega el autor Momblanc (2024) que resulta imprescindible que se respeten los principios éticos que inspiran el desarrollo y la

implementación de la IA en la sociedad y, en particular, en sus ámbitos de actuación. De ahí que se considere ineludible la intervención y supervisión humana, así como también la garantía de solidez y seguridad técnicas, a fin de que estos objetos inteligentes puedan convivir armónicamente con los ciudadanos.

- Conforme a la teoría del delito, los sistemas con IA fuerte, por muy avanzados que sean —tanto en términos cualitativos como cuantitativos— y aunque logren emular razonamientos, emociones y capacidades intelectuales iguales o superiores a las de un ser humano, no podrían ser considerados sujetos susceptibles de reproche penal. Se comparte así lo expuesto por Camilo Momblanc (2024), quien sostiene que, desde la teoría del delito, los sistemas inteligentes son entes incapaces de acción. Por lo tanto, no resulta necesario continuar con el análisis de los restantes elementos: tipicidad, antijuridicidad y culpabilidad. Así, la función de la acción como elemento de la teoría del delito consiste en ofrecer el soporte mínimo del edificio del delito, el cual, en cualquier caso, se funda en una conducta humana.

La problemática resulta equivalente cuando indagamos en la tipicidad. Los sistemas inteligentes (IA fuerte) no pueden actuar de forma dolosa o culposa, ya que su conducta no puede ser evaluada desde un plano causal, sino que requiere un proceso ejecutado por una persona humana con voluntad (finalidad). Siguiendo este esquema, la conducta típica —supongamos— de un sistema inteligente no podrá ser justificada, ya que los permisos y eximentes previstos en el ordenamiento jurídico se aplican únicamente a personas físicas, pues su fundamento es precisamente evitar injusticias fuera del derecho. Asimismo, no se podrá examinar el último componente de la teoría del delito: la culpabilidad, la cual constituye el reproche penal que la sociedad hace de una conducta tipificada y antijurídica. Sus condiciones están estructuradas y diseñadas para personas naturales con voluntad, capaces de motivarse y autodeterminarse conforme a la norma penal.

No existe evidencia científica sobre la posibilidad de que las máquinas inteligentes lleguen a desplegar una conciencia de sí mismos que les permita comprender el valor social negativo de sus acciones y captar la función preceptiva del derecho penal. Esta es una capacidad lógicamente humana aún no emulada ni por la IA más avanzada (Camilo Momblanc, 2024).

La culpabilidad, tal como la concibe la dogmática penal más prestigiosa, está desarrollada y pensada para seres humanos, pues son los únicos sujetos capaces de pensamiento, razonamiento y motivación conforme al derecho vigente, actuando con voluntad suficiente según el imperativo de la norma jurídica o ante la coacción de una pena. Para finalizar, diremos que el estudio de la teoría del delito y de los elementos que la integran: acción, tipicidad, antijuridicidad y culpabilidad, demuestra que los sistemas con IA fuerte no pueden ser considerados sujetos susceptibles de adjudicación penal.

La evidencia científica y normativa nos lleva a concluir indefectiblemente que estos entes (robots humanoides y otros sistemas inteligentes) no pueden ser considerados sujetos susceptibles de responsabilidad penal, de conformidad con el principio de legalidad y el principio de culpabilidad que fundamentan la teoría del delito, independientemente del grado de avance o progreso que alcancen en la emulación de las capacidades humanas. En este sentido, los robots, máquinas u objetos análogos carecen de voluntad, no comprenden la significación jurídica de sus actos ni poseen capacidad de autodeterminación, circunstancias todas ellas esenciales para atribuir responsabilidad penal.

En consecuencia, la acción de un sistema inteligente (materializada en un aparato de cualquier índole tecnológica) estará siempre determinada por programaciones, instrucciones o mandatos humanos. De este modo, la tipicidad tampoco puede admitirse, ya que estos entes, al no ser capaces de razonar como un ser humano, no pueden orientar sus conductas ni ser evaluados en términos de dolo o culpa. Asimismo, no responden a los presupuestos exigidos por la antijuridicidad y la culpabilidad, institutos dogmáticos diseñados exclusivamente para las personas físicas.

En conclusión, la responsabilidad penal recae necesariamente sobre los seres humanos, ya sea el creador, programador, ingeniero o usuario que haya intervenido en el hecho. El auténtico reto se centra, entonces, en el ámbito probatorio, donde corresponderá demostrar quién fue el sujeto que determinó el hecho lesivo y bajo qué modalidad lo hizo al momento de implementar, programar o manipular el funcionamiento del sistema inteligente (robots, máquinas u objetos análogos). Por el contrario, si el hecho lesivo (dañoso) no resulta previsible, ni evitable por el autor, no se le podrá cargar con la responsabilidad penal, puesto que, implicaría habilitar la responsabilidad objetiva (*versari in re illicita*).

5. CONCLUSIONES

Este trabajo examinó la relación entre la inteligencia artificial y el derecho penal, analizando tanto situaciones actuales como posibles escenarios futuros. Se abordó su concepto, características e impacto en la "sociedad inteligente", con especial atención a los problemas que plantea en materia de responsabilidad penal y teoría del delito. Se concluye que, a medida que la tecnología avance, los riesgos asociados a los sistemas inteligentes aumentarán, lo que exigirá equilibrar beneficios y costos mediante decisiones individuales y regulaciones estatales orientadas a prevenir daños.

Bibliografía final

Ana Dobratinich, G. (2021). *Derecho y Nuevas Tecnologías*. 1º ed., Ciudad Autónoma de Buenos Aires: La Ley.

Almazán Salazar, E. (2021). *Personalidad Electrónica de los Robots. Implicaciones y Retos Legales de la Inteligencia Artificial. Propuesta Lege Ferenda de Lex Robótica*. 1º ed., España: Juristas con Futuro.

Camilo Momblanc, L. (2024). Inteligencia Artificial y Derecho Penal. ¿Será necesario un Nuevo Concepto de Delito?, en *Revista de la Facultad de Derecho y Ciencia Política de la Universidad Alas Peruanas*, Vol. 22, Nº 34, 2024, pp. 183-210.

Domínguez Henaín, D. H. (s. f.). *Módulo de Derecho Penal. Parte General*. Corrientes.

González, R. L. (2018). *Derecho Penal: Parte General*. 1º ed., Ciudad Autónoma de Buenos Aires: Astrea.

Jakobs, G. (1997). *Derecho Penal. Parte General. Fundamentos y Teoría de la Imputación*. 2º ed., Madrid: Parcial Pons.

Ley 22.278 del Régimen Penal de la Minoridad, B.O. 25/08/1980.

Quintero Olivares, G. (2017). La robótica ante el derecho penal. El vacío de respuesta jurídica a las desviaciones incontroladas, en *Revista Electrónica de Estudios Penales y de la Seguridad*, Nº 1, España: REEPS.

Roxin, C. (1997). *Derecho Penal Parte General. Fundamentos. La Estructura de la Teoría del Delito*. 1° ed., Madrid: Civitas.

Valls Prieto, J. (2022). Sobre la responsabilidad penal por la utilización de sistemas inteligentes, en *Revista Electrónica de Ciencia Penal y Criminología*, Nº 24, Universidad de Granada, España, pp. 1-35.

Sección IV

Derecho internacional

Nuevas tecnologías en el paradigma del cuidado social

Por Cristina Monereo Atienza (*)

Resumen: El presente trabajo analiza la crisis estructural de los cuidados en el marco del capitalismo neoliberal y el papel que desempeñan las nuevas tecnologías en este contexto. En primer lugar, se describe cómo la creciente demanda de cuidados, la mercantilización de la atención y la feminización precarizada de este trabajo han configurado un modelo desigual y externalizado, que transforma el cuidado en un bien de mercado en lugar de un derecho colectivo. Se introduce la noción de tecnopoder para mostrar cómo las tecnologías aplicadas al cuidado, desde la robótica social hasta la inteligencia artificial médica, no son neutrales, sino que suelen reproducir lógicas de control, eficiencia y deshumanización. Posteriormente, el texto explora el surgimiento de un nuevo paradigma del cuidado, impulsado por la teoría crítica feminista, que lo concibe como una actividad esencial para la sostenibilidad de la vida y no como una tarea privada o secundaria. Este paradigma se fundamenta en nociones como la autonomía relacional, el reconocimiento, la vulnerabilidad y la solidaridad, ampliando el concepto de cuidado hacia una dimensión social y colectiva, incluso más allá de lo humano. En este marco, las tecnologías pueden desempeñar un rol emancipador y transformador si se insertan en una ética del cuidado orientada a la interdependencia, la reciprocidad y la justicia social. Finalmente, el trabajo concluye que el reto no consiste en rechazar la tecnología, sino en reorientar su función para que, lejos de mercantilizar, refuerce las redes afectivas y comunitarias, contribuyendo a una democracia del cuidado que incluya humanos, no humanos y entornos en un entramado común de vida.

Palabras clave: nuevas tecnologías; tecnopoder; teoría feminista; cuidado social.

Abstract: This paper examines the structural crisis of care within the neoliberal capitalist framework and the role that new technologies play in this context. First, it outlines how the growing demand for care, the commodification of caregiving, and the precarious feminization of this work

(*) Catedrática de Filosofía del Derecho, Universidad de Málaga.

have shaped an unequal and externalized model, turning care into a market commodity rather than a collective right. The concept of technopower is introduced to show how technologies applied to care—ranging from social robotics to medical artificial intelligence—are not neutral, but often reproduce logics of control, efficiency, and dehumanization. The paper then explores the emergence of a new paradigm of care, developed by feminist critical theory, which conceives care as an essential activity for the sustainability of life rather than a private or secondary task. This paradigm is grounded in notions such as relational autonomy, recognition, vulnerability, and solidarity, expanding the concept of care toward a social and collective dimension that even goes beyond the human. Within this framework, technologies can acquire a transformative and emancipatory role if embedded in an ethics of care centered on interdependence, reciprocity, and social justice. Finally, the study argues that the challenge is not to reject technology but to reorient its function so that, instead of reinforcing commodification, it strengthens affective and community networks, contributing to a democracy of care that embraces humans, nonhumans, and environments within a common web of life.

Key words: new technologies; technopower; feminist theory; social care.

1. INTRODUCCIÓN: LAS CRISIS DE LOS CUIDADOS EN LA ERA TECNOLÓGICA

1.1. La crisis estructural de los cuidados

En las últimas décadas, múltiples voces están señalando la presencia de una crisis estructural en el cuidado, que no es una emergencia coyuntural sino una consecuencia estructural profunda del modelo político, jurídico y socioeconómico neoliberal.

Esta crisis se manifiesta a través de varios procesos paralelos. El primero es la creciente demanda de cuidados, provocada fundamentalmente por el envejecimiento de la población, las enfermedades crónicas y el aislamiento en las ciudades.

Luego está la disminución en el tiempo, los recursos y la capacidad social para ofrecer esos cuidados de manera sostenible y también justa. La incorporación de las mujeres al mercado laboral, la baja natalidad y la dispersión geográfica de las familias han llevado a la mercantilización de los cuidados. Esto ha generado una división de clases porque se ofrecen cuidados personalizados y atentos para quienes pueden pagar bien, mientras que aquellos que dependen de los escasos fondos públicos reciben una atención

mínima y estandarizada. El cuidado se ha transformado, por tanto, en un lujo en lugar de ser un bien común.

En último lugar, hay una invisibilización y feminización del trabajo de cuidados. Este trabajo se ha encontrado siempre relegado al contexto privado y doméstico (que era femenino y no reconocido ni remunerado). Si bien actualmente se encuentra también mercantilizado, sigue siendo realizado en su mayoría por mujeres, muchas de ellas inmigrantes, que lo llevan a cabo de manera precaria (con salarios bajos, inestabilidad, y con una carga emocional abrumadora). De hecho, estas condiciones se ven acentuadas con la uberización del trabajo de cuidados a través de plataformas que convierten la atención a mayores o dependientes en "servicios a la carta". De esta forma, el cuidado se convierte en una actividad que es posible delegar o externalizar y, sin la debida atención, puede comprometer el tejido afectivo, relacional y ético que la sostiene.

Todo esto viene a dar cuenta de que el crecimiento de los cuidados no ha ido acompañado de una redistribución justa del trabajo ni de un reconocimiento económico y político del cuidado como actividad básica para la sostenibilidad de la vida, sino que se ha generado un modelo mercantilizado, externalizado, profundamente desigual y disociado de una actividad que genera lazos afectivos. El cuidado se ha convertido en un producto del mercado en vez de entenderse como un derecho, un valor o una función colectiva básica para mantener la vida. Como enuncia Silvia Federici, el cuidado ha sido funcionalizado por el mercado, pero sigue sostenido por cuerpos invisibilizados, feminizados y racializados (Federici, 2013, 147, 205 y ss.).

La clave es que esta crisis es logística y económica, pero ante todo es moral y política (Tronto, 1993; 1998). Nuestras sociedades son dependientes de unos cuidados que no se reconocen, no se valoran y no se distribuyen. Las cadenas de cuidados globales, que son sostenidas por mujeres inmigrantes en las condiciones de explotación, son el síntoma más claro de esta asimetría estructural (sobre ello, véase Monereo Atienza, 2025).

1.2. El tecnopoder: las tecnologías no son neutras

En esta crisis estructural, las nuevas tecnologías (desde robots cuidadores, a IA médica o sensores de monitoreo remoto) aparecen como una solución funcional a la "crisis de tiempo" o "falta de mano de obra", pero al final están respondiendo al mismo patrón y reproduciendo la misma lógica: automatizar lo afectivo para sostener la eficiencia del sistema, no su justicia. Es así como el denominado tecnopoder se está desplegando en su función más opresora.

El término *tecnopoder* puede definirse en general como la forma en que el poder se articula, distribuye y ejerce *en* y *a través* de dispositivos, infraes-

tructuras y saberes tecnocientíficos (Foucault, 1976; Haraway, 1991; Barad, 2007; Preciado, 2008).

Este concepto arranca de las genealogías foucaultianas de biopoder en las que se desenmascara el dominio político-jurídico sobre los cuerpos. Por tanto, es una noción que incide en un poder opresor, ahora trasladado a la esfera privada económica que utiliza las tecnologías para el control tanto de la corporalidad como, sobre todo, de la psique humana. En esta línea, Byung-Chul Han defiende que se ha dejado atrás la biopolítica y que el poder ya no solo se ejerce sobre los cuerpos sino que pasa a ser un psicopoder que vigila, controla y conduce a los seres humanos desde dentro. Es un poder que tiene la capacidad de hacer creer que los sujetos toman libremente sus decisiones cuando en realidad no son conscientes de la guía e influencia en sus decisiones por parte de aquellos poseedores de la información recabada en la red. Así es cómo los sujetos son *cosificados* o *reificados* (Honneth, 2007, 17 y ss.). La cosificación de las relaciones humanas olvida el reconocimiento de los otros y de uno mismo como seres dignos, y con ello la cooperación participativa se vuelve muy instrumental y vacía de emociones y sentimientos humanos. Esto supone en realidad una merma de la libertad y autonomía individual solo posibilitada en relación, interconexión e interdependencia con otros.

Desde esta perspectiva, la biomedicina, las tecnologías digitales o algorítmicas son utilizadas por empresas privadas que adquieren un alto control sobre los individuos. El tecnopoder produce entonces nuevas subjetividades administradas bajo la promesa de su autonomía cuando lo cierto es que su finalidad es la guía y la eficiencia económica (en este sentido está, por ejemplo, el concepto de "farmacopornopoder" de Preciado, 2008, 46, 58). Los individuos están profundamente gobernados por un régimen tecnocientífico, que impone una visión mercantilizadora de la vida y que los hace profundamente dependentes de unos productos que no siempre necesitan.

Esta dimensión negativa del tecnopoder desvela que las tecnologías no son neutras y actualmente están insertadas en un discurso filosófico, político y socio-económico patriarcal y capitalista que persiste en conseguir la mayor eficiencia aun despojando a los individuos de su propia humanidad.

En resumen, este poder se despliega en las relaciones de cuidado aprovechando su crisis estructural e imponiéndose como una solución a los problemas generados por el propio sistema, adoptando la misma dinámica de mercantilización y eficiencia e ignorando muchas veces el papel esencial de los cuidados para la existencia u ofreciendo solo una fachada superficial de afecto.

2. El nuevo paradigma de los cuidados y el papel de las tecnologías

Para repensar el papel de las tecnologías en los cuidados es necesario insertarlas en un nuevo paradigma que cuestione la productividad como medida del valor humano, revalorice la interdependencia y reclame una política del cuidado que atraviese todas las dimensiones de lo social, lo económico, lo tecnocientífico y ambiental.

Lo prioritario es ahora hacer visible la red que sostiene a todos, poniendo el foco en los sujetos implicados en los cuidados, tanto el sujeto de cuidado, como también el sujeto cuidador. Solo así, las tecnologías podrán desembarazarse de la visión del tecnopoder en su sentido negativo, el que deshumaniza y controla a los individuos, haciéndoles creer falsamente que son liberados.

2.1. Breve genealogía del nuevo paradigma de los cuidados

La conformación del paradigma actual de los cuidados, tal y como lo presenta la teoría crítica feminista, ha sido fruto de varias luchas frente a la concepción moderna liberal y tradicional del cuidado en el que esta actividad fue manifiestamente ignorada por la política y el derecho, considerada femenina, afectiva y secundaria, además de fruto de una decisión libre de las mujeres (Kittay, 1996, 76 y ss.; Whitney, 2011, 554-574; Leach, 2014, 213 y ss.).

Frente a esta concepción tradicional, en la década de los setenta las corrientes feministas iniciaron el debate en torno a la noción de "trabajo doméstico" al que se atacaba en razón a la división sexual del trabajo realizada por el capitalismo. Se contraponía el trabajo productivo y visible de los hombres *versus* el trabajo reproductivo e invisible de las mujeres (Benería, 1979, 203-225).

Más tarde, se esgrimió el concepto de "trabajo reproductivo" con el objetivo de visibilizar la labor llevada a cabo por las mujeres en la esfera privada como un trabajo necesario socialmente y, por tanto, también productivo (Himmmelweit, 2000). El cuidado desde entonces se considera una actividad necesaria que hay que realizar. Si la persona encargada no la lleva a cabo, tienen que encargarse de ella bien otros familiares (generalmente otras mujeres de la familia que sustituyen a la mujer) o bien una persona ajena a la familia (normalmente de nuevo otra mujer quien acomete la actividad a cambio de una remuneración, Batthyány, 2015, 9). El problema entonces está en que la liberación de ciertas mujeres de una parte de la carga de lo doméstico y familiar se ha resuelto a través de otras mujeres que asumen esta tarea. Esto ha ocurrido tanto en el marco de relaciones familiares basadas en mecanismos de reciprocidad y solidaridad, como en el marco de las relaciones mercantiles atípicas que rigen en el mundo de lo privado (Tobío, 2006, 28). En el caso de la mercantilización de los cuidados surgen

además situaciones de especial vulnerabilidad, ya que las trabajadoras del cuidado suelen ser féminas en situación de pobreza y necesidad y generalmente inmigrantes en situación laboral muy precaria.

Por eso, recientemente, se ha promovido un nuevo desplazamiento conceptual hacía el término "trabajo de cuidados", que enriquece y amplía la noción de trabajo reproductivo con las tareas realizadas desde el ámbito laboral y estatal. Es una noción centrada en los aspectos positivos de generación de bienestar de los cuidados (Esquivel y otros, 2012, 11; Esquivel, 2013, 27). Este término persigue desprecarizar el trabajo de cuidados y asumir la responsabilidad social sobre esta tarea.

En esta línea se ha acuñado igualmente el término de "cuidado social", cuyo punto clave es considerar al cuidado como actividad más allá de los contextos particulares y poner el foco en la interacción entre familia, estado y mercado, esto es, en cómo se reparten las tareas y las responsabilidades entre todos estos sectores (Tobío, 2008, 87-104). En términos morales, se propone la conjunción de la ética de la justicia con la ética del cuidado (West, 1997, 22-93), dirigida a reestructurar instituciones y prácticas como las familiares, las laborales y la separación entre espacio público y privado.

Es aquí donde se da el salto revolucionario fundamental al conformarse un nuevo paradigma que invita a pensar el cuidado como un elemento esencial para la reproducción de la vida, siendo un eje estructurante de las relaciones políticas y sociales. El cuidado se concibe como una actividad mucho más general, que está en todas partes (Fisher y Tronto, 1990, 35-62), que engloba a toda la labor dirigida a reparar, mantener y hacer el mundo más habitable incluyendo a nuestro cuerpo, nuestra mente y el entorno en el que se vive y nos interrelacionamos con los demás. Todos somos interdependientes y, por tanto, la responsabilidad del cuidado es un acto colectivo (Tronto, 2015, 252-271; 1998, 15-20).

El cuidado pasa a ser un problema colectivo que necesita de la intervención estatal y que incluso da lugar a un derecho al cuidado (Pautassi, 2007; 2028, 777-742; Batthyány, 2015, 12), justificado en una nueva ética pública que reconozca la interdependencia, la vulnerabilidad y la agencia distribuida entre humanos y no-humanos.

2.2. Las bases del cuidado social: autonomía relacional, reconocimiento, vulnerabilidad y solidaridad

Una de las conceptualizaciones más influyentes de este paradigma es la realizada por Joan Tronto quien define al cuidado como una actividad cotidiana que implica "todo aquello que hacemos para mantener, continuar y reparar nuestro mundo de manera que podamos vivir en él lo mejor posible" (Tronto, 1993, 103). Con esta definición amplia, Tronto abarca

tanto la dimensión afectiva como material del cuidado, proponiendo una lectura política de su distribución social. Para ello distingue cinco fases del cuidado: *caring about, taking care of, caregiving, care-receiving* y *caring with* en las que se aprecia cómo el cuidado es un proceso de reconocimiento, relacional y situado que involucra responsabilidad frente a la vulnerabilidad, acción y reciprocidad. Es una tarea atenta, sensible y basada en el afecto, pero no es solo una tarea individual sino clave para el sustento social de la vida.

Considerado así, el cuidado demanda salir del ámbito privado para ser reconocido como pilar de la vida democrática. Hasta ahora, las sociedades neoliberales han invisibilizado y devaluado el cuidado y es necesario que se posicione como un principio organizador de las instituciones públicas. Por esta razón, Tronto defiende una "democracia del cuidado" en la que los sujetos de cuidado no sean solo los dependientes y pasen a ser todos los individuos en tanto que seres interdependientes y situados en redes de reciprocidad mutua *(Tronto, 1993, 133)*.

El foco de los cuidados se traslada entonces a varios conceptos fundamentales, rompiendo el molde de un sujeto autónomo y autosuficiente frente a otro sujeto dependiente. Estos conceptos renovadores son la autonomía relacional, el reconocimiento, la vulnerabilidad y la solidaridad recíproca.

El punto de partida es la crítica a la noción individualista moderna de autonomía en la que se ha basado erróneamente el paradigma clásico de los cuidados y que la reduce a la autosuficiencia y la mera capacidad individual de búsqueda del propio concepto del bien y la vida buena, sin interferencia alguna de factores externos. La autonomía, como la posibilidad de desarrollar unas capacidades mínimas y elaborar el propio concepto del bien y la vida buena, no se ejerce de manera autosuficiente porque el contexto social tiene un papel esencial en las elecciones del individuo. Así pues, la autonomía es capacidad y lo importante son las opciones reales que están disponibles para esa persona (Mackenzie, 2014a, 43; Sen, 1979, 195-220; Nussbaum, 1995, 61-104).

Además de capacidad, la autonomía es igualmente un estatus: el de reconocerse autónomo como los demás. Como explica Axel Honneth, este estatus se genera a partir de relaciones intersubjetivas de reconocimiento, que pueden fallar en algunos casos de discriminación o humillación (género, raciales, de personas con discapacidad, etc.…). La falta de reconocimiento afecta a la confianza en uno mismo, el respeto hacía la propia persona y la autoestima. Esto impide la plena autonomía que solamente se puede alcanzar bajo condiciones sociales de apoyo, porque el ser humano es un ser vulnerable por definición. Es decir, existe un diálogo necesario con uno mismo y con otros seres en la determinación del propio modelo del bien y la vida buena, y ese diálogo se basa en la idea de que todo ser humano es vulnerable así como en la idea de reconocimiento mutuo de esa vulnera-

bilidad que genera lazos de solidaridad (Anderson and Honneth, 2005, 131; Honneth, 2014, 165 y ss.).

Por tanto, la vulnerabilidad conlleva una dimensión positiva al acentuar la importancia del yo en conexión con los otros, lleva aparejada la idea de comunión y por tanto permite entender al ser humano como un ser interrelacionado con los demás con los que comparte la misma condición y con quienes dará lugar a redes e instituciones para afrontar esa situación adversa.

Como defiende Martha Albertson Fineman, la vulnerabilidad tiene consecuencias éticas en cuanto conlleva una apertura hacia el otro y, conjuntamente, atañe a cuestiones jurídicas de Justicia social al generar responsabilidad hacía las situaciones vulnerables. La vulnerabilidad es una característica compartida aunque es asimétrica, por lo que existe una responsabilidad directa de las instituciones político-jurídicas y sociales ante esa asimetría (Fineman, 2011, 251-275; Stalsett, 2005, 45-46).

Este concepto de vulnerabilidad es útil para reactivar la responsabilidad pública al permitir distinguir al menos tres dimensiones (Fineman, 2013, 20 y ss.; Mackenzie, 2014a, 33-59; 2014b, 15-41). En primer lugar, la vulnerabilidad es universal y una condición humana inevitable y constante en el ser humano. En segundo lugar, la vulnerabilidad se experimenta de forma diferente por cada individuo dependiendo de su contexto (en relación al ámbito médico, véase Turner, 2006, 35 y ss.). Por último, la vulnerabilidad puede ser provocada por la estructura socio-política y económica, y no depender directamente de la condición humana. En suma, existen tres tipos de vulnerabilidades: la inherente, la contextual y la patogénica. Todas ellas configuran lo que llama un "sujeto vulnerable".

Asimismo, existen otras tesis que amplían el alcance de este modelo de los cuidados. María Puig de la Bellacasa extiende el cuidado hacia lo más-que-humano, incorporando las dimensiones materiales, tecnológicas y afectivas de la existencia. Para ella, el cuidado "implica una disposición afectiva hacia el mundo que se traduce en una forma específica de atención, respuesta y responsabilidad" (Puig de la Bellacasa, 2011, 90). Su propuesta cuestiona la separación entre sujeto y objeto, naturaleza y cultura, y plantea una **ética especulativa** del cuidado donde la producción de conocimiento se ve atravesada por prácticas afectivas. Puig de la Bella Casa propone el concepto de *matters of care* para pensar cómo los ecosistemas no solamente deben entenderse en términos de eficacia, sino en función de relaciones afectivas y ecológicas.

Desde esta perspectiva, el sujeto de cuidado ya no es únicamente el ser humano vulnerable, lo son igualmente los ecosistemas, los sistemas tecnológicos, las bacterias, los suelos, las máquinas, esto es, un entramado de

entidades con capacidad de afectarse mutuamente. Puig propone entonces una ecología política del cuidado, donde el conocimiento y la ciencia no pueden ser neutrales, estando siempre implicadas en relaciones de cuidado o descuido.

Visto de este modo, el nuevo paradigma implica que el cuidado es esencial para la vida humana y también para la no humana. El cuidado comporta responsabilidad ética y pública, y debe adquirir un lugar central en las estructuras políticas, sociales y económicas. Se trata de deconstruir la lógica patriarcal que ha separado lo racional de lo afectivo, lo público de lo privado, lo humano de lo no humano, en una ruptura de los binarismos clásicos en los que se ha basado la mirada moderna de la vida.

En este modelo el sujeto de cuidado emerge como una figura compleja. Ya no es un sujeto pasivo, sino una entidad que participa en relaciones dinámicas de afecto, poder, agencia y reciprocidad. El cuidado no es unidireccional, es una práctica relacional, situada y comprometida con la sostenibilidad de la vida en todas sus formas.

2.3. Las nuevas tecnologías para los sujetos de cuidado

Las tecnologías pueden adquirir un papel esencial en los cuidados si se replantea su función en este nuevo marco de los cuidados. Como se afirmaba más arriba, hasta ahora se han presentado como una solución mercantilizada para perpetuar el modelo clásico que amplifica la responsabilidad privada y no considera esta actividad como central y colectiva en el marco del sistema del capitalismo tardío.

Un buen ejemplo de ello está en el campo de la robótica social, cuyo desplazamiento del ámbito industrial a la esfera de la asistencia doméstica tiene que ver con la creciente necesidad de mano de obra en el sector asistencial (Dumouchel y Damiano, 2017; véase el análisis de este trabajo realizado por Fersini, 2025).). Las empresas desarrolladoras de robots han visto en este sector una fuente de ingresos en la que invertir, aprovechando la coyuntura ocasionada por la disminución de la asistencia informal provista por las familias y el incremento de los costes económicos y burocráticos de contratación de las trabajadoras del cuidado. Desde la consideración de estas personas como personal laboral con plenos derechos (reclamación justa, por otra parte), muchos demandantes no puedan costearse los cuidados como desearían quedando a merced de la caridad de la escasa y precaria ayuda pública.

El uso de la robótica en el cuidado aparece como una alternativa frente a esa situación. Aunque está aún en período de experimentación, existen ya algunas propuestas dirigidas al seguimiento, tratamiento y entretenimiento de personas vulnerables, especialmente ancianos. Por ejemplo, está el ro-

bot Paro (una foca terapéutica utilizada en diversos estudios para mejorar la salud mental y reducir el estrés en personas mayores); también está el robot AIBO (un perro que puede ser utilizado como mascota digital, ofreciendo compañía y estimulación); el robot Elli Q (un robot de compañía con voz y pantalla táctil que puede recordar citas, gestionar medicación y realizar videollamadas); Garmi (un robot que puede servir comida, abrir botellas de agua y brindar apoyo en caso de caídas, también puede ayudar en videollamadas con familiares); así entre otros muchos.

El problema de estas propuestas no está en las tecnologías *per se*, cuyo despliegue tienen que ver igualmente con un cambio cultural más amplio de hecho, a día de hoy sería tremendamente ingenuo negar su existencia y sus posibilidades para el futuro del cuidado. No obstante, su incorporación a este ámbito debe hacerse con la mirada crítica necesaria para frenar las disfuncionalidades del sistema socio-económico y no perder de vista la base afectiva aneja a esta labor.

Por eso, es interesante recuperar discursos alternativos sobre el tecnopoder que tengan una mirada más benévola hacia las tecnologías. Frente al discurso patriarcal y neoliberal, el pensamiento crítico feminista ha incidido en una visión completamente diferente de las tecnologías dotando al tecnopoder de una función transformadora y emancipadora. Autoras como Karen Barad o Donna Haraway muestran que el cuerpo no es simplemente moldeado por el poder, sino que está *materializado* por ensamblajes tecnopolíticos que pueden dar lugar a nuevas subjetividades más inclusivas y superadoras del deficiente sujeto moderno de derechos (véase Haraway, 1999, 150; Barad, 2007, 146).

Ciertamente, estas tesis se introducen en un modelo que se ha denominado "posthumano" cuyas propuestas son provocadoras y seguramente problemáticas en el ámbito jurídico, pero son formulaciones sugerentes y enriquecedoras en relación a los cuidados porque inciden en *otra visión más amable* de las tecnologías de tal manera que estas, estando inevitablemente presentes en la existencia, pueden adquirir un papel esencial en un modelo en el que cuidar no sea un acto privado propio de la esfera femenina, profundamente sentimental e irracional. Los cuidados se reconceptualizan como un acto social colectivo reconocido públicamente, que genera una red de relaciones afectivas entre los seres humanos y también en simbiosis entre estos y la naturaleza (Puig de la Bella Casa, 2017).

El rol de las tecnologías permuta y se configuran como herramientas al servicio de un modelo afectivo e interdependiente, como un elemento más en la red de cuidados que promueve la diversidad y la ayuda mutua entre lo humano y lo no humano.

De este modo, el foco tecnológico cambia y está, en primer lugar, en las personas en situación de vulnerabilidad que han de ser cuidadas, pero dentro de un forma de pensar el sujeto cuidado no como vulnerable pasivo, sino como agente situado, afectado y afectante en una red densa de interdependencia, una subjetividad que rompe con las lógicas neoliberales del individuo autónomo y autosuficiente. Además, el foco tecnológico se encuentra, asimismo y en segundo lugar, en el sujeto cuidador (todavía mujeres). Los robots y otras tecnologías (el conocido botón rojo, las cámaras de vigilancia, etc..) no tienen que verse como sustitutos de las tareas de cuidado; son asistentes al propio cuidador, facilitando y extendiendo el trabajo allí donde no es posible su realización.

Además, si los cuidados son considerados un acto social colectivo (reconocido públicamente) que genera una red de relaciones afectivas entre los seres humanos y también en simbiosis entre estos y la naturaleza, entonces las tecnologías pueden introducirse como un tercer pilar que promueva la ayuda mutua e incluso sirvan para el reconocimiento de la diversidad.

Al hilo de esta última idea, seguramente no se pueda obviar el debate de fondo planteado por el posthumanismo. Hay que tener en cuenta que se está trabajando en la humanización de los robots cuidadores con resultados cada vez más sorprendentes, además de que ya existen casos de seres sintéticos (medio-humanos) producidos gracias a los nuevos avances técnicos. Este debate ya se está produciendo en el ámbito de las personas con discapacidad y las prótesis insertadas en sus cuerpos.

Todo ello provoca una reveladora discusión acerca de la definición de lo humano que puede conllevar un replanteamiento (para algunos una "superación") del sujeto de derechos, basándose en lo innecesario de plantear el tema de la igualdad de derechos en un mundo en el que todos son uno y están conectados inseparablemente (Haraway, 2019). Este nuevo paradigma revoluciona el humanismo liberal al comprenderse la existencia de un lazo común que une al conjunto de los seres en un *continuum* inseparable entre lo humano, lo natural, y lo tecnológico. De esta manera el sujeto no es entendido de manera individual o separada, y la interrelación entre todos los seres y toda la Naturaleza sería la esencia que construye al sujeto. El sujeto no es comparativamente igual o diferente en derechos. Se conformaría como una entidad transversal diferenciada que engloba lo humano, animales y toda la Tierra (Braidotti, 2013, 10; 2022, 85 y ss.; Haraway, 1999, 121-163). Todos los seres son intrínsecamente co-construcciones inseparables. No se podría hablar de un proceso de humanización de otros, ni de una especificación o extensión de derechos a otros seres al ser todos parte de ese sujeto.

La idea en el ámbito jurídico no es deshacerse de forma drástica de las conquistas ilustradas (entre ellas de los derechos), pero sí debe ser estar abiertos al debate de todas posibilidades planteadas dentro de epistemolo-

gías diferentes, algo que ya se está haciendo en muchos lugares en torno a la subjetividad y los derechos de la naturaleza [(1)].

3. A MODO DE CONCLUSIÓN

El nuevo paradigma del cuidado social invita a imaginar el cuidado más allá del antropocentrismo, del capacitismo, del binarismo afecto-tecnología. Recuerda que cuidar no es solamente un gesto sentimental, sino una práctica de mundo, es una apuesta ética y política por la sostenibilidad de lo común, incluso (y especialmente) en entornos tecnológicos.

En esta línea, se trata de recuperar el concepto de tecnopoder en su forma transformadora. El pensamiento tecnocientífico se torna un instrumento útil dentro del nuevo paradigma de los cuidados entendidos como una lógica organizativa de la Vida en común y no como recurso escaso que se optimiza.

Todo ello plantea un debate crítico del modelo todavía liberal-capitalista (no lo suficientemente corregido) del sujeto y los derechos impuesto desde la modernidad. La meta es introducir poco a poco las transformaciones oportunas que caminen hace una idea de sujeto y de derechos que comprenda la importantísima interrelación, relación, reciprocidad, correspondencia y complementariedad de todos y de todo (Ávila Santamaría, 2011).

BIBLIOGRAFÍA

Anderson, Joel and Honneth, Axel (2005). "Autonomy, Vulnerability, Recognition, and Justice". En Christman, Joseph, and Anderson, Joel (eds.), *Autonomy and the Challenges to Liberalism: New Essays*. New York: Cambridge University Press, 127-149.

Ávila Santamaría, Ramiro F. (2011). "El derecho a la Naturaleza: fundamentos". En Espinosa Gallegos-Anda, Carlos y Pérez Fernández, Camilo (eds.). *Los derechos de la naturaleza y la naturaleza de sus derechos*. Quito: Ministerio de Justicia, derechos humanos y culturales, 35-73.

Barad, Karen (2007). *Meeting the Universe Halfway: Quantum Physics and the Entanglement of Matter and Meaning*. Durgham (North Carolina, USA): Duke University Press.

Batthyány, Karina (2015). *Las políticas y el cuidado en América Latina. Una mirada a las experiencias regionales*. CEPAL- Serie Asuntos de Género.

(1) Piénsese, por ejemplo, en las constituciones ecuatoriana y colombiana, así como la jurisprudencia de esos países al respecto.

Benería, Lourdes (1979). "Reproduction, Production and the Sexual Division of Labor". *Cambridge Journal of Economics*, Vol. 3(3), 203–225.

Braidotti, Rosi (2013). *The Posthuman*. Cambridge, UK: Polity Press.

- (2022). *Feminismo posthumano*. Trad. Serra Lopes- Sion. Barcelona: Gedisa, 2022.

Dumouchel, Luisa y Damiano, Paul (2017). Living with Robots. Trad. M. Deveboise. Cambridge: Harvard Universitiy Press.

Esquivel, Valeria. (2013). *El cuidado en los hogares y en las comunidades*. Informes de Investigación Oxfam.

Esquivel, Valeria; Faur, Eleonor y Jelin, Elisabeth (2012). "Hacia la conceptualización del cuidado: familia, mercado y estado". En Esquivel, Valeria, Faur, Eleonor y Jelin, Elisabeth. (eds.), *Las lógicas del cuidado infantil. Entre las familias, el mercado y el Estado*, Editorial IDES, 11-43.

Federici, Silvia (2013). *Revolución en punto cero. Trabajo doméstico, reproducción y luchas feministas*. Trad. C. Fernández y P. Martín. Madrid: Traficantes de Sueños.

Fersini, Maria Pina, "El desplazamiento de la robótica del ámbito industrial al ámbito social y las nuevas fronteras del cuidado: desajustes, desafíos y análisis literarios prospectivos para nombrar el porvenir" (en prensa).

Fineman, Martha A. (2011). "The Vulnerable Subject and The Response of The State". *Emory Law Journal,* vol. 60, n. 2, 251-275.

- (2013). "Equality, Autonomy, and the Vulnerable Subject in Law and Politics". En Fineman, Martha A. and Grear, Anne, *Gender in Law, Culture, and Society: Vulnerability: Reflexions on a New Ethical Foundation for Law and Politics*. Surrey (England)/Burlington (USA): Ashgate Publishing Ltd, 13-27.

Fisher, Bérénice y Tronto, Joan (1990). "Towards a Theory of Caring". En Abel, Emily and Nelson, Margaret (eds.), *Circles of Care: Work and Identity in Women's Lives*. New York: State University of New York Press, 35-62.

Foucault, Michel (1995, 2ª ed.). *Historia de la sexualidad, Vol. I: La voluntad de saber*. Trad. T. Segovia, Madrid: Siglo XXI Editores.

Han, Byung-Chul (2014), *En el enjambre*. Trad. R. Gabás. Barcelona: Herder.

Haraway, Donna (1991). *A Cyborg Manifesto: Science, Technology, and Socialist-Feminism in the Late Twentieth Century*. En *Id. Simians, Cyborgs, and Women: The Reinvention of Nature*. Nueva York: Routledge.

- (1999). "Las promesas de los monstruos: una política regeneradora para otros inapropiados/bles". *Política y Sociedad*, núm. 30, 1999, 121–163.

-(2019). *Seguir con el problema: generar parentesco en el Chthulucene*. Trad. H. Torres. Madrid: Consonni.

Himmelweit, Susan (2000). *Inside the household from labor to care*. London: Macmillan.

Honneth, Axel (2007). *Reification: A New Look at an Old Idea*. Oxford: Oxford University Press.

- (2014). *El derecho de la libertad. Esbozo de una eticidad democrática*. Trad. G. Calderón. Madrid: Clave intelectual, 165 y ss.

Kittay, Eva Feder (1996). Love's Labor: Essays on Women, Equality, and Dependency. New York: Routledge.

Leach, Joan (2014). "The Politics of Care in Technoscience". En Van der Velden & White (eds.), *Technoscience and Postphenomenology*, Lexington Books, 213-229.

Mackenzie, Catriona (2014a). "The Importance of Relational Autonomy and Capabilities for the Ethics of Vulnerability". En Mackenzie, Catriona, Rogers Wendy and Dodds, Susan, *Vulnerability: New Essays in Ethics and Feminist Philosophy*. Oxford/ New York: Oxford University Press, 33-59.

- (2014b). "Three Dimensions of Autonomy: A Relational Analysis". En Piper, Mark and Veltman, Andrea (eds.), *Autonomy, Oppresion and Gender*. Oxford/New York: Oxford University Press, 15- 41.

Monereo Atienza, Cristina (2025), "La subjetivización de los cuidados: vulnerabilidad, inserseccionalidad y reconocimiento". En Vila Viñas, David (coord.), *Nuevos escenarios del cuidado desde una perspectiva socio-jurídica. Entre el reconocimiento de derechos y la mercantilización en el empleo de hogar y de establecimientos turísticos*, Madrid: Dykinson, 217-238.

Nussbaum, Martha C. (1995), "Human capabilities, Female Human Being". En Nussbaum, Martha C. and Glover, Jonathan (eds.), *Women, Culture and Development. A Study of Human Capabilities*, Oxford: Clarendon Press, 61-104.

Pautassi, Laura (2007). *El cuidado como cuestión social desde el enfoque de derechos. Serie mujer y desarrollo,* CEPAL, núm. 87.

Preciado, Paul B. (2008). Testo Yonqui: Sexo, drogas y biopolítica. Barcelona: Espasa Libros.

Sen, Amartya (1979). "Equality of what?", *The Tanner Lecture on Human Values*. Stanford University, 195-220.

Stalsett, Sturla J. (2005). "Vulnerabilidad, dignidad y justicia: valores éticos fundamentales en un mundo globalizado". En Kliksberg, Bernardo (ed.), *La agenda ética pendiente de América Latina*. Buenos Aires: Fondo de Cultura Económica de Argentina, 43-59.

Tobío Soler, Constanza (2006) "Nuevas formas familiares: Las madres que trabajan". *Paisajes: Revista De Pensamiento Contemporáneo*, núm. 19, 25-32.

- (2008). "Redes familiares, género y política social en España y Francia". *Política y Sociedad*, vol. 45, núm. 2, 87-104.

Tronto, Joan (1993). Moral Boundaries: a Political Argument for an Ethic of Care. New York: Routlege.

- *(1998).* "An ethic of care". *Ethics and Aging: Bringing the Issues Home*, vol. 22, núm. 3, 15-20.

- (2015). "Theories of care as a challenge to Weberian paradigms in social science". En Engster, Daniel y Hamintong, Maurice, *Care Ethics and Political Theory*. Oxford: Oxford University Press, 252-271.

Turner, Bryan S. (2006). *Vulnerability and Human Rights*. Pensilvania: Pensilvania State University Press/ University Park.

West, Robin (1997). *Caring Justice*. New York/London: New York University Press.

Whitney, Lisa (2011), "The Politics of Care". *Hypatia*, 26(3), 554–574.

The European Union Artificial Intelligence Act and the elevated role of soft law

Por Stanley Greenstein (*)

Abstract: This article examines the European Union Artificial Intelligence Act (EU AI Act) and highlights the increasingly significant role of soft law in regulating AI technologies. It explores the Act's main objectives, including the establishment of a risk-based framework, transparency obligations, and accountability mechanisms for AI systems. The analysis underscores how soft law instruments —such as guidelines, standards, and codes of conduct— complement binding regulations by promoting ethical practices, innovation, and adaptive governance. The paper argues that the interplay between hard and soft law is essential to ensure both legal certainty and technological flexibility, paving the way for a more robust and dynamic regulatory ecosystem in the EU.

Keywords: Artificial Intelligence Act; European Union law; soft law; AI governance; ethical standards; risk-based regulation; compliance.

El Acta de inteligencia artificial de la Unión Europea y el papel elevado del soft law

Resumen: Este artículo analiza el Artificial Intelligence Act de la Unión Europea (UE) y subraya el papel cada vez más relevante del *soft law* en la regulación de las tecnologías de inteligencia artificial. Examina los principales objetivos del Acta, entre ellos el marco basado en el riesgo, las obligaciones de transparencia y los mecanismos de rendición de cuentas para los sistemas de IA. El análisis destaca cómo los instrumentos de *soft law* —tales como directrices, normas y códigos de conducta— complementan las regulaciones vinculantes al fomentar prácticas éticas, innovación y gobernanza

(*) Jur. Dr. Associate Professor (Docent) in Law and Information Technology at the Faculty of Law, Stockholm University, Sweden. He is Chairman of the Board, Swedish Law and Informatics Research Institute (IRI, <https://irilaw.org/>) and Chairman of the Trust for Legal Information (<https://rattsinfo.se>). His main area of expertise is the study of the interaction between law and technology, with a focus on the legal regulation of emerging technologies including machine learning and artificial intelligence, promoting proactive legal design, and studying the legal and ethical implications of the increased reliance on digital technologies for society. Orcid: <https://orcid.org/0000-0002-0694-768X>. greenstein@juridicum.su.se

adaptativa. El trabajo sostiene que la interacción entre el *hard law y el soft law* es fundamental para garantizar seguridad jurídica y flexibilidad tecnológica, impulsando así un ecosistema regulatorio más sólido y dinámico en la UE.

Palabras clave: Artificial Intelligence Act; derecho de la Unión Europea; *soft law*; gobernanza de la IA; estándares éticos; regulación basada en el riesgo; cumplimiento normativo.

1. Introduction

On the 1st of August 2024 the eagerly anticipated Artificial Intelligence Act (AI Act) came into effect. (1) The Artificial Intelligence Act is unique to the extent that it is the first hard law worldwide to define the concept artificial intelligence, albeit via the concept 'AI system'. It is argued that a trend that is gaining traction in relation to lawmaking and the attempt to regulate emerging technologies such as machine learning and artificial intelligence, is the increased reliance on soft law instruments as an integral compliment to the actual legal texts of the legislative frameworks. In other words, it no longer suffices to rely merely on the literal reading of legislation and a large portion of the normative law-making process is in turn encapsulated the soft law-making process. This trend is not new and soft law instruments have always been relied upon to explain legislative texts. Yet it is becoming increasing apparent, in the face of the attempt to legislate emerging digital technologies that are characterised by mathematical complexity and 'black box' attributes, that the black-letter law on its own does not suffice. Legislation has always had the trait of sometimes being vague or open to multiple interpretations. This is in part a technique whereby a discretional leeway is intentionally conferred upon the judiciary by the legislator. However, it is becoming more predominant as legal development lags behind technological development in the context of regulating emerging technologies and considering the technically complex ecosystem within which these technologies operate, there is an increasing reliance on the development of soft law instruments in order to interpret the ambit of legislative acts regulating emerging technologies, thereby making the law in general more responsive to the challenges posed by technology.

This article briefly examines the trend within the context of the European Union of making the legislating process conditional upon the utilisation of soft law instruments. This encompasses the strategy of making

(1) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance) PE/24/2024/REV/1*OJ L, 2024/1689, 12.7.2024*.

the laws that regulate complex technologies, general and technologically neutral, leaving the details, especially the technical details, to the bodies that develop the soft law. More specifically, this article investigates how this trend is occurring in relation to the development of standards as incorporated within the Artificial Intelligence Act. The main aim of this article is to illuminate the above-mentioned trend and elevate the extent to which soft law instruments are increasingly being relied upon as part of the legislative process.

2. The New Legislative Framework

An initial point of departure is the recognition of the fact that the trend towards utilising soft law instruments in order to develop the meaning of legislative texts is based on a conscious strategy of the European Union. In promoting the internal European market but also attempting to regulate products being placed on the European market, a new legislative strategy was termed 'The New Legislative Framework' (NLF). It is recognised as a cornerstone of the European Union's approach to regulating products within the Single Market. Introduced in 2008 through Regulation 765/2008 (setting out the requirements for accreditation and the market surveillance of products) and Decision 768/2008/EC (on a common framework for the marketing of products, which includes reference provisions to incorporate in product legislation revisions) the NLF was designed to bring greater coherence, transparency, and effectiveness to EU product legislation. (2) Before its adoption, sector-specific rules often varied in structure and terminology, creating confusion for businesses and uneven enforcement across Member States. The NLF responded to these shortcomings by creating a common toolbox of legal concepts and procedures that can be applied across different sectors, thereby making EU product legislation more consistent and reliable. (3) A legislative technique following the NLF is the promulgation of secondary legislation from the EU that are broad in nature, the idea being that they will be supplemented by the creation of harmonised standards. (4) In other words, whereas legislation previously could be described as relatively detailed, e.g. in relation to technical specifications, the NLF promoted the transition of primary European Union legislation (Directives and Regulations) to slimmed down frameworks, where general principles were laid down with the intention that these would be complemented by soft law

(2) European Commission, 'New Legislative Framework' <https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en?utm_source=chatgpt.com> (accessed 19 September 2025).

(3) Ibid.

(4) CEN, 'The "New Approach"' <https://boss.cen.eu/reference-material/guidancedoc/pages/newapproach/> (accessed 19 September 2025).

instruments that in tur laid down the technical specifications, usually in a more technical language.

An example of the application of the NLF is seen in relation to a Directive from the European Union regulating the accessibility to websites and mobile applications of public sector bodies (Web Accessibility Directive). (5) Recital 37 states that four principles must be adhered to, namely perceivability, operability, understandability and robustness. It the briefly outlines what these principles entail:

[Perceivability, meaning that information and user interface components must be presentable to users in ways they can perceive; operability, meaning that user interface components and navigation must be operable; understandability, meaning that information and the operation of the user interface must be understandable; and robustness, meaning that content must be robust enough to be interpreted reliably by a wide variety of user agents, including assistive technologies. (6)

Recital 37 then proceeds to state that the principles of accessibility can be achieved by adhering to testable criteria forming the basis of the European standard EN 301 549 V1.1.2 'Accessibility requirements suitable for public procurement of ICT products and services in Europe' (2015-04) [European standard EN 301 549 V1.1.2 (2015-04)]. (7)

The notion of the conformity assessment is an integral aspect of the NLF. A product must undergo a so-called conformity assessment before it is placed on the market, which occurs by means of it being demonstrated that it meets all legislative requirements, including testing, inspection and certification, with the applicable product legislation specifing the procedure for each product. (8) A manufacturer can only place a product on the EU market when it meets all the applicable requirements with the European Commission's main objective being to help ensure that unsafe or otherwise non-compliant products do not find their way to the EU market. (9) Finally, as part of the conformity assessment, the manufacturer must draw up an declaration of conformity containing all information to identify the product, the legislation according to which it is issued, the manufacturer or the

(5) Directive (EU) 2016/2102 of the European Parliament and of the Council of 26 October 2016 on the accessibility of the websites and mobile applications of public sector bodies (Text with EEA relevance) OJ L 327, 2.12.2016, 1–15 <https://eur-lex.europa.eu/eli/dir/2016/2102/oj/eng> (accessed 29 September 2025).

(6) Ibid, Recital 37.

(7) Ibid.

(8) European Commission, 'Conformity Assessment' <https://single-market-economy.ec.europa.eu/single-market/goods/building-blocks/conformity-assessment_en> (accessed 29 September 2025).

(9) Ibid.

authorised representative, the notified body if applicable, a reference to harmonised standards or other normative documents, where appropriate. (10)

The NLF is not a static initiative and as of 2022 an evaluation highlighted the need for its modernization in order to, among other things, address the digital transformation. (11) In other words, it is a process dynamic in nature and continually being updated in relation to legislative demands.

3. HARD LAW VERSUS SOFT LAW

Two concepts that to a certain degree define each other and that are relevant in the context of this article are that of 'hard law' on the one hand and 'soft law' and on the other hand. The term hard law can alternatively be referred to as 'traditional law' or 'black letter law' an in essence refers to the central sources of law that we usually have in mind when conceptualising the law. Soft law, while it has no universal definition and can be difficult precisely to define, can be defined in terms of its characteristics. (12) In the field of international law, soft law is argued to have three main characteristics: first, it is not binding, second, it consists of general norms or principles, which are open-textured and general in their content and wording and third, it is not enforceable through binding dispute resolution. (13) This can be seen in relation to traditional law or hard law, which is binding and based on rules that incorporate specific commitments. (14) The characteristic nature of soft law is demonstrated by two additional concepts, namely, 'legal soft law' and 'non-legal soft law'. Legal soft law is comprised of non-binding or voluntary codes of conduct as accepted by international and regional organisations. In other words, it is made up of binding regulations with a vague content and with no concrete obligations being conferred. (15) Non-legal soft law is described as statements by individuals, which promote general principles. (16) Put another way, it provides advice and guidelines, sometimes including the ethical dimension, yet remains outside the traditional development of the law. (17)

---

(10) Ibid.

(11) Product Compliance Institute, 'EU: Product legislation – revision of the New Legislative Framework (NLF)' <https://www.productcomplianceinstitute.com/2025/07/22/eu-product-legislation-revision-of-the-new-legislative-framework-nlf/> (accessed 19 September 2025).

(12) Van der Sluijs, Jessika, *Soft Law – an International Concept in a National Context*, in Wahlgren, Peter (ed), *Soft Law*, Scandinavian Studies in Law, Volume 58, Stockholm Institute for Scandinavian Law, 2013, 287.

(13) Boyle, A. E., 'Some Reflections on the Relationship of Treaties and Soft Law', International and Comparative Law Quarterly, Vol. 48, 1999, 901-902.

(14) Ibid.

(15) Van der Sluijs (n 9) 286.

(16) Chinkin, C. M., 'The Challenge of Soft Law: Development and Change in International Law', International and Comparative Law Quarterly, Vol. 38, 1989, 851.

(17) Van der Sluijs (n 9) 286.

Consequently, having the 'soft law' in focus, it is a term used to represent that body of law falling outside of traditional law, yet which has a regulatory function. Whereas the hard law traditionally would reference central sources of law, such as laws, preparatory works, case law and doctrine, the soft law notion represents other regulatory instruments that have not attained the hierarchical level associated with hard law, yet play a regulatory role, e.g. codes of conduct, standards, communications, opinions and to a certain degree even technology itself. (18)

4. STANDARDISATION

Standards have an important function within society. They promote economic development and enable globalization. For example, the most well-known standard is the A4 size paper used for printing. This standard has many advantages as the producers of printers know that their printers will need to be able to load and print on this size paper and it prevents organizations that require printing services with having to deal with the logistics of having different printers only working with different sized paper. Standards are used every day by businesses, manufacturers, public bodies and other organisations as a tool for managing vital issues such as trade, regulation, quality, health and safety, new technologies energy efficiency, environmental impact, connectivity and interoperability. (19)

The main aim of standardisation, therefore, is to reduce costs by and increase trade by means of coming to a common understanding. (20) In other words, standards are an important tool in business as they facilitate communication between the different transacting parties. They essentially are a common language for use between the suppliers and producers of products, services and processes so that there is a common understanding regarding the requirements that these products, services and processes should fulfil. (21) Therefore, not only do standards facilitate business, especially in an era of globalization, but they also ensure that the products produced for consumption are in fact safe. There are different types of standards that have different functions: terminology standards (create a common language); compatibility standards (reduce variation and ensure that different parts fit together); performance standards (specify operation requirements for, health, safety and the environment); measurement and test standards (products can be tested using a common measure) and managements stan-

(18) Lessig, Lawrence, *Code Version 2.0* (Basic Books 2006). Here Lessig refers to the regulatory modalities of law, norms, the market and architecture (representing technology).

(19) Danish Standards Foundation, *A World Built on Standards – A Textbook for Higher Education*, 2015 <https://www.ds.dk/media/px5jhney/a-world-built-on-standards.pdf> 3.

(20) European Commission, 'Guide for referencing standards in public procurement in Europe', CEN, 2018, 5.

(21) Danish Standards Foundation (n 15) 33.

dards (a tool for organizations top effectively manage their efforts for improvement e.g. information security). (22)

Consequently, not only do standards promote commerce, but they are an important tool for coordination how society is organized and also for protecting society from risks and vulnerabilities. It is with this in mind that standards fulfil the aims of the European Union, namely promoting the internal market and protecting the consumer from the risks associated with emerging technologies. Standards are therefore increasing in importance as society becomes more complex and technical in nature. This is to ensure that technology has a certain degree of robustness but also ensures the interoperability of these technologies. However, as the technology becomes more complex, it is becoming more difficult for most people to have meaningful insight and this means that it is becoming more difficult to place demands – legal or other – on these complex technologies. This is important from the legal perspective, and a catalyst for the legislative developments referred to above, where instead of a law demanding that a certain technology have certain specific characteristics, the law merely states that a technology should be built or developed according to a specific standard.

4.1. What is a Standard?

Having examined the importance of standards above, ii is worthwhile to address what exactly a standard is. There are a number of definitions of what a standard is and a regular dictionary definition describes it as, 'something set up and established by authority as a rule for the measure of quantity, weight, extent, value, or quality'. (23) Alternatively, standards can be described in terms of their function: [s]tandards are a way of communicating –a kind of common language– in the form of a technical specification. Often it is an ability to communicate across languages to avoid misunderstandings. Standards only exist because we use them, and because we benefit from using them. If they did not bring value, their justification for existing would no longer apply. (24)

The ISO/IEC also has a definition of standards, namely a, 'document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context'. (25) The EU describes standards as:

---

(22) Ibid 33-35.

(23) Merriam-Webster online dictionary, definition of 'standard' <https://www.merriam-webster.com/dictionary/standard> (accessed 29 September 2025)

(24) Danish Standards Foundation (n 15) 13.

(25) ISO/IEC Guide 2:2004(en), 'Standardisation and related activities – General vocabulary' <https://www.iso.org/obp/ui/en/#iso:std:iso-iec:guide:2:ed-8:v1:en> (accessed 29 September 2025).

> "[...] a technical specification, adopted by a recognised standardisation body, for repeated or continuous application, with which compliance is not compulsory, and which is one of the following: (a) 'international standard' means a standard adopted by an international standardisation body; (b)'European standard' means a standard adopted by a European standardisation organization; (c) 'harmonised standard' means a European standard adopted on the basis of a request made by the Commission for the application of Union harmonisation legislation; (d) 'national standard' means a standard adopted by a national standardisation body". (26)

Finally, according to the German Institute for Standardisation (DIN) a standard is, '[...] a document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results.' (27) Therefore, as is evident from all the above definitions, standards are a mechanism that is voluntary in nature that are used to facilitate the attainment of objectives by means of communication. The objective may be cost effectiveness in commerce but also determining appropriate technical levels for goods and services in order that the health and safety of society is put first:

> "The term 'standardisation', therefore, refers to the process of developing a standard. It is defined as, '... a systematic process by which tangible or intangible subjects are harmonized by the joint efforts of the stakeholders for the benefit of society as a whole. Because of its well-proven full consensus processes, standardisation has a high democratic legitimation and complies with anti-trust legislation'" (28).

Therefore, standards are important societal instruments that facilitate communication in the commercial context but that also ensure that the risks with technological development are identified and mitigated. They can be described as the best way of doing something as agreed on by many experts from around the world.

4.2. Two Types of Standards

Two terms often mentioned in the context of standards are the 'de facto standard' and 'de jure' standard. De facto standards are standards that have received a certain degree of acceptance and dominance by virtue of market forces or a certain level of public acceptance, this taking place without the standard having been developed via formal mechanisms. These standards are usually widely accepted by a certain industry and its customers, and is

---

(26) Regulation (EU) No 1025/2012.

(27) German Institute of Standardisation, 'Introduction to Standardisation' <https://www.din.de/resource/blob/195038/64b75612aae6d6e7341e815becadb5d9/an-introduction-to-standardisation-data.pdf> 13.

(28) Ibid.

therefore also referred to as market-driven standards. (29) De jure standards, in contrast to de facto standards, refer to standards that have received this status by law or some official practice or procedure i.e. their existence is endorsed by formal standards organizations, which ratifies it and gives it a stamp of approval. (30) Consequently, de jure standards can be said to have the following characteristics: they are voluntary and market-driven, they come about via consensus and they are recognized by an authorized body. (31) For example, de jure standards have a more formal in nature having been developed via a formal standardisation body. It is within this context that the standards referred to within legislative texts developed by the European Union can be considered 'de jure' standards. The harmonised standards referred to in the European Union legal frameworks are given a certain degree of legitimacy due to the fact that they are specifically referred to in the legislative documents as a legitimate mechanism to be used in order to interpret the black-letter-law laid down in the legislative text.

4.3. How Standards Come About

Standards can come into being for a number of reasons. Firstly, there may be a need for that standard, e.g. the need may be to protect people by outlawing the use of certain chemicals or there may be a need to standardise the size of paper in order to simplify the printing processes of organizations. The second alternative, especially in relation to 'de jure' standards, is that the legislator, via a legislative text such as a European Union Directive or Regulation, promotes the development of such a standard. The main actors in the standardisation process are the various standardisation entities that occur on the national, regional and international levels. If it is identified that this need for a standard applies only to one country then its development will be promoted at the national level. (32) However, should the need arise on the regional level then the corresponding standardisation organizations will promote the development of standards from within these standardisation coordinating actors.

Standardisation bodies can be international, regional and national. International bodies include the International Organization for Standardisation (ISO) (33) and the International Electrotechnical Commission (IEC) (34), regional bodies include the European Committee for Standardisation

(29) Electronic Design, *What's the Difference Between De Jure and De Facto Standards?* <https://www.electronicdesign.com/technologies/embedded-revolution/article/21796209/whats-the-difference-between-de-jure-and-de-facto-standards> (accessed 29 September 2025).

(30) Ibid.

(31) Danish Standards Foundation (n 15) 13.

(32) European Commission (n 17).

(33) <https://www.iso.org/home.html> (accessed 29 September 2025).

(34) <https://www.iec.ch/> (accessed 29 September 2025).

(CEN) [35] and the European Committee for Electrotechnical Standardisation (CENELEC) [36]and finally national bodies include entities such as the German Institute for Standardisation (DIN, Germany) [37] and the Swedish Institute for Standards (SIS, Sweden) [38] to name but a few. The International Organization for Standardisation (ISO) is an independent non-governmental international organization to which membership is granted. [39] Through its members, it 'it brings together experts to share knowledge and develop voluntary, consensus-based, market relevant International Standards that support innovation and provide solutions to global challenges'. [40]

An initial examination concerning how standards are established is facilitated by identifying the parties that are involved in the standardisation process and their roles. In addition, the process of the development of a standard is the result of a certain amount of interaction between the different standardisation organizations on an international, regional and national level. For example, in the European context, all standards adopted by CEN must in turn be adopted by the national standardisation bodies. Furthermore, standards adopted on the international level can automatically be adopted on the European level after a voting process. Then there is the national delegation principle that states that with the development of a European standard, each country can send a delegation to represent that country's viewpoints. To facilitate this, the committees established on a European or International level to develop a standard are mirrored in the national standardisation bodies. Also, the national body can take on the position of a secretariat of a working committee at the European or international level, thereby giving the national body lots of influence in the standardisation process.

If one takes ISO as an example, then the following entities are relevant: the 'experts' are the people that will use the standards and be affected by then, these people coming from industry, government consumer organizations, academia and non-government organization to mention but a few; the 'members', comprising the national standardisation bodies, have the role of identifying the experts from around the world and finally, the standardisation organization, ISO in this case, coordinates the development process and publishes the standards. [41] Most of the work takes place in the committees. For example, the ISO working group five technical committee, which includes international specialists from data protection authorities, security

(35) <https://standards.cen.eu/> (accessed 29 September 2025).
(36) <https://www.cenelec.eu/> (accessed 29 September 2025).
(37) <https://www.din.de/en> (accessed 29 September 2025).
(38) <https://www.sis.se/en/> (accessed 29 September 2025).
(39) <https://www.iso.org/about-us.html> (accessed 29 September 2025) 16.
(40) <https://www.iso.org/about-us.html> (accessed 29 September 2025).
(41) ISO, 'ISO in Brief', edition 7, 2019 <https://www.iso.org/publication/PUB100007.html> (accessed 29 September 2025) 10.

agencies, academia and industry was responsible for developing ISO 27001 that has become influential within the area of information security. The process that takes place at the ISO is best described by the following:

ISO (the International Organization for Standardisation) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardisation. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work ... The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote. (42)

The process of developing a standard is a voluntary one and relies on a consensus between the experts, the intention being that these standards meet the needs of industry, business and other parties that could have an interest in the development of the standard. The main idea is to bring together as many people as possible to develop the standard, including manufacturers, consultants, distributors and authorities. (43) If any of these actors are interested in participating in the development of a standard, a first step is to contact their national standardisation authority. The basic process for developing a standard can be divided into nine steps: 1) there is an idea, 2) a decision is made to develop a standard, 3) the development process starts, 4) a draft of a proposed standard is produced, 5) a consultation takes place, 6) comments are received, 7) there is a vote, 8) the standard is adopted and 9) the standard is ready for use and evaluation. (44)

Standards, generally speaking, are development after meetings between a wide variety of societal actors. However, it must be highlighted that the majority of these actors are large commercial entities that have a stake in ensuring that a standard will best reflect their commercial needs.

5. The Relationship Between Standards and the Law

It can be said that there is a mutually dependent relationship between standards and the law. Standards refer to laws and laws in turn reference

(42) International Organization for Standardisation (ISO) < https://www.iso.org/obp/ui/#iso:std:iso-iec:27001:ed-2:v1:en> (accessed 29 September 2025)

(43) European Commission (n 17).

(44) Ibid 7.

standards. The attainment of information security in the context of data protection provides a clear example of the interaction between standards and the law. It is stated that standards facilitate the development and application of the law. (45) This is evident from the data protection and security context. For example, Article 32 GDPR states that compliance with the information security demands set by Article 32 can be demonstrated by means of reference to 'an approved certification mechanism as referred to in Article 42'. Article 43 in turn clarifies and elaborates on these certifications mechanisms and adds that the European Commission, '… may adopt implementing acts laying down technical standards for certification mechanisms…'. (46)

The use of standards is voluntary, however, and as seen above, there is a strong connection between standards and the law. For example, standards may be referred to in a law or regulation, they may be referenced in a contract, they may be used in order to achieve clarity when attempting to settle a dispute and courts often refer to then as an evidentiary tool in order to determine whether, in the manufacturing process, the correct technical rules were followed thus showing that the manufacturer displayed 'due diligence'. (47) So in other words, in the area of commercial business, courts can use standards, and the adherence to standards, in order to determine if a product has been developed with due care and is 'fault-free'. In other words, the adherence to a standard can have an effect on which party the evidentiary burden is placed in court proceedings, which is interesting to the extent that the adherence to standards is not mandatory. Consequently, a reason to follow a standard may be strategic in that it may give the manufacturer of a product an evidentiary advantage at an eventual court case. Also, just because a standard is followed, it does not mean that the law has automatically been followed, especially in cases where the law sets demands, that go beyond the scope of a standard. (48) Finally, the adherence to a standard can affect the manner in which damages are allocated. (49)

Another legal principle in European law is called 'the presumption of conformity'. It is often the case that European Union law will require that something is produced or developed according to certain rules but that these rules are specified by referring to an identified standard. The presump-

(45) Swedish Institute for Standards <https://www.sis.se/en/idc/> (accessed 29 September 2025).

(46) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) OJ L 119, 4.5.2016, pp. 1–88 <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng> (accessed 29 September 2025), Article 43(9).

(47) German Institute of Standardisation, 'Introduction to Standardisation' <https://www.din.de/resource/blob/195038/64b75612aae6d6e7341e815becadb5d9/an-introduction-to-standardisation-data.pdf> (accessed 29 September 2025) 32.

(48) Ibid, 33.

(49) Ibid.

tion of conformity principle therefore has the effect that if a manufacturer adheres to a standard specified in the law, then there is a presumption that the product is in conformity with the law and can consequently be placed on the European market. (50) The conformity assessment is an integral part of the NLF and as will be seen below, is central to the Artificial Intelligence Act.

It has been stated that standards share some characteristics with the law in that they are a rule-based system with the goal of enhancing transparency and predictability. (51) Having said this, it is also noteworthy that the process of developing a standard may also exhibit some differences in relation to the traditional law-making process. For example, the standardisation process may not have as many control stations along the way, the process may not be as open to public scrutiny as compared to the traditional law-making process, the is a monetary cost associated with becoming a member of a standardisation institution and standards cost money (they must be bought) as compared with laws that are freely accessible to the public. Finally, standards can be very technical in nature which may be an obstacle to their being comprehended by the general public.

6. The Artificial Intelligence Act

There are a number of soft law instruments promoted under the European Artificial Intelligence Act. For example, in relation to General Purpose AI models, a code of practice is available for providers of these models, which is in turn complemented with Commission guidelines. (52) Harmonised standards are another soft law mechanism available under the Artificial Intelligence Act. On the 24th of October 2024 the European Commission published a brief addressing the issue of harmonised standards in order to facilitate compliance with the Artificial Intelligence Act. (53) In the standardisation request, the European Commission requested that the harmonised standards cover ten concrete aspects of artificial intelligence, namely risk management, data governance and quality, record keeping, transparency, human oversight, accuracy, robustness, cybersecurity, quality management and conformity assessment. (54)

While a form of traditional standardisation, harmonised standards are to be distinguished from regular standards to the extent that their use and

(50) Ibid.

(51) Swedish Institute for Standards (n 41).

(52) European Commission, 'The General-Purpose AI Code of Practice' <https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai> (accessed 30 September 2025).

(53) European Commission, 'Harmonised Standards for the European AI Act' <https://publications.jrc.ec.europa.eu/repository/handle/JRC139430> (accessed 30 October 2025).

(54) Ibid.

the formalia surrounding their use are specifically referred to in European Union legislation. In other words, harmonised standards are a specific mechanism whereby standards become 'officially approved "manuals"' that, if followed by commercial entities or other actors, indicate compliance with safety regulation. (55) That said, their main aim remains the same, namely, to support European Union legislation and transform sometimes vague legal terms into concrete technical equivalents. (56) In relation to harmonised standards, there are a five main actors that play a role in bringing about a standard, namely, the Commission, the European Standards Organisations, the National Standards Bodies, European stakeholder organisations and harmonized standards consultants. (57) While it is the Commission that drafts a standardisation request provided to the standards organizations, the other actors also play a role, national standards bodies representing member states (government, industry, civil society), European stakeholder organisations representing certain interests within the EU (e.g. SME's, trade unions) and harmonised standards consultants, which comprise private entities providing additional expertise. (58)

Harmonised standards have an inherent legal effect to the extent that if they are followed by a party, e.g. by a commercial entity, it does not absolve that party from legal liability but it does create a presumption of conformity. (59) This means that when a party adheres to a standard, it is presumed to have followed the law as represented by that standard, and the product can thus be placed on the market. In the context of the Artificial Act then, high-risk AI systems or general-purpose AI models in conformity with these harmonised standards or parts thereof (published in the Official Journal of the European Union in accordance with Regulation (EU) No 1025/2012) shall be presumed to be in conformity with the requirements set out in Section 2, Chapter III of the AI Act (Art. 40(1), AI Act).

The process for creating harmonised standards is strictly regulated. It begins with the Commission issuing a standardisation request to the EU standardisation organisations (Art.40(2), AI Act). There are three European Standards Organisations, namely CEN (European Committee for Standardisation), CENELEC (European Electrotechnical Committee for Standardisation) and ETSI (European Telecommunications Standards Institute). (60)

---

(55) *EU Artificial Intelligence Act, 'Standard Setting'* <https://artificialintelligenceact.eu/standard-setting/> (accessed 1 May 2025).

(56) *Ibid.*

(57) *Ibid.*

(58) *Ibid.*

(59) DIN, DIHK and ZDH, '*An Introduction to Standardisation: A Practical Guide for Small Business*' (DIN, 2019) <https://www.din.de/resource/blob/195038/64b75612aae6d6e7341e815becadb5d9/an-introduction-to-standardisation-data.pdf> (accessed 1 May 2025).

(60) CEN <https://www.cencenelec.eu/about-cen/> (accessed 1 May 2025); CENELEC <https://www.cencenelec.eu/about-cenelec/> (accessed 1 February 2025); ETSI, <https://www.etsi.org/> (accessed 1 May 2025).

The Commission's standardisation requests can include a wide range of demands and require deliverables such as reporting and documentation on processes to improve AI systems' resource performance, e.g. reducing the high-risk AI system's consumption of energy, and on the energy-efficient development of general-purpose AI models.

In addition, the Commission shall also consult the AI Board and relevant stakeholders, including the advisory forum when preparing a standardisation request and can request the European standardisation organisations to provide evidence of their best efforts to fulfil the objectives. The participants in the standardisation process must also seek to promote investment and innovation in artificial intelligence, including through increasing legal certainty, as well as the competitiveness and growth of the Union market, to contribute to strengthening global cooperation on standardisation and taking into account existing international standards in the field of artificial intelligence that are consistent with Union values, fundamental rights and interests, and to enhance multi-stakeholder governance ensuring a balanced representation of interests and the effective participation of all relevant stakeholders (Art. 40(3), AI Act). Following the request, the national standards board's provide technical experts who form a technical committee, which in turn has relatively free hands with which to create a 'working group', including stakeholders and observers, to draft the actual document. (61) Thereafter, national standards boards collect feedback from stakeholders and consultants review the standard, which is subsequently published in the Official Journal of the EU (OJEU) and eventually adopted by national standards boards into their national standards. (62)

Harmonised standards can be viewed as an interesting regulatory mechanism. They play an important regulatory role to the extent that commercial entities, wishing to comply with legislative documents, can adhere to the standard instead, it being far easier to apply technical specifications to the development of technology as opposed to the sometimes-vague langue of the law. Notwithstanding this function, it cannot be overlooked that the standardisation process lacks the same openness, transparency and public participation as compared to the ordinary legislative process and is therefore inferior compared to the democratic processes associated with the ordinary legislative process. For example, the notion of sending out a public remit is an integral part of the legislative process whereas this public participation mechanism is lacking in relation to the development of standards. In addition, the publicity aspect of standards is diminished by the fact that

(61) *EU Artificial Intelligence Act, 'Standard Setting'* < https://artificialintelligenceact.eu/standard-setting/> (accessed 1 May 2025).

(62) *Ibid.*

they are only publicly available for a substantial fee. (63) Finally, it is noteworthy that at the time of the writing of this article, the harmonised standards for the Artificial Intelligence Act had still not been finalised.

7. Conclusion

This article set out with the main aim of illumination the extent to which soft law instruments are increasingly being relied upon to complement the literal legislative instrument texts. This trend has arisen out of necessity in a time when digitalisation within society is occurring at a rapid rate and the transformation of society to the use of complex tools encompassing emerging technologies is putting pressure on the legislative process as a regulatory instance. The pace at which technologies are developing and the demand to maintain technological neutrality in order to remain relevant over time has increased the utility of soft law instruments. One such soft law mechanism investigated above in the article is that of the standard. The advent of the European Artificial Intelligence Act has elevated the status of the standard as a soft law mechanism. The complexity of the technology as well as the complexity of the law has highlighted the need for harmonised standards as a mechanism with which to attain compliance with the Artificial Intelligence Act. However, there are drawbacks with using standards as a form of regulation: they are merely voluntary, their development in essence occurs behind closed doors and to a limited extent the legislator is relinquishing part of its legislative powers to the commercial entities that the standards will in turn regulate. Nevertheless, it would seem that, taken the lay of the technological landscape, the European Union legislator has few options open and it is argued that standards, and more specifically harmonised standards, will continue to play an important function within the European Union's regulatory toolbox.

---

(63) It is noteworthy that in a recent ruling by the Court of Justice of the European Union, the court found that 'overriding public interest' justifird the free disclosure of technical standards, Fortuna, Gerado, 'Court Ruling Casts Doubt on Bloc's Standards Bodies' Funding' <https://www.euronews.com/next/2024/03/07/court-ruling-casts-doubt-on-blocs-standards-bodies-funding> (accessed 30 September 2025).

La protección de los datos personales en el proceso penal chileno ante la utilización de inteligencia artificial

Por Valeska Fuentealba Sepúlveda (*)

Resumen: El presente trabajo tiene por objeto estudiar la configuración del derecho a la autodeterminación informativa o de protección de datos personales en un proceso penal chileno en el que se utiliza inteligencia artificial y valorar críticamente la regulación existente. Utilizando una metodología dogmática, el texto analiza la situación del derecho a la autodeterminación informativa previo al proceso penal y durante su desarrollo, de modo de delimitarlo y compatibilizarlo con los sistemas de inteligencia artificial que pueden auxiliar al *ius puniendi* estatal. Lo anterior resulta relevante considerando que, en la actualidad, no hay normas que regulen el uso de esta herramienta en un proceso jurisdiccional, a pesar de que es usada por los intervinientes. Por lo anterior, será relevante analizar cómo se configura el derecho fundamental mencionado, así como las leyes que regulan el tratamiento de datos personales.

Palabras clave: Derecho a la protección de datos personales; datos personales; sistema penal; tratamiento de datos personales.

Abstract: The purpose of this paper is to examine the configuration of the right to informational self-determination, or the protection of personal data, within the context of a Chilean criminal proceeding in which artificial intelligence is employed, and to critically assess the existing regulatory framework. Using a dogmatic methodology, the text analyzes the status of the right to informational self-determination both prior to and during the criminal process, with the aim of defining its scope and reconciling it with artificial intelligence systems that may assist the State's *ius puniendi*. This analysis is particularly relevant given that, at present, there are no legal provisions expressly regulating the use of such tools in judicial proceedings, despite their use by participants in the process. Accordingly, it becomes

(*) Abogada, licenciada en Ciencias Jurídicas, magíster en Derecho Penal y Ciencias Penales y candidata a doctora en Derecho, todos por la Pontificia Universidad Católica de Valparaíso. Profesora investigadora de la Escuela de Derecho, Facultad de Ciencias Humanas, de la Universidad Bernardo O'Higgins, Chile. Correo electrónico: v.fuentealba.sepulveda@gmail.com

essential to examine how this fundamental right is shaped, as well as the legislation governing the processing of personal data.

Keywords: Right to personal data protection; data privacy; criminal justice system; processing of personal data.

1. INTRODUCCIÓN

El ordenamiento jurídico se erige sobre valores e ideas fundamentales que resultan relevantes para la convivencia social y el desarrollo de las personas. Estos valores, así como la persona y la misma sociedad, no son estáticos, sino que su valoración y determinación están históricamente condicionados (Muñoz y García, 2015, p. 64).

En la medida que la sociedad se complejiza, el ordenamiento jurídico irá dando respuesta a sus problemas más sensibles. Hay diversos ejemplos de esto, pero, sin dudas, uno que lo demuestra de manera palmaria es el desarrollo y uso de las tecnologías, adentrándose en ámbitos tan sensibles como el ejercicio de la función jurisdiccional (Allende, 2019, p. 187-189, 198). En Chile, aunque existen antecedentes normativos previos (Fuentealba y Quiero, 2023, pp. 227-228), el uso intensivo de tecnologías se acelera tras la reforma procesal penal del año 2000 (Romero, 2021, p. 404).

En este contexto, la protección de datos personales emerge como una dimensión del derecho a la privacidad que exige especial atención, en tanto la recopilación y tratamiento de información de los intervinientes puede impactar en sus derechos fundamentales y en el equilibrio entre eficacia persecutoria y garantías procesales (Ramírez, 2016, p. 59). Tal como advierte Guerrero (2020, p. 34), una deficiente regulación de protección de datos personales conseguiría afectar el ejercicio mismo de la función jurisdiccional.

En ese sentido, la hipótesis de este trabajo es que la utilización de nuevas tecnologías y, particularmente, de inteligencia artificial en el proceso penal genera un escenario dinámico en el que la afectación de derechos fundamentales –especialmente el de protección de datos personales– resulta difícil de prever y prevenir, cediendo con frecuencia ante los imperativos de eficiencia de la persecución penal.

2. LA PROTECCIÓN DE DATOS PERSONALES COMO DERECHO A RESPETAR EN EL PROCESO PENAL CHILENO

Toda persona posee la prerrogativa de decidir quién accede a su vida privada, comunicaciones, documentos y datos personales, configurando un ámbito libre de injerencias (Figueroa, 2020, pp. 129-131), sin embargo, no existe en Chile una regulación armónica y sistemática del presunto dere-

cho a la privacidad [1], donde se hallaría el derecho a la autodeterminación informativa. Este déficit normativo dificulta determinar su contenido y los límites que el legislador le reconoce.

El derecho a la privacidad (Donoso y Reusser, 2022, pp. 4-5), originado en la idea de "ser dejado solo" de Warren y Brandeis (1980, pp. 193-219), hoy se entiende como un poder de control sobre la información personal, ámbito en el que no participan otras personas y que protege la dignidad y libertad humana (Lara, Pincheira y Vera, 2014, p. 12), lo que significaría que es el "derecho del individuo a decidir por sí mismo en qué medida quiere compartir con otros sus pensamientos y sentimientos, así como los hechos de su vida personal" (Cerda, 2012, p. 7). La evolución de este derecho –que cambia según el contexto y época histórica en el que se analiza (Merkel, 2022, pp. 254-255)– ha incorporado el concepto de "autodeterminación informativa", entendida como la facultad de controlar los datos personales y decidir su tratamiento (Sancho, 2021, p. 771) [2]. Dicha facultad es esencial para garantizar la libertad futura del individuo (Contreras, 2020, p. 100; Reusser, 2021, p. 60).

No obstante, existe dispersión normativa al momento de regular el tema, encontrándose disposiciones a nivel constitucional, legal [3] e, incluso, infralegal [4]. A esto se suma la influencia de estándares internacionales como el Reglamento Europeo de Protección de Datos [5] o las Directrices de

---

(1) "Presunto", en cuanto no es un derecho aceptado por la totalidad de la doctrina. La dificultad no es, por cierto, solamente dogmática, sino que también normativa: nuestra Carta Fundamental no alude de manera expresa a esta garantía. Ahora, ello no es óbice para considerarla consagrada entre nosotros, piénsese, por ejemplo, en el derecho al debido proceso, nomenclatura no utilizada por el constituyente.

(2) En un sentido similar, Donoso y Reusser (2022, pp. 10-11) entienden que sería "el derecho del individuo a controlar la obtención, tenencia, tratamiento y transmisión de datos relativos a su persona, decidiendo en cuanto a los mismos las condiciones en que dichas operaciones pueden llevarse a cabo".

(3) Por nombrar solo algunas, a nivel legal encontramos la ley ° 19.628, que Protege la Vida Privada, e modificada por la ley 21.719, que se encuentra en periodo de vacancia; la ley 20.285, Sobre Acceso a la Información Pública; ley 20.575, Sobre Principio de Finalidad en el Tratamiento de Datos Personales; ley 20.584, Sobre Derechos y Deberes de los Pacientes; ley 21.096, que Facilita el Intercambio Recíproco de Información con otros Países; y, más recientemente –aunque la regulación del derecho es tangencial– la ley 21.430, Sobre Garantías y Protección Integral de los Derechos de la Niñez y Adolescencia; y la ley 21.459, que Establece Normas sobre Delitos Informáticos.

(4) A nivel infralegal son relevantes el Acta 44-2022 de la Excma. Corte Suprema, Auto Acordado Sobre Criterios de Publicidad de Sentencias y Carpetas Electrónicas; y la Resolución Exenta N° 33 del SERNAC, que Aprueba Circular Interpretativa Sobre Protección de los Consumidores Frente al Uso de Sistemas de Inteligencia Artificial en las Relaciones de Consumo.

(5) Con respecto a esto, Donoso y Reusser (2022, p. 40) sostienen que "a pesar de que parece aplicarse solo a Europa, en realidad tiene un alcance global que lo lleva a tener impactos concretos incluso en los países más alejados de ese continente, como el nuestro, tanto por la influencia de la normativa y el desarrollo de la protección de datos como derecho fundamental en todo el orbe, como también por normas que afectan las relaciones contractuales y comerciales de quienes se relacionan, directa o indirectamente, con productos y servicios que fluyen hacia el Viejo Continente o que provienen del mismo; es decir, el Reglamento introdujo reglas de aplicación extraterritorial que afectan a entidades fuera de la Unión Europea, las que incluso pueden quedar sujetas a su régimen sancio-

la OCDE sobre Protección de la Privacidad, los que ofrecen principios de recolección y uso legítimo (OCDE, 2003, p. 39), e inspiran normas internas como la ley 19.628 (Velásquez y Donoso, 2013, pp. 174-181).

En este contexto se ha consolidado una "cultura de protección de datos" (Zeballos, 2013, p. 52), que reconoce el carácter esencial de la información personal de un sujeto. Sin embargo, el legislador puede disponer excepciones por razones de interés público como ocurre en el ámbito penal (Bronfman y Fuentealba, 2023, pp. 222-223).

La afectación de la privacidad debe evaluarse a partir de su potencial incidencia en la libertad y dignidad de la persona (Donoso y Reusser, 2022, p. 10). Así, el dato personal, en tanto es un medio para conocer la vida de la persona, exige garantías proporcionales al riesgo que supone su uso en procesos donde el Estado ejerce coerción penal asistido por tecnología y, particularmente, por inteligencia artificial.

3. La inteligencia artificial en el proceso penal chileno

La relevancia del derecho a la protección de datos personales se acrecienta cuando los sistemas de inteligencia artificial procesan información de los intervinientes en el proceso penal –por ejemplo, los datos de atención de urgencias (DAU) de imputado y víctima, o la información sobre una sentencia condenatoria firme que afecta al imputado, los que son considerados datos sensibles (Ruiz, 2022, pp. 22-23)–, dada su capacidad para tratar grandes volúmenes de datos y generar decisiones automatizadas (Dunstan, Maas y Tobar, 2022, p. 32).

Si bien en Chile no existe una regulación legal específica, la Política Nacional de Inteligencia Artificial (Ministerio de Ciencia, 2024, pp. 9, 70) la define como un "sistema computacional que puede, para un determinado conjunto de objetivos definidos por humanos, hacer predicciones y recomendaciones o tomar decisiones que influyen en entornos reales o virtuales". Para ello emplea como insumo fundamental los datos, los que son, según el mismo documento, "una observación, representación o sampleo de fenómenos reales, que funcionan como materia prima para los algoritmos, y que permite entrenar y mejorar algoritmos y modelos apuntando a explicar o predecir los fenómenos representados o sampleados u observados". Ellos son, entonces, la materia prima de los algoritmos, entendidos como un

---

nador, que no se caracteriza precisamente por su laxitud". De todos modos, esto se ve reforzado en el hecho de que Chile suscribió el Acuerdo por el que se Establece una Asociación entre la Comunidad Europea y sus Estados Miembros (Decreto N° 28 de 2003, del Ministerio de Relaciones Exteriores), donde se acuerda dar al procesamiento de datos personales un elevado nivel de protección, el que se ha entendido que se ve satisfecho con el Reglamento ya mencionado.

"conjunto ordenado y finito de operaciones que permite hallar la solución de un problema" (6).

De acuerdo con Corvalán (2020, p. 33), la inteligencia artificial replica procesos cognitivos humanos a mayor escala y velocidad, lo que explica su creciente uso por razones de eficiencia y aparente objetividad (Azuaje y Finol, 2020, p. 114). Sin embargo, la distinción entre datos personales y no personales –relevante para el derecho– pierde sentido en el ámbito tecnológico, pues ambos pueden alimentar modelos predictivos capaces de construir perfiles individuales (Bronfman y Fuentealba, 2023, p. 225). Este perfilado puede comprometer la vida privada y la autodeterminación informativa, incluso a partir de datos no protegidos formalmente.

Los riesgos asociados a la inteligencia artificial incluyen la existencia de sesgos invisibles que reproducen discriminaciones existentes en cuanto al género, etnia, condición socioeconómica u otras (Palma, 2021, pp. 289-291; Boulamnwini y Gebru, 2018, pp. 1-15; Sobrino, 2020, pp. 2-3). Asimismo, la opacidad algorítmica dificulta la rendición de cuentas del Estado, generando tensiones entre eficiencia pública y transparencia judicial (Araya, 2020, pp. 260-262; Azuaje y Finol, 2020, p. 147). En este sentido, la "justicia predictiva" puede derivar en falsos positivos o decisiones basadas en correlaciones espurias, afectando derechos fundamentales de los intervinientes, sobre todo del imputado.

En suma, la inteligencia artificial ofrece eficiencia en la persecución penal, pero su despliegue puede erosionar garantías esenciales de las personas, como son la dignidad y libertad, máxime ante la ausencia de normas que delimiten la responsabilidad por el mal uso de datos personales por órganos como el Ministerio Público o las policías (Fuentealba y Quiero, 2025, p. 165-173).

3.1. La utilización de inteligencia artificial antes de iniciado el proceso

En etapas previas al proceso penal, la inteligencia artificial puede emplearse con fines preventivos o de vigilancia, lo que plantea serios riesgos de afectación de la privacidad y la libertad individual (Cavoukian, 2013, pp. 31 y 37; Blanco, 2020, pp. 283, 285 y 289).

La jurisprudencia constitucional chilena ha sostenido que la ley puede limitar este derecho por razones legitimas (7), aunque ha rechazado faculta-

(6) En otro sentido, Nieva (2018, p. 21) plantea que sería un "esquema ejecutivo de la máquina almacenando todas las opciones de decisión en función de los datos que se vayan conociendo".

(7) Sentencias del Tribunal Constitucional Chileno roles N° 1683-10; N° 1732-10; N° 1990-11; y N° 460-05.

des estatales ilimitadas en materia de investigación [8]. Un eventual uso de sistemas predictivos de criminalidad atentaría contra el principio de culpabilidad y de responsabilidad por el hecho (Fuentealba y Falcone, 2022, pp. 86, 88), derivando en un "Derecho penal de la seguridad" (Barona, 2019, p. 43).

La utilización de tecnologías de vigilancia –drones, cámaras, sistemas de geolocalización y posicionamiento global o de reconocimiento facial– evidencia una tensión entre seguridad pública y protección de datos personales (Malamud, 2018, 139; Ramírez, 2016, p. 59; Fuentealba y Quiero, 2023, pp. 231-234) que ya ha tenido lugar en otras latitudes (Hairabedián, 2020, 184). En el derecho chileno, la videovigilancia estatal antes del proceso ha sido permitida bajo criterios de proporcionalidad (Donoso y Reusser, 2022, pp. 172-177), aunque puede implicar la captación masiva de información de personas no imputadas. Tales prácticas, equiparables a diligencias intrusivas, deberían requerir de autorización judicial conforme al artículo 9° del Código Procesal Penal chileno. En un sentido similar se pronuncia, por ejemplo, Hairabedián (2020, p. 185)

3.2. La utilización de inteligencia artificial en el proceso penal

Durante el desarrollo de un proceso penal puede utilizarse inteligencia artificial en diversas etapas y por parte de diferentes intervinientes o sujetos procesales. Si bien el tema puede ser, por tanto, abordado desde diversas aristas, para efectos del trabajo solo se mencionará la fase de investigación.

El Código Procesal Penal chileno regula actuaciones que inciden en la privacidad –a modo ejemplar pueden verse los artículos 197 y 198 (exámenes corporales o médicos), 199 bis (pruebas de ADN), 203 (pruebas caligráficas), 213 (medidas de vigilancia) y 226 (otros medios técnicos de investigación, como fotografía, filmación u otros medios de reproducción de imágenes)– y que, mediante herramientas tecnológicas, adquieren una nueva dimensión (Blanco, 2020, p. 725). Un ejemplo paradigmático de ello es el "agente encubierto en línea", introducido por el artículo 12 de la ley 21.459, sobre delitos informáticos, que autoriza la infiltración digital de funcionarios policiales en redes cerradas de comunicación (Mayer y Vera, 2022, p. 42).

Esta diligencia se asemeja al allanamiento remoto (Salt, 2020, p. 152), permitiendo acceder subrepticiamente a datos informáticos, personales o no [9], con el fin de recabar evidencia. Aunque respaldada por la ley 19.974,

---

(8) Sentencias del Tribunal Constitucional Chileno roles N° 389-03; N° 521-06; N° 1894-11; y N° 1939-11.

(9) El dato informático puede, en ocasiones, ser un dato personal, idea que tangencialmente (en cuanto se aludía al derecho a la privacidad) fue discutida durante la tramitación de la ley. Ver Bcn, *Historia de la Ley N° 21.459*, pp. 21, 23, 27, 28, 46, entre varias otras.

sobre Sistema de Inteligencia del Estado, su ejecución debe observar el principio de proporcionalidad, tanto para autorizar como para valorar la evidencia obtenida. Así, la advertencia del juez Brandeis en el caso "Olmstead v. United States" (en el año 1928) sigue vigente: el progreso tecnológico facilita al Estado penetrar en los ámbitos más íntimos del individuo (Blanco, 2020, p. 330).

El uso de inteligencia artificial en la investigación y tratamiento de datos exige sistemas transparentes, imparciales y auditables (Lemontech, sf., pp. 21-23) [10] [11]. En el extranjero existen experiencias como Prometea (Argentina), Fiscal Watson (Colombia) o COMPAS (Estados Unidos), que ilustran la incorporación de herramientas de inteligencia artificial en funciones judiciales y fiscales (Le Fevri, 2022, pp. 36-42; Roa y Sanabria-Moyano, 2022, pp. 275-310), pero también alertan sobre los riesgos de delegar decisiones sensibles en sistemas automatizados.

En Chile, a través del Ecosistema Heredia, el Ministerio Público también ha comenzado a utilizar esta clase de tecnología, estando aún pendiente en análisis por parte de la doctrina de esta herramienta, sobre todo en lo tocante a la base de licitud que la autorizaría [12], a pesar de que hay quienes sostienen que, con respecto al tratamiento de datos, "la frontera de lo aceptable se ha movido considerablemente" (Abuselme, 2025, p. 16).

4. Conclusiones

El uso de las tecnologías ha irrumpido en diversas actividades, incluyendo la función jurisdiccional, configurando nuevas tensiones entre la seguridad pública y los derechos fundamentales. En Chile, la ausencia de un marco legal sobre inteligencia artificial deja desprotegido el derecho a la autodeterminación informativa frente al uso estatal de tecnologías predictivas y de vigilancia.

---

(10) Lemontech, *Justicia digitalizada y descentralizada: ¿en qué consiste y cómo adaptarse?*, pp. 21-23 [visible en internet: https://recursos.lemontech.com/landing/ebook-justicia-digitalizada-y-descentralizada].

(11) De todos modos, a pesar de que ya existen casos de utilización de inteligencia artificial como apoyo para la dictación de sentencias en otras áreas del derecho, no se han documentado casos en los que sea el software, sin intervención humana, la que resuelva el caso. Se pueden ver apoyos en la dictación de sentencias en las realidades de Estonia (2019) y Beijing (con su Tribunal de Internet, inaugurado el 2019) (Cárdenas, 2021, pp. 3-4).

(12) Es discutible que la ley haya considerado el tratamiento dentro de las funciones que le encomienda al Ministerio Público. Por un lado, porque, de conformidad al principio de legalidad, si bien no se requiere una mención expresa, sí sería necesario que la hipótesis de tratamiento de datos pueda comprenderse connatural a las funciones que la Constitución o la ley encomienda al Ministerio Público. Por otro, porque tratándose de las policías, auxiliares del Ministerio Público, existen normas que, de manera más clara, permitirían fundamentar el uso de la inteligencia artificial (al menos en lo tocante al tratamiento de datos biométricos para el control de las fronteras, como puede verse en la reciente ley 21.734, que vino a modificar el DL N° 2.460 de la Policía de Investigaciones de Chile que, en todo caso, ya amparaba dicha hipótesis).

La hipótesis de este trabajo –que la utilización de nuevas tecnologías y, particularmente, de inteligencia artificial en el proceso penal genera un escenario dinámico en el que la afectación de derechos fundamentales es difícil de identificar y de precaver, cediendo su protección ante la eficacia de la persecución penal– se confirma ante la falta de mecanismos de control y responsabilidad institucional. Si bien la inteligencia artificial promete mayor eficiencia en la persecución penal, su empleo sin garantías adecuadas puede erosionar la dignidad humana, la libertad y el control individual sobre la información personal.

En definitiva, la protección de datos personales en el proceso penal chileno demanda una regulación que armonice la eficacia esperada del *ius puniendi* con el respeto a la autodeterminación informativa, incorporando principios de transparencia, proporcionalidad y rendición de cuentas en el uso de sistemas de inteligencia artificial.

Bibliografía

Abusleme, C. (2025). El diablo está en los detalles: cinco nudos legales de los que depende el éxito de la nueva ley de protección de datos personales en el sector público. En D. Zaror (Ed.), *Expertos en acción: protegiendo datos personales bajo la nueva legislación*. Rubicón Editores.

Allende, J. A. (2019). Tribunales civiles en línea: Una propuesta para introducirlos sin afectar el derecho a acceder a la justicia de quienes no están conectados a internet. *Revista Chilena de Derecho y Tecnología, 8*(1).

Araya, C. (2020). Desafíos legales de la inteligencia artificial en Chile. *Revista Chilena de Derecho y Tecnología, 9*(2).

Arrabal, P. (2021). La ilicitud de la prueba tecnológica por la vulneración del derecho fundamental a la protección de datos en el ámbito penal. En S. Pereira & F. Ordóñez (Dirs.), *Investigación y proceso penal en el siglo XXI. Nuevas tecnologías y protección de datos*. Thomson Reuters.

Azuaje, M., & Finol, D. (2020). Transparencia algorítmica y la propiedad intelectual e industrial: Tensiones y soluciones. *Revista La Propiedad Inmaterial, 30*.

Barona, S. (2019). Inteligencia artificial o la algoritmización de la vida y de la justicia: ¿Una solución o problema? *Revista Boliviana de Derecho, 28*.

Biblioteca del Congreso Nacional de Chile. (s. f.). Historia de la Ley N.º 21.459. https://www.bcn.cl/historiadelaley/nc/historia-de-la-ley/8018/

Blanco, H. (2020). *Tecnología informática e investigación criminal*. Thomson Reuters.

Boulamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research, 81*. https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

Bronfman, A. & Fuentealba, V. (2023). La protección de datos personales en una etapa preliminar al proceso penal frente a la utilización de inteligencia artificial. En L. Mayer, G. Oliver & J. Vera (Ed.), *Hacia un derecho penal centrado en la persona. Libro homenaje al profesor Luis Rodríguez Collao*. Editorial Jurídica de Chile.

Cárdenas, R. (2021). ¿Jueces robots? Inteligencia artificial y derecho. *Revista Justicia & Derecho, 4*(2).

Cavoukian, A. (2013). Surveillance, then and now: Securing privacy in public spaces. Information and Privacy Commissioner of Ontario. https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-surveillance.pdf

Cerda, A. (2012). *Legislación sobre protección de las personas frente al tratamiento de datos personales*. Material de estudio del Centro de Estudios en Derecho Informático. Facultad de Derecho, Universidad de Chile. https://www.researchgate.net/publication/342183451_Legislacion_sobre_Proteccion_de_las_Personas_frente_al_Tratamiento_de_Datos_Personales

Contreras, P. (2020). El derecho a la protección de datos personales y el reconocimiento de la autodeterminación informativa en la Constitución chilena. *Estudios Constitucionales, 18*(2).

Corvalán, J. G. (2017). Inteligencia artificial y derechos humanos (parte I). *Diario DPI*. https://dpicuantico.com/sitio/wp-content/uploads/2017/07/Juan-Gustavo-Corvalan-Constitucional-03.07.2017.pdf

Corvalán, J. G. (2020). *Perfiles digitales humanos*. Thomson Reuters.

Donoso, L., & Reusser, C. (2022). *La protección de los datos personales en Chile*. DER Ediciones.

Dunstan, J., Maas, A., & Tobar, F. (2022). *Una mirada a la era de los datos*. Editorial Universitaria.

Figueroa, R. (2020). Derecho a la privacidad. En P. Contreras & C. Salgado (Eds.), *Curso de Derechos Fundamentales*. Tirant lo Blanch.

Fuentealba, V., & Falcone, D. (2022). Legitimidad del derecho penal: Principios y límites. En V. Fuentealba (Coord.), *Esquemas de derecho penal chileno. Parte general* (Vol. I). Tirant lo Blanch.

Fuentealba, V., & Quiero, F. (2025). Responsabilidad civil extracontractual derivada del tratamiento indebido de datos personales. En G. Garbaccio & N. Souza, *Construyendo puentes: colaboración jurídica entre Chile y Brasil*. Instituto Brasileiro de Ensino. https://repositorio.idp.edu.br/handle/123456789/5374

Fuentealba, V., & Quiero, F. (2023). ¿Es importante hablar de protección de datos personales en el proceso penal chileno? En C. Droguett & N. Walker (Eds.), *Derecho digital y privacidad en América y Europa*. Tirant lo Blanch.

Guerrero, B. (2020). Protección de datos personales en el Poder Judicial: Una mirada al principio de publicidad de las actuaciones judiciales. *Revista Chilena de Derecho y Tecnología, 9*(2).

Hairabedián, M. (2020). El uso de drones sobre domicilios en la investigación penal. En D. Dupay (Dir.), *Cibercrimen II*. B de F.

Irti, N. (1991). *La edad de la descodificación* (L. Rojo, Trad.). José María Bosch.

Lara, J. C., Pincheira, C., & Vera, F. (2014). *La privacidad en el sistema legal chileno*. ONG Derechos Digitales. https://www.derechosdigitales.org/wp-content/uploads/pp-08.pdf

Le Fevri, E. (2022). Informe 6: Uso estratégico de datos e inteligencia artificial en la justicia. *CAF*. https://scioteca.caf.com/handle/123456789/1932

Lemontech. (s. f.). Justicia digitalizada y descentralizada: ¿En qué consiste y cómo adaptarse? https://recursos.lemontech.com/landing/ebook-justicia-digitalizada-y-descentralizada

Malamud, S. (2018). Videovigilancia y privacidad. Consideraciones en torno a los casos "Globos" y "Drones". *Revista Chilena de Derecho y Tecnología, 7*(2).

Mayer, L., & Vera, J. (2022). La nueva ley de delitos informáticos. *Revista de Ciencias Penales*.

Merkel, L. (2022). *Derechos humanos e investigaciones policiales. Una tensión constante*. Marcial Pons.

Ministerio de Ciencia, Tecnología, Conocimiento e Innovación. (2021). *Política nacional de inteligencia artificial*. https://minciencia.gob.cl/areas-de-trabajo/inteligencia-artificial/politica-nacional-de-inteligencia-artificial/

Muñoz, F., & García, M. (2015). *Derecho penal. Parte general* (9.ª ed.). Tirant lo Blanch.

Nieva, J. (2018). *Inteligencia artificial y proceso judicial*. Marcial Pons.

OCDE. (2003). *Privacy online: OECD guidance on policy and practice*. OECD Publications. https://read.oecd-ilibrary.org/science-and-technology/privacy-online_9789264101630-en#page39

Palma, J. M. (2021). Inteligencia artificial y lucha contra la delincuencia: Potencialidad y peligros en el mundo global. En F. Llano & J. Garrido (Eds.), *Inteligencia artificial y derecho. El jurista ante los retos de la era digital*. Thomson Reuters.

Ramírez, T. (2016). Nuevas tecnologías al servicio de la seguridad pública y su impacto en la privacidad: Criterios de ponderación. *Revista Chilena de Derecho y Tecnología, 5*(1).

Reusser, C. (2021). *Derecho al olvido. La protección de datos personales como límite a las libertades informativas*. DER Ediciones.

Roa, M., & Sanabria-Moyano, J. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. *Revista Brasileira de Direito Processual Penal, 8*(1).

Romero, S. (2021). Juicios penales mediante videoconferencia y su repercusión en los derechos del imputado: La experiencia tras la pandemia en Chile. En S. Pereira & F. Ordóñez (Dirs.), *Investigación y proceso penal en el siglo XXI. Nuevas tecnologías y protección de datos*. Thomson Reuters.

Ruiz, A. (2022). De la transparencia a la protección de los datos de personas condenadas. Comentario a la sentencia Rol N.º 26276-2019 de la Excelentísima Corte Suprema. En P. Contreras et al., *Privacidad y protección de datos personales*. DER Ediciones.

Salt, M. (2020). Allanamiento remoto: ¿Un cambio de paradigma en el registro y secuestro de datos informáticos? En D. Dupay (Dir.), *Cibercrimen II*. B de F.

Sancho, M. (2021). Tratamiento de datos genéticos y vulneraciones de derechos fundamentales. En S. Pereira & F. Ordóñez (Dirs.), *Investigación y proceso penal en el siglo XXI. Nuevas tecnologías y protección de datos*. Thomson Reuters.

Sobrino, W. (2020). *Contratos, neurociencias e inteligencia artificial*. Thomson Reuters.

Warren, S., & Brandeis, L. (1890). The right to privacy. *Harvard Law Review*. https://www.cs.cornell.edu/~shmat/courses/cs5436/warren-brandeis.pdf

Zaballos, E. (2013). *La protección de datos personales en España: Evolución normativa y criterios de aplicación* [Tesis doctoral, Universidad Complutense de Madrid]. https://eprints.ucm.es/id/eprint/22849/1/T34733.pdf

Câmeras corporais nas polícias brasileiras: entre transparência e desigualdade federativa

Por Augusto Jobim do Amaral [(*)], Eduardo Baldissera Carvalho Salles [(**)] y Paula Fernanda Failace Antunes de Oliveira [(***)]

Resumo: Este artigo apresenta os primeiros resultados de uma investigação empírica conduzida entre os meses de julho e dezembro de 2024, realizada por meio de pedidos encaminhados com base na Lei de Acesso à Informação. O objetivo central foi compreender o uso institucional das câmeras corporais pelas polícias dos estados e também pela União, analisando de que maneira esses equipamentos têm sido incorporados às rotinas das diferentes corporações policiais. Para isso, foram elaborados e enviados questionários padronizados direcionados às Polícias Militares, às Polícias Civis, às Polícias Penais, bem como à Polícia Federal e à Polícia Rodoviária Federal. As respostas recebidas —quando efetivamente fornecidas— revelam um cenário marcado por improviso administrativo, significativa opacidade e, sobretudo, profunda desigualdade federativa. Cada ente policial opera a partir de regras próprias, sem a existência de diretrizes mínimas compartilhadas nem de mecanismos estruturados de monitoramento, controle e avaliação de resultados. Paralelamente, observa-se que, ao mesmo tempo em que são investidos milhões de reais na aquisição ou na locação dos equipamentos, permanece a ausência de métricas públicas claras e de fiscalização organizada. Tal lacuna ameaça transformar uma ferramenta

---

(*) Doutor em Altos Estudos Contemporâneos pela Universidade de Coimbra (2012); Doutor em Ciências Criminais pela Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS (2014). Professor Permanente dos Programas de Pós-Graduação em Ciências Criminais e em Filosofia da PUCRS. Líder do Grupo de Pesquisa "Politicrim" (PUCRS). Currículo Lattes: http://lattes.cnpq.br/4048832153516187. ORCID: https://orcid.org/0000-0003-0874-0583.

(**) Doutor em Ciências Criminais pela PUCRS (2022), em cotutela com a Universidad de Sevilla (Espanha), com bolsa CAPES-PrInt. Pesquisador de Pós-Doutorado junto à Escola de Humanidades da PUCRS. Pesquisador do Grupo de Pesquisa Politicrim (PUCRS). Currículo Lattes: http://lattes.cnpq.br/9552058719966354. ORCID: https://orcid.org/0000-0001-6342-5733.

(***) Graduanda em Direito pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Integrante do Grupo de Pesquisa Politicrim (PUCRS) e do programa G+1 em Ciências Criminais (PUCRS). Bolsista de Iniciação Científica — CNPq (2021 2022), FAPERGS (2023–2025) e, atualmente, CNPq (2025–2026) Currículo Lattes: https://lattes.cnpq.br/5752482309065903. ORCID: https://orcid.org/0009-0006-0706-9221.

originalmente concebida como promissora para ampliar a transparência e o controle da atividade policial em mera vitrine tecnológica. Assim, os achados indicam não apenas uma política em processo de construção, mas também uma política em constante disputa, cujos rumos e alcances permanecem indefinidos.

Palavras-chave: câmeras corporais; polícias brasileiras; segurança pública; controle; políticas públicas; letalidade policial.

Cámaras corporales en las policías brasileñas: entre transparencia y desigualdad federativa

Resumen: Este artículo presenta los primeros resultados de una investigación empírica realizada entre los meses de julio y diciembre de 2024, por medio de solicitudes formuladas a través de la Ley de Acceso a la Información. El estudio se centra en el uso institucional de las cámaras corporales por parte de las policías de los estados y también de la Unión. Para ello, fueron enviados cuestionarios estandarizados a diferentes corporaciones: policías militares, policías civiles, policías penales, además de la Policía Federal y de Carreteras Federal. Las respuestas obtenidas —cuando efectivamente fueron proporcionadas— ponen en evidencia un escenario marcado por improvisación, escasa transparencia y una profunda desigualdad federativa. Cada ente policial actúa de manera aislada, estableciendo sus propias reglas, sin la existencia de directrices mínimas comunes ni de mecanismos sólidos y efectivos de control, monitoreo y evaluación de los resultados. Al mismo tiempo en que se invierten millones de reales en la adquisición o alquiler de los equipos, la ausencia de métricas públicas claras y de una fiscalización estructurada genera el riesgo de que una herramienta inicialmente prometedora de control y rendición de cuentas termine reducida a una mera vitrina tecnológica. En síntesis, los hallazgos revelan no solo una política pública aún en construcción, sino también una política en abierta disputa, cuyos rumbos y alcances permanecen indefinidos.

Palabras clave: cámaras corporales; policías brasileñas; seguridad pública; control; políticas públicas; letalidad policial.

Body-worn cameras in brazilian police forces: between transparency and federative inequality

Abstract: This article presents the initial results of an empirical investigation conducted between July and December 2024, carried out through requests submitted under the Access to Information Law. The main objective was to examine the institutional use of body-worn cameras by state police forces as well as by federal law enforcement agencies, analyzing the ways in which these devices have been integrated into the daily routines and operational procedures of different police organizations across Brazil. To

achieve this, standardized questionnaires were carefully designed and sent to the Military Police, Civil Police, Penitentiary Police, the Federal Police, and the Federal Highway Police. The responses obtained—when actually provided—reveal a landscape characterized by administrative improvisation, notable opacity, and, above all, significant federal inequality. Each law enforcement body operates according to its own internal rules, in the absence of minimum shared guidelines or structured mechanisms for monitoring, evaluation, and accountability. At the same time, while substantial public funds are invested in the acquisition or leasing of these devices, there is still a lack of transparent metrics and organized oversight mechanisms. This structural gap risks transforming a tool originally conceived to promote transparency, accountability, and oversight into a mere technological showcase. Consequently, the findings indicate not only a policy under construction but also a field of continuous contestation, whose scope, effectiveness, and future direction remain uncertain.

Keywords: body-worn cameras; Brazilian police; public security; oversight; public policy; police lethality.

1. Introdução

A apresentação deste relatório de pesquisa trata de uma etapa inicial do levantamento quantitativo relativo ao estado da arte acerca da implementação dos dispositivos de câmeras corporais nas polícias brasileiras. Para este momento, optou-se deliberadamente por não se antecipar um enfrentamento crítico dos dados produzidos, em especial a partir análise dos próprios pressupostos que amparam a produção das formas tecnopolíticas de controle social que vem sendo produzidas, na direção de um aparente solucionismo tecnológico, não raro, afeitas à relegitimação/moderação do exercício do poder de polícia, em sua intrínseca gestão desigual de violência na população [(1)].

Assim, desde esta amostra inicial, de modo rigoroso, entre julho e dezembro de 2024, foi realizada uma investigação empírica voltada a compreender como o uso de câmeras corporais tem sido incorporado às rotinas das polícias no Brasil. O objetivo central foi mapear as práticas institucionais relacionadas a esses dispositivos: quem grava, quando grava, como se grava, quem acessa as imagens, por quanto tempo elas são armazenadas e com qual finalidade.

Para isso, foram enviados questionários com 23 perguntas padronizadas a todas as Polícias Militares, Civis, Penais, bem como à Polícia Federal e à Polícia Rodoviária Federal. As respostas, quando fornecidas, variaram em completude e qualidade: algumas instituições atenderam prontamente

(1) Tal manancial crítico pode ser antecipado desde algumas emergências por nós já produzidas. Em especial, relativo ao assunto, cf. Amaral (2020, 2024); Amaral e Ferreira (2024, p. 45-58); Amaral et al. (2022a, 2022b); Amaral e Medina (2022, p. 221-242) e Amaral, Elesbão e Dias (2021, p. 667-698).

com dados detalhados; outras, com respostas vagas, genéricas ou intencionalmente evasivas. Em certos casos, foi necessário acionar o Ministério Público para assegurar o cumprimento da Lei de Acesso à Informação.

A análise baseou-se exclusivamente nas respostas recebidas até o final de 2024, compondo assim um recorte temporal claro e metodologicamente delimitado [2]. As informações coletadas foram sistematizadas em matrizes comparativas e examinadas de forma qualitativa, com especial atenção à heterogeneidade de critérios adotados quanto ao uso, ao acesso e à finalidade das gravações.

Trata-se, portanto, de um retrato parcial de uma política pública ainda em construção. Os resultados apresentados constituem etapa intermediária de uma pesquisa em andamento, desenvolvida no âmbito do Grupo de Pesquisa "Politicrim" da PUCRS e financiada pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul. O trabalho busca contribuir para o debate público e acadêmico sobre os desafios normativos, operacionais e éticos do uso institucional dessas tecnologias no campo da segurança pública.

Com vistas a garantir transparência e permitir a verificação dos dados produzidos por parte da comunidade científica, o conjunto integral de respostas recebidas utilizadas na análise está publicamente disponível em pasta compartilhada [3].

2. Panorama geral de uso e distribuição das câmeras corporais

O primeiro dado que salta aos olhos diz respeito à desigualdade de implementação da tecnologia entre os entes e entre os diferentes tipos de forças policiais. As Polícias Militares concentram quase a totalidade dos dispositi-

(2) A Polícia Penal do Estado do Ceará (CE) não respondeu aos ofícios durante o período de coleta de dados, sendo necessário abrir um processo administrativo para obter respostas ao pedido de acesso à informação sobre o uso de câmeras corporais nas Unidades Prisionais do Estado do Ceará. Em 18 de fevereiro de 2025, a 106ª Promotoria de Justiça de Fortaleza/CE – 1ª Promotoria de Justiça de Corregedoria de Presídios e Penas Alternativas – MPCE determinou que a Secretaria de Administração Penitenciária e Ressocialização (SAPR) prestasse esclarecimentos sobre a recusa de resposta à pesquisa. Em 25 de fevereiro de 2025, em cumprimento à determinação do Ministério Público, no Procedimento Administrativo 01.2025.00002843-8, obtivemos retorno. Por essa razão, as informações não estão integradas aos infográficos. Por outro lado, a Polícia Militar do Estado do Rio de Janeiro (PMERJ) negou acesso às informações sobre o uso de câmeras corporais, justificando-se com a Resolução SEPM 278/2020, que impõe exigências como apresentação de projeto de pesquisa, autorização do reitor da instituição e compromisso de fornecer cópia do estudo final mencionando a contribuição da Secretaria. A 6ª Promotoria de Justiça de Tutela Coletiva do Patrimônio Público e da Cidadania do Ministério Público do Estado do Rio de Janeiro (MPRJ) considerou tais exigências ilegais e incompatíveis com a Lei de Acesso à Informação (Lei 12.527/2011) e o Decreto Estadual 46.475/2018. Em 2 de setembro de 2024, a promotoria encaminhou ofício ao Tribunal de Contas do Estado do Rio de Janeiro (TCERJ) e ao MPRJ, solicitando a apuração da legalidade da norma e a suspensão de seus efeitos, além da uniformização das práticas de transparência entre os órgãos do Estado.

(3) Disponível em: https://abre.ai/mskO.

vos em uso no País: são 28.182 câmeras ativas, incluindo aquelas utilizadas em projetos-piloto. É nelas que se encontra o grosso dos investimentos e das iniciativas voltadas à gravação de abordagens e rotinas operacionais.

As Polícias Civis, por outro lado, aparecem de forma bastante tímida: apenas três estados —Acre, Pará e Rio de Janeiro— informaram o uso de câmeras corporais, totalizando 132 equipamentos. Já entre as Polícias Penais, foram identificadas 554 câmeras em uso, sendo a imensa maioria delas concentradas no estado do Pará, com 500 dispositivos, seguido por Rio Grande do Sul (32) e Amazonas (12). O Ceará, embora tenha informado a existência de equipamentos, não apresentou a quantidade exata durante o período de coleta.

Nas polícias da União, o cenário ainda é incipiente. Nem a Polícia Federal, nem a Polícia Rodoviária Federal, tampouco a Força Nacional informou números concretos de câmeras em uso. Todas alegaram estar em fase de estudos ou de projetos-piloto, o que indica que o uso institucional da tecnologia permanece, até o momento, em estágio preliminar na esfera da União.

3. Quem acessa as imagens?

Uma das perguntas mais relevantes sobre o uso de câmeras corporais —e talvez uma das menos discutidas— é quem, afinal, pode assistir às imagens gravadas. A pesquisa revelou um cenário disperso e sem diretrizes comuns. Em alguns estados, o acesso dos agentes às gravações é imediato, sem necessidade de autorização prévia, como nas Polícias Militares de São Paulo, Minas Gerais, Pará, Rio de Janeiro e Amazonas. Em outros, há uma camada intermediária de controle: o agente precisa solicitar permissão superior, como ocorre nas Polícias Militares do Paraná, Acre, Pernambuco e Roraima. Há ainda casos em que o acesso é completamente vedado, como nas Polícias Militares do Rio Grande do Sul e do Rio Grande do Norte, e na Polícia Penal do Rio Grande do Sul.

As diferenças entre esses regimes de acesso não são apenas burocráticas: dizem respeito ao modelo de controle interno que cada corporação adota — e, principalmente, ao equilíbrio entre fiscalização e autonomia. Um sistema em que o próprio policial pode assistir às imagens imediatamente, inclusive da própria atuação, é substancialmente diferente de outro em que esse conteúdo só pode ser acessado por instâncias superiores ou departamentos internos. A ausência de uma política nacional sobre o tema abre margem para distorções: em alguns estados, o agente pode revisar a própria atuação antes mesmo de prestar depoimento. Em outros, ele sequer tem acesso ao que foi gravado. Essa assimetria levanta dúvidas sobre a eficácia dos controles e sobre o grau de independência dos mecanismos internos de correição.

4. O que realmente está sendo gravado?

Nem toda câmera ligada está, de fato, gravando. Esse foi um dos principais pontos de atenção ao longo da pesquisa. Embora vários estados afirmem adotar o chamado "uso contínuo" das câmeras corporais —como as Polícias Militares do Rio Grande do Sul, Paraná, Rio de Janeiro, Pará, Rondônia e Roraima—, na prática, o que existe é o funcionamento constante do dispositivo, mas não necessariamente a gravação e o armazenamento integral das imagens. Em muitos casos, apenas interações específicas, como abordagens, perseguições ou ordens superiores, são gravadas em alta resolução, com áudio, e efetivamente salvas. O restante do tempo é preenchido por registros de baixa qualidade que, após poucos dias, são automaticamente sobrescritos, a menos que sejam acionados manualmente pelo agente.

Por outro lado, há estados em que o uso é deliberadamente restrito a eventos determinados. As Polícias Militares de Minas Gerais, Pernambuco, Sergipe e Rio Grande do Norte, por exemplo, permitem a interrupção da gravação em situações sensíveis — como a presença de crianças, crimes sexuais ou a pedido da vítima. Já as Polícias Civis do Rio Grande do Sul, Rio de Janeiro e Acre utilizam as câmeras apenas em ações específicas, como cumprimento de mandados e grandes operações. No sistema penal, o uso também é limitado a procedimentos de revista ou controle de acesso, como no Amazonas.

5. Quanto tempo os dados ficam armazenados?

A durabilidade das imagens é outro ponto de grande variação entre os estados. Há corporações que mantêm os registros por apenas trinta dias, como a Polícia Penal do Amazonas e a Polícia Militar de Pernambuco — período extremamente curto, sobretudo em casos em que a denúncia ou a investigação demora a ser formalizada. Outras forças, como as Polícias Militares do Paraná, São Paulo, Rio de Janeiro, Pará e Rio Grande do Sul, adotam prazos intermediários, com retenção entre sessenta dias e um ano. Em alguns desses estados, esse intervalo pode variar conforme o tipo de gravação: interações consideradas "intencionais" ou de interesse institucional tendem a ser armazenadas por mais tempo do que gravações feitas em modo automático ou de rotina.

Em menor número, alguns estados optam por armazenamentos prolongados. As Polícias Militares do Rio Grande do Norte e de Sergipe informaram manter os registros por até três anos. Já as Polícias Civis do Acre e do Pará lideram nesse quesito, com políticas que preveem a guarda das imagens por até 5 anos. A falta de padronização nacional nesse aspecto tem efeitos práticos importantes. Quando o tempo de retenção é reduzido, o potencial probatório da câmera se esvazia: provas podem desaparecer antes

mesmo de serem requisitadas e o controle institucional torna-se simbólico. Em resumo, não basta filmar — é preciso garantir que a imagem esteja disponível quando for necessária.

6. Quanto custam as câmeras corporais?

O investimento público na implementação das câmeras corporais também revela um cenário marcado por contrastes. Há estados que optaram pela compra direta dos equipamentos, como as Polícias Militares de Minas Gerais, Rio Grande do Norte, Rio Grande do Sul e Sergipe, além das Polícias Penais do Pará e do Rio Grande do Sul. Essa modalidade tende a exigir um desembolso inicial maior, mas oferece maior controle sobre os dispositivos e os dados gerados. Por outro lado, estados como São Paulo, Rio de Janeiro, Pará, Rondônia e Roraima adotaram o modelo de locação, geralmente vinculado a contratos que incluem serviços de armazenamento em nuvem e manutenção — o que, a médio e longo prazo, pode gerar custos recorrentes muito superiores. Há ainda casos pontuais de doação, como na Bahia e no Rio Grande do Norte, onde o número de equipamentos é pequeno e a autonomia sobre a política, bastante reduzida.

Os valores também variam de forma expressiva. A Polícia Militar do Rio de Janeiro lidera o ranking, com um contrato de R$ 125,4 milhões. Já estados como Roraima (PM), Pernambuco (PM) e o Rio Grande do Norte (PM) registraram investimentos abaixo de R$ 600 mil. No caso das Polícias Civis, o Rio de Janeiro firmou um contrato de R$ 888 mil com vigência de 30 meses, enquanto o Acre destinou R$ 72 mil para um contrato de 6 meses. Entre as Polícias Penais, o Rio Grande do Sul investiu pouco mais de R$ 34 mil em sua política. A disparidade nos custos, somada à ausência de métricas públicas de avaliação de impacto, levanta uma questão incômoda: o que justifica tamanha variação? Sem indicadores de eficácia, o investimento em câmeras pode acabar servindo mais como vitrine do que como ferramenta real de controle.

7. Onde as câmeras estão sendo usadas?

Nem todo policial que vai para a rua usa câmera. E, entre os que usam, nem todos estão envolvidos nas situações mais críticas. A pesquisa revelou que, em muitos estados, as câmeras são distribuídas prioritariamente às patrulhas regulares, deixando de fora justamente os batalhões táticos e grupos especializados — aqueles que atuam em operações de maior complexidade e risco. É o caso das Polícias Militares do Rio Grande do Sul, Paraná e Pará, além da Polícia Penal do Rio Grande do Sul, onde o uso da tecnologia ainda está restrito a agentes em ações ostensivas convencionais.

Por outro lado, alguns estados decidiram equipar também os grupos táticos e especializados. As Polícias Militares de São Paulo, Minas Gerais, Rio de Janeiro, Bahia, Roraima, Pernambuco e Rio Grande do Norte, bem como as Polícias Civis do Acre e do Rio de Janeiro, informaram o uso das câmeras nesses contextos. Essa escolha tem implicações práticas significativas: limitar o uso das *bodycams* a patrulhas de rotina pode fazer com que as situações mais críticas — justamente aquelas que mais exigem controle externo e transparência— escapem do registro visual. A alocação seletiva, quando não é tecnicamente justificada, enfraquece o próprio sentido da política pública. Afinal, o controle que evita ser visto dificilmente será efetivo.

8. Quais os critérios para o uso das câmeras?

Apesar de se tratarem de uma tecnologia com alto potencial de impacto sobre a atuação policial, os critérios que definem quando, como e por que as câmeras corporais devem ser utilizadas variam amplamente entre os estados. Identificamos quatro modelos principais: (a) uso contínuo durante todo o turno, (b) uso apenas durante ocorrências, (c) uso restrito a missões previamente planejadas e (d) uso condicionado à quantidade de equipamentos disponíveis. Cada modelo carrega implicações operacionais distintas — e riscos distintos também.

Estados como Rio Grande do Sul, Paraná, São Paulo, Rio de Janeiro, Pará e Roraima (todos no âmbito das Polícias Militares) informaram adotar o uso contínuo, ativando os dispositivos no início da jornada e mantendo-os ligados até o final do turno. Já as Polícias Militares de Minas Gerais, Pernambuco e Rio Grande do Norte, assim como as Polícias Civis do Acre, Rio de Janeiro e Rio Grande do Sul e a Polícia Penal do Amazonas restringem o uso a determinadas ocorrências ou operações específicas. Nesses casos, a gravação pode ser acionada ou interrompida pelo próprio agente, o que abre margem para lacunas relevantes nos registros. Em algumas unidades, como a Polícia Militar do Rio Grande do Norte, o critério é ainda mais simples: as câmeras são usadas por quem tiver acesso, dependendo da disponibilidade dos equipamentos. A ausência de padronização nacional, mais uma vez, cria um quadro de insegurança e desigualdade — tanto para os agentes quanto para os cidadãos.

9. As câmeras são algum resultado?

Apesar dos investimentos significativos e do crescente uso das câmeras corporais, poucos estados possuem algum tipo de estudo avaliativo sobre os impactos concretos da política. Raríssimas instituições realizaram análises sistemáticas sobre o efeito das gravações na redução de abusos, no número de investigações internas ou mesmo na percepção pública sobre a atividade-

de policial. Onde existem relatórios, eles são parciais, pouco detalhados ou metodologicamente frágeis.

A Polícia Militar de Minas Gerais, por exemplo, informou ter feito uma análise de variação nos números de procedimentos disciplinares, mas sem estabelecer vínculo direto com o uso das câmeras. A Polícia Civil do Acre disse não ter identificado alterações nos índices. A Polícia Militar do Rio de Janeiro realizou estudos internos, mas não compartilhou dados objetivos. Na maior parte dos estados, no entanto, não há sequer tentativa de mensurar resultados. Sem dados, o que se tem é apenas suposição — e uma política pública sem avaliação sistemática corre o risco de ser apenas uma boa intenção sem eficácia comprovada. Em um cenário de gastos elevados e desafios estruturais, seguir investindo sem medir resultados é, no mínimo, uma temeridade.

10. Onde os projetos foram descontinuados?

Santa Catarina, que foi o primeiro estado brasileiro a implementar o uso das *bodycams* em ações policiais, também foi o primeiro a abandonar o projeto. Segundo informações oficiais, a decisão se deu por falhas técnicas, limitações operacionais e pelo não atingimento dos objetivos esperados. Mas também, é importante que se diga, houve interrupção do aporte de recursos do Tribunal de Justiça de Santa Catarina (TJSC). O projeto, que chegou a ser apresentado como exemplo de inovação, foi encerrado sem qualquer relatório público de avaliação de impacto ou lições aprendidas.

Em Rondônia, o cenário foi semelhante. Após a adoção inicial das câmeras em caráter experimental, os dispositivos acabaram entrando em desuso por desgaste natural e ausência de renovação contratual. À época da pesquisa, o estado se encontrava em fase de aquisição de novos equipamentos, tentando retomar o programa. Esses casos reforçam a necessidade de planejamento, acompanhamento técnico e definição clara de objetivos. Sem isso, a política de câmeras corre o risco de se tornar um ciclo de entusiasmo inicial seguido de abandono silencioso — com alto custo e pouca entrega.

11. Quais entes federativos estão em fase de implementação ou expansão?

Ao mesmo tempo em que alguns entes federativos descontinuam o uso das câmeras, outros estão justamente no movimento oposto: iniciando ou ampliando seus projetos. A Polícia Federal, a Polícia Rodoviária Federal e a Polícia Penal Federal informaram estar em fase de estudos para adoção da tecnologia. Já no plano estadual, as Polícias Militares do Rio Grande do Sul, Espírito Santo, Minas Gerais, Mato Grosso do Sul, Acre, Amazonas, Pará,

Tocantins, Bahia, Ceará e Pernambuco relataram processos de expansão, seja pela aquisição de novos dispositivos, seja pela ampliação do número de agentes que os utilizam.

Alguns estados já iniciaram implementações recentes, como Distrito Federal, Alagoas e Maranhão. No âmbito das Polícias Civis, os estados de Goiás e Alagoas estão em fase de ampliação, enquanto o Rio de Janeiro afirmou estar consolidando sua implementação. Entre as Polícias Penais, os projetos de expansão foram relatados por Paraná, Goiás, Acre, Alagoas, Maranhão, Roraima e Paraíba. A expectativa é que esses números cresçam rapidamente a partir do "Projeto Nacional de Câmeras Corporais", iniciativa do Governo Federal que destinou R$ 102 milhões à aquisição de mais de 52 mil equipamentos para as Polícias Militares de 20 estados e do Distrito Federal. A depender da forma como essa expansão for conduzida —com ou sem diretrizes nacionais claras—, ela poderá reduzir desigualdades ou simplesmente replicar os mesmos problemas já identificados na pesquisa.

12. Conclusão

A análise preliminar conduzida neste levantamento empírico revela que o uso de câmeras corporais pelas polícias brasileiras ainda não constitui uma política pública minimamente estruturada e homogênea em nível nacional. O que se observa, ao contrário, é um cenário profundamente fragmentado, marcado por assimetrias operacionais, lacunas regulatórias e ausência de padronização quanto a critérios de ativação, acesso, armazenamento e finalidade das imagens captadas. A heterogeneidade institucional é visível tanto no plano quantitativo —com estados que operam milhares de dispositivos ao lado de outros que sequer iniciaram projetos-piloto— quanto no plano qualitativo, refletido nas distintas formas de controle interno, transparência e supervisão das imagens. Em certos estados, as câmeras são utilizadas de forma contínua durante todo o turno; em outros, sua ativação depende da discricionariedade do próprio agente. O tempo de retenção das imagens varia de trinta dias a cinco anos, sem justificativas públicas claras. O investimento financeiro, por sua vez, oscila de pouco mais de R$ 30 mil a contratos superiores a R$ 125 milhões — muitas vezes sem qualquer correspondência com a escala ou o impacto da política implementada.

A escassez de avaliações formais sobre os efeitos das câmeras também chama atenção. Poucas instituições se dispuseram a mensurar impactos sobre práticas policiais, responsabilização por abusos ou redução de violações de direitos. Na maioria dos casos, sequer há dados disponíveis. Essa ausência de monitoramento compromete a capacidade do poder público de aferir a efetividade da política, tornando difícil distinguir entre iniciativas simbólicas e instrumentos reais de controle e prevenção.

Importa destacar, no entanto, que esta pesquisa oferece apenas uma fotografia parcial da realidade nacional. Em primeiro lugar, porque o próprio objeto de análise encontra-se em constante transformação: novos projetos estão em fase de implementação, expansão ou reformulação, o que impede conclusões estáticas ou generalizadoras. Em segundo lugar, porque a pesquisa ainda está em curso, e seus resultados finais serão apresentados em versão ampliada e sistematizada em publicações futuras.

Apesar dessas limitações, os dados aqui sistematizados fornecem uma base empírica robusta para o debate acadêmico e institucional sobre os rumos da política de câmeras corporais no Brasil.

Referências

Amaral, A. J. (org.). (2022). *A cidade como máquina biopolítica*. São Paulo: Tirant lo Blanch.

Sabariego, J., Amaral, J. J. do, Carvalho, E. B., & Salles-Amaral, A. (Orgs.). (2021). *Algoritarismos*. São Paulo: Tirant lo Blanch

Amaral, A. J., & Dias, F. V. (2024). *Tecnopolítica criminal*. São Paulo: Tirant lo Blanch.

Amaral, A. J., Elesbão, A. C. S., & Dias, F. V. (2021). Governamentalidade algorítmica e novas práticas punitivas. *Revista Derechos en Acción*, v. 30, n. 20, p. 667-698. https://doi. org/10.24215/25251678e549

Amaral, A. J., & Medina, R. (2022). As máquinas de visão cibernéticas e o advento de um novo regime de verdade. *Revista de Estudios Foucautianos*, *n. 12, p. 221-242*. https://doi.org/10.5281/zenodo.6786135

La inteligencia artificial en el derecho paraguayo: oportunidades y desafíos

Por Jorge Luis Usher Armoa [(*)] *y Ángela Gisel Gamarra Schulz* [(**)]

Resumen: El presente trabajo analiza la incipiente adopción de la inteligencia artificial (IA) en el sistema judicial de Paraguay, explorando las tensiones entre la promesa de eficiencia y los riesgos estructurales que plantea para el Estado de derecho. Se examina cómo la transformación digital, acelerada por la expansión del Expediente Judicial Electrónico, ha sentado las bases para la integración de la inteligencia artificial, una tecnología que trasciende la mera gestión documental para optimizar procesos cognitivos como la redacción de escritos y el análisis de sentencias. El documento aborda la adopción de herramientas como "AuroraLex" tanto en el sector público, a través del Ministerio de la Defensa Pública, como en el privado, y la respuesta institucional del Poder Judicial mediante la capacitación ética de magistrados. Se destaca la paradoja de una implementación fragmentada y prematura, que avanza sin un marco legislativo de gobernanza plenamente definido. La doctrina paraguaya se presenta como un contrapeso de cautela, señalando preocupaciones críticas en materia de ética, privacidad y responsabilidad, evidenciadas por la sanción al uso indebido y la autorregulación del sector privado ante la debilidad institucional. Se analizan desafíos fundamentales como la asimetría procesal derivada de la brecha tecnológica, el riesgo de sesgo algorítmico que podría perpetuar injusticias sociales, la brecha digital que amenaza el acceso universal a la justicia y la potencial erosión de la independencia judicial. Finalmente, se concluye que, ante la ausencia de una regulación específica, el sistema judicial recurre a mecanismos de control ético y disciplinario como respuestas institu-

---

(*) Abogado por la Universidad Nacional de Itapúa, licenciado en Ciencias de la Comunicación con énfasis en Periodismo y especialista en Docencia Universitaria, ambos por la Facultad de Filosofía y Ciencias Humanas de la Universidad Católica "Nuestra Señora de la Asunción" Campus Itapúa. Abogado litigante. Coordinador académico de la Facultad de Filosofía y Ciencias Humanas de la Universidad Católica "Nuestra Señora de la Asunción" Campus Itapúa. abg.lic.ju@gmail.com

(**) Abogada, notaria y escribana pública por la Facultad de Ciencias Jurídicas de la Universidad Nacional de Itapúa. Especialista en Docencia Universitaria, especialista en Derecho Civil, por la Universidad Nacional de Itapúa, magister en Tributación y Asesoría Impositiva (MAETAI) por el Instituto Superior de Formación Tributaria y Empresarial (FOTRIEM). Actuaria judicial del Poder Judicial en la Tercera Circunscripción Judicial de la República del Paraguay. schulzgamarra@gmail.com

cionales para garantizar la primacía del control humano y la integridad del proceso judicial.

Palabras clave: inteligencia artificial; sistema judicial; ética; derecho; regulación.

Abstract: This paper analyzes the incipient adoption of Artificial Intelligence (AI) within the Paraguayan judicial system, exploring the tensions between the promise of efficiency and the structural risks it poses to the rule of law. It examines how the digital transformation, accelerated by the expansion of the Electronic Judicial File, has laid the groundwork for the integration of AI, a technology that transcends mere document management to optimize cognitive processes such as legal drafting and the analysis of judicial sentences. The document addresses the adoption of tools like "AuroraLex" in both the public sector, through the Public Defender's Ministry, and the private sector, and the institutional response of the Judiciary through the ethical training of magistrates. It highlights the paradox of a fragmented and premature implementation, which is advancing without a fully defined legislative governance framework. Paraguayan doctrine is presented as a cautious counterweight, pointing out critical concerns regarding ethics, privacy, and liability, evidenced by the sanctioning of improper use and the self-regulation of the private sector in the face of institutional weakness. Fundamental challenges are analyzed, such as procedural asymmetry derived from the technological gap, the risk of algorithmic bias that could perpetuate social injustices, the digital divide that threatens universal access to justice, and the potential erosion of judicial independence. Finally, it is concluded that, in the absence of specific regulation, the judicial system is resorting to ethical and disciplinary control mechanisms as institutional responses to ensure the primacy of human oversight and the integrity of the judicial process.

Key Words: Artificial Intelligence; judicial system; ethics; law; regulation.

1. Introducción

El sistema judicial paraguayo se ha ido ajustando a las transformaciones tecnológicas en los últimos años, principalmente en su faz operativa (Caballero Galeano, 2024), siendo uno de sus antecedentes directos el proceso de paulatina expansión del Expediente Judicial Electrónico que, a raíz de la pandemia del COVID-19 por decisión de la Corte Suprema de Justicia se ha descentralizado e implementado en las 16 circunscripciones de la República del Paraguay, beneficiando a Juzgados de zonas urbanas y rurales, sentando, a su vez, la base para las reformas y actualizaciones siguientes (Díaz et al., 2022).

Siendo la inteligencia artificial, un evidente salto cualitativo que supera la mera gestión documental y los expedientes digitales pues, para enfocarse

en la potenciación y optimización de los meros procesos cognitivos como la redacción de documentos jurídicos y el análisis de sentencias judiciales, venturando aliviar la carga procesal (Calvo de Mora Rivas & Factor Santoveña, 2024) Por lo que, la doctrina paraguaya ha identificado a este nuevo fenómeno como un conjunto de desafíos y oportunidades para el futuro jurídico del país, con la mira en las implicancias desde la ética, la legislación, la privacidad y la responsabilidad civil yacentes. (Caballero Galeano, 2024; Ortiz Cristaldo, 2022).

Por consiguiente, la promesa de eficiencia en el uso de la inteligencia artificial, se encuentra con los riesgos que la academia ha señalado para el Estado de derecho, siendo una preocupación y critica común la ausencia de un marco legislativo robusto, tendiente a la creación de vulnerabilidades significativas en cuanto a la protección de los derechos fundamentales (Calvo de Mora Rivas & Factor Santoveña, 2024). Resultando al inteligencia artificial, no solo una herramienta técnica, sino un desafío estructural para los principios consagrados en la Carta Magna.

2. Adopción institucional incipiente de la inteligencia artificial

La adopción de la inteligencia artificial se da tanto el ámbito público como en el privado. A nivel institucional, el Ministerio de la Defensa Pública del Paraguay ha incorporado herramientas tecnológicas como el proyecto "AuroraLex" para optimizar la labor de los defensores, posibilitándoles la redacción de escritos y el análisis de documentos de manera más ágil, fortaleciendo el acceso a la justicia (Ministerio de la Defensa Pública, 2024). Al igual que otras herramientas específicas como "Prometea" que, a pesar de su origen argentino, ha sido analizado su uso por la Sala Constitucional (Sequera Buzarquis, 2021). En el sector privado, la ya mencionada "AuroraLex" ofrece múltiples funciones de asistencia legal, también para todos los abogados matriculados del Paraguay (AuroraLex, 2025).

Y en atención a ello, como respuesta a los riesgos inminentes que pudiesen devenirse por su aplicación, el Poder Judicial ha tomado como medida la capacitación de magistrados en el uso ético de la inteligencia artificial (UNESCO, 2024). Reconociendo así la magnitud del desafío que representa para el Estado de derecho y subrayando la necesidad de que el factor humano mantenga el control en su empleo.

Empero, esta utilización está marcada por una paradoja de implementación fragmentada y prematura, vale decir que, la evidente necesidad de mejorar la eficiencia judicial (Ministerio de la Defensa Pública, 2024) impulsa la adopción de sistemas de inteligencia artificial antes de que la legisla-

ción de gobernanza esté plenamente definida (MITIC, 2025; Calvo de Mora Rivas & Factor Santoveña, 2024).

3. La inteligencia artificial como vehículo de eficiencia y optimización

El panorama nacional reconoce a la inteligencia artificial como un envión de desarrollo, resaltado por autores paraguayos que ponderan consistentemente su potencial transformador para la práctica legal, por su capacidad para optimizar tareas repetitivas y mejorar la eficiencia operativa en despachos y tribunales (Caballero Galeano, 2024; Calvo de Mora Rivas & Factor Santoveña, 2024). Esta visión optimista como oportunidad (Caballero Galeano, 2024) se alinea con la intención del Ministerio de Tecnologías de la Información y Comunicación (MITIC) de impulsar un proyecto de ley enfocado en promover el uso de la inteligencia artificial para el desarrollo nacional (MITIC, 2025).

4. Aspectos críticos de la cautela doctrinal: Ética, Privacidad y Responsabilidad

Frente a la promesa de eficiencia, por otra parte, la doctrina paraguaya actúa como un contrapeso de cautela, urgiendo la creación de una legislación que garantice la supervisión humana y la protección de derechos fundamentales (Calvo de Mora Rivas & Factor Santoveña, 2024). Esta necesidad de regulación ha sido probada de facto en el ámbito nacional con la aparición de casos concretos, como la sanción impuesta a un abogado por el uso inapropiado de inteligencia artificial, que evidencia la necesidad urgente de disposiciones claras al respecto (Calvo de Mora Rivas & Factor Santoveña, 2024).

Ante la ausencia de un marco legal específico, el Poder Judicial se ha visto obligado a recurrir a su facultad disciplinaria (Ortiz Cristaldo, 2022; El Observador, 2025) para llenar el vacío normativo. Este proceso conlleva de emergencia al uso del Derecho Disciplinario o Sistema Disciplinario del Poder Judicial como regulador de facto del uso de la inteligencia artificial en el litigio, definiéndose las medidas caso por caso mediante sanciones ex post, lo cual genera inseguridad jurídica para los operadores, ya que carecen de directrices claras con anterioridad.

En el ámbito de la privacidad, el debate es activo. A pesar de que la protección de datos personales es reconocida como un derecho personalísimo en el Paraguay (TEDIC, 2024), la debilidad institucional en la gobernanza ha llevado al sector privado a adoptar medidas de mitigación. La plataforma AuroraLex, por ejemplo, restringe su uso a abogados matriculados en Paraguay y enfatiza explícitamente que los datos proporcionados son privados y

no se utilizan para entrenar su modelo, con el fin de generar confianza ante el vacío de control estatal (AuroraLex, 2025). Esta acción demuestra que el mercado se ve impelido a asumir una responsabilidad ética que debería ser garantizada por el Estado.

5. La propiedad intelectual como espejo de la personalidad jurídica de la IA

Un punto crucial de la mirada nacional es la definición de la naturaleza jurídica de la inteligencia artificial., pues la Ley 1.328/98 de Derecho de Autor y Derechos Conexos establece que la autoría está reservada exclusivamente a personas físicas (*Revista Científica UNIDA*, 2024; UPE, 2025). Este marco legal la restringe a un rol de herramienta o auxiliar. Por consiguiente, si genera contenido de forma totalmente autónoma, sin intervención humana sustancial, esa obra carece de autor legalmente reconocido en Paraguay y, por ende, no goza de protección legal (UPE, 2025). Esta postura, alineada con tratados internacionales en la materia, es vital, ya que reafirma la primacía del control humano sobre la producción jurídica y creativa.

6. Desafíos fundamentales de la inteligencia artificial y su impacto en el proceso judicial paraguayo

Uno de los desafíos más críticamente analizados por la academia paraguaya es el riesgo de asimetría procesal o desigualdad de herramientas. Los autores han señalado que la inteligencia artificial introduce una clara desventaja tecnológica cuando un litigante utiliza recursos de inteligencia artificial para procesar rápidamente grandes volúmenes de documentos, mientras que la contraparte se ve obligada a realizar el mismo trabajo de forma manual (Allen, s.f.; SOURDIN & CORNES, 2018, citado en Allen, s.f.).

La existencia de servicios de inteligencia artificial específicos y de pago en el mercado local, como AuroraLex (AuroraLex, 2025), agrava esta preocupación, ya que el acceso a la eficiencia procesal se convierte en una función de la capacidad económica. Esta situación tiende a generar un fenómeno de privatización del debido proceso, donde el estándar de diligencia procesal avanzado es dictado por el poder adquisitivo, poniendo en jaque el principio de acceso equitativo a la justicia. La incorporación de herramientas en el Ministerio de la Defensa Pública (Ministerio de la Defensa Pública, 2024) es una medida necesaria para mitigar esta brecha, buscando asegurar que la asistencia legal pública pueda competir tecnológicamente con el sector privado.

7. Sesgo algorítmico y discriminación socio-judicial

La amenaza del sesgo algorítmico no es solo un problema técnico; es la digitalización de la injusticia social. La investigación académica nacional llama a un deber colectivo no solo de detectar el sesgo, sino de "abordar y remediar sus orígenes sociales" (LATAM DIGITAL, 2022).

Para el contexto paraguayo, esto implica una dificultad considerable, pues los algoritmos de asistencia judicial inevitablemente serán entrenados con datos históricos reflejados de un sistema judicial que arrastra problemas crónicos de acceso y percepción de corrupción (ABC Color, 2023; Judiciales.net, s.f.) Al automatizar decisiones basadas en estos datos, existe el riesgo de que la inteligencia artificial perpetúe y escale las estructuras de discriminación socioeconómica, de género o geográfica preexistentes.

Para contrarrestar esta causalidad histórica, su implementación debe ir más allá de las auditorías técnicas, por lo que es indispensable adoptar un enfoque de diseño sensible al valor (*Value Sensitive Design* - VSD), que exige que los valores constitucionales paraguayos (como la no discriminación) sean incorporados explícitamente en la fase de conceptualización y diseño del algoritmo, mitigando activamente los sesgos que provienen de los datos históricos.

8. Brecha digital y acceso universal a la justicia

A pesar de los logros en la expansión del Expediente Judicial Electrónico a 16 circunscripciones de la República (Díaz et al., 2022), la brecha digital sigue siendo un obstáculo crucial para el acceso universal a la justicia, pues estudios realizados continúan señalando la existencia de esta distancia, afectando especialmente a grupos vulnerables como las personas mayores (Mendieta et al., 2023).

A lo que, una justicia cada vez más dependiente de la inteligencia artificial debe considerar la perspectiva del ciudadano como "no-usuario" (LATAM DIGITAL, 2022) fundando en que si bien el Expediente Judicial Electrónico amplió el uso de tecnologías existentes (Díaz et al., 2022), la inteligencia artificial requiere un nivel superior de alfabetización digital y conectividad que podría ser inaccesible para una porción significativa de la población paraguaya. Por cuanto que, una automatización judicial que no sea intrínsecamente inclusiva inevitablemente agudizará la marginación de los ciudadanos menos conectados, comprometiendo la universalidad del acceso a la justicia.

9. Riesgos institucionales: independencia judicial ante la dependencia algorítmica

El Poder Judicial de Paraguay goza de independencia y autarquía plena, con mecanismos claros para la designación de sus miembros a través del Consejo de la Magistratura (Poder Judicial, s.f.). Sin embargo, la irrupción de la inteligencia artificial introduce una sutil, pero potente, amenaza a esta independencia, pues las herramientas de inteligencia artificial utilizadas, especialmente los predictivos o recomendadores, pueden llevar a una colonización algorítmica del criterio humano. A lo que, si un juez se vuelve dependiente de las sugerencias del algoritmo para maximizar la eficiencia o reducir el riesgo de que sus sentencias sean apeladas, se produce una erosión de su autonomía cognitiva, incluso si la decisión final formalmente le sigue perteneciendo (Urueña, 2021; Reyes Torres, 2025). Este riesgo se agrava si los algoritmos son de código cerrado o propiedad de empresas externas, lo cual compromete la soberanía cognitiva judicial.

Otro aspecto de necesaria atención es que la doctrina paraguaya, al citar la normativa internacional, ha reconocido el déficit de la explicabilidad de la inteligencia artificial (XAI), señalando que actualmente no existe un "derecho a la explicación" de las decisiones automatizadas en la legislación paraguaya (Caballero Galeano, 2024; Wachter et al., 2017, citado en Caballero Galeano, 2024).

A esto, la opacidad inherente a muchos modelos de *machine learning* (Urueña, 2021) plantea un serio obstáculo para la responsabilidad. La falta de transparencia y auditabilidad en los procesos algorítmicos impide asignar la responsabilidad civil o disciplinaria de manera clara cuando la IA comete errores (Calvo de Mora Rivas & Factor Santoveña, 2024). Por cuanto, para garantizar el principio de garantía judicial, la futura regulación debe establecer taxativamente el requisito de que todos los modelos de inteligencia artificial utilizados en la función jurisdiccional sean auditables, transparentes y permitan una justificación humana comprensible de la decisión asistida (Miranda, 2022; Simón, 2023, citado en Urueña, s.f.).

10. Control ético y práctico como respuesta institucional

Ante la carencia de un marco legislativo completo y concreto, el Poder Judicial está utilizando los mecanismos de control existentes y proactivos, como la colaboración con la UNESCO para la formación ética de los jueces (UNESCO, 2024) que demuestra una postura institucional que prioriza el control humano y la deliberación ética sobre la mera implementación técnica, reforzando el argumento doctrinal de que la supervisión humana es la

clave de bóveda para una inteligencia artificial judicial responsable (Calvo de Mora Rivas & Factor Santoveña, 2024).

Simultáneamente, la aplicación del poder disciplinario del juez (Ortiz Cristaldo, 2022), evidenciada en la sanción a abogados por uso indebido de IA (Calvo de Mora Rivas & Factor Santoveña, 2024 3), es el mecanismo de control ex post actual para asegurar que se mantenga la integridad y la ética en la práctica legal asistida por tecnología.

11. Aplicación de la inteligencia artificial por actores de justicia

La incipiente aplicación de la inteligencia artificial en el sistema judicial paraguayo se concentra actualmente en la optimización de procesos (eficiencia) y la asistencia técnica al operador jurídico, por ningún motivo en la sustitución de la toma de decisiones, respetando todo principio ético fundamental (UNESCO, 2024; Revistas.pj.gob.pe, s.f.). Siendo la implementación asimétrica y dependiendo de los fueros y las capacidades institucionales.

La inteligencia artificial en el rol de los jueces y la Corte Suprema de Justicia se orienta primariamente a la gestión de macrodatos y la aceleración de tareas repetitivas, cuyo principal beneficio es el ahorro de tiempo (Poder Judicial, 2025).

De igual modo esta máxima instancia tiene la oportunidad de utilizarla para el manejo de la minería de datos jurisprudenciales a gran escala, lo cual es esencial para identificar tendencias, evaluar el rendimiento institucional y lograr una unificación más rápida y efectiva de los criterios judiciales. Esto optimiza la planificación estratégica y la gestión de la calidad judicial (Poder Judicial, 2025 24).

En cuanto a su potencial aplicación multifueros, el Sistema Integral de Tramitación Judicial Electrónica (SITJE) se enfocó inicialmente en los Juzgados de Ejecución Penal (DPL News, 2025). En este sentido, la inteligencia artificial podría ser clave para el seguimiento y control de las Personas Privadas de Libertad (PPL) y el cumplimiento de medidas alternativas (DPL News, 2025 25). No obstante, el riesgo de utilizarla para la evaluación de riesgo procesal (fianzas o libertad condicional) es alto, debido a que el sesgo algorítmico podría perpetuar prejuicios históricos (Urueña, 2021 26).

Para fueros como el civil y laboral caracterizados por la sobrecarga judicial (Calvo de Mora Rivas & Factor Santoveña, 2024 3), el uso potencial se centra en la automatización de despachos de mero trámite y la asistencia en el cálculo de liquidaciones judiciales (siguiendo modelos regionales como Prometea, en la Argentina, que demostró alta eficiencia). Por lo que la inteligencia artificial liberaría al juez para concentrarse en la ponderación de la

prueba, un acto que requiere la sana crítica y no puede ser delegado a una máquina (Poder Judicial, 2025).

En cuanto a su uso por otros actores de justicia, como ya se ha mencionado el Ministerio de la Defensa Pública ha liderado la implementación de inteligencia artificial con el proyecto AuroraLex (Ministerio de la Defensa Pública, 2024). En el Fuero Penal (su ámbito principal, junto a Niñez), esta herramienta digital permite a los defensores acceder rápidamente a información relevante, analizar grandes volúmenes de documentos y redactar escritos de manera más efectiva (Ministerio de la Defensa Pública, 2024). Teniendo como objetivo mejorar la eficiencia y precisión en la resolución de casos, fortaleciendo el acceso a la justicia (Ministerio de la Defensa Pública, 2024).

Por su parte, los estudios jurídicos privados o particulares, utilizan la inteligencia artificial para la gestión de litigios masivos y el análisis predictivo de casos (similares a Lex Machina en otros contextos), principalmente en los fueros Civil y comercial. Si bien esto representa una oportunidad de eficiencia (Calvo de Mora Rivas & Factor Santoveña, 2024), genera la principal amenaza a la equidad procesal, cual es la asimetría tecnológica (Allen, s.f.).

Y aunque no se ha detallado un proyecto masivo de inteligencia artificial en el Ministerio Público, el uso potencial reside en la clasificación y pre análisis de denuncias y la asistencia en la revisión de evidencia digital durante la etapa de investigación penal. No obstante, su utilización para la priorización de casos debe manejarse con extrema cautela para evitar la reproducción de sesgos (Urueña, 2021).

De igual modo, El Poder Judicial dentro de la promoción a la mediación (Poder Judicial, 2025) ve en la inteligencia artificial una fortaleza para poder asistir a los mediadores en la identificación de rangos de acuerdo y la propuesta de soluciones extrajudiciales basadas en precedentes, acelerando así los Mecanismos Alternativos de Solución de Conflictos (MASC) en los fueros Civil y Laboral.

Como eslabón principal, el ciudadano se beneficia indirectamente de la inteligencia artificial a través de la mayor eficiencia administrativa (Ministerio de la Defensa Pública, 2024). En términos de uso directo, este podría interactuar con *chatbots* o asistentes virtuales para obtener orientación procesal básica y la clarificación de sentencias en lenguaje sencillo, tal como se ha explorado en jurisdicciones como Perú (Revistas.pj.gob.pe, s.f.; Calvo de Mora Rivas & Factor Santoveña, 2024). Sin embargo, la principal amenaza para el ciudadano sigue siendo la falta de explicabilidad (XAI) en las decisiones asistidas por IA (Caballero Galeano, 2024 1), lo que compromete su derecho a una defensa efectiva frente a un sistema opaco (Urueña, s.f.).

12. Comparativa con países de la región

Brasil ha sido pionero en la informatización (Ley Federal 11.419/2006) y el uso de la inteligencia artificial a través de su Consejo Nacional de Justicia (CNJ) (Unidos en Red, s.f.) utilizando sistemas de *machine learning* configurados para trabajar con el Proceso Judicial Electrónico (PJE).

Por una parte, el sistema permitió la racionalización del gasto público, el esfuerzo del personal y el uso del tiempo (Unidos en Red, s.f.) Por otra, la experiencia brasileña demuestra que los sistemas de inteligencia artificial, basados en la clasificación y predicción, inevitablemente generan errores, conocidos como "falsos positivos" y "falsos negativos" (Unidos en Red, s.f.), realidad que subraya que el operador judicial debe mantener una actitud crítica, entendiendo que la herramienta digital potencia, pero no sustituye, el cerebro humano (Unidos en Red, s.f. 24).

Argentina, cuenta con Prometea, el sistema de inteligencia artificial más conocido en la región, utilizado por el Poder Judicial argentino, y que ha arrojado resultados positivos en programas piloto de inteligencia artificial generativa (Argentina.gob.ar, 2025).

Esta herramienta digital ha demostrado una alta eficiencia operativa, pues los reportes indican que en la generación de dictámenes, el sistema alcanzó una eficiencia de hasta el 75% (BID, s.f.). Además, de garantizar la ausencia de errores de tipeo o cálculo en la confección de despachos (Gauna et al., 2021).

Colombia, ha protagonizado un hito jurisprudencial, pues el caso más emblemático involucra a un juez que utilizó Chat GPT para sustentar un fallo en un caso de protección al derecho fundamental a la salud de un menor (Franco & Wittenstein, s.f., citado en CETyS, 2024).

La Corte Constitucional de Colombia revisó el caso y determinó que la inteligencia artificial es admisible únicamente como una herramienta de apoyo para la gestión de tareas y la asistencia en la redacción de decisiones judiciales (UNESCO, 2024 30; Revistas.pj.gob.pe, s.f.). No obstante, la Corte prohibió la dependencia excesiva, enfatizando el principio de no sustitución de la racionalidad humana y la necesidad de aplicar juicios de valor que la máquina no puede delegar (UNESCO, 2024). Al citar explícitamente la recomendación de la UNESCO sobre la Ética de la inteligencia artificial (UNESCO, 2024), la Corte demostró la utilización del derecho blando internacional para generar salvaguardas constitucionales inmediatas frente al vacío regulatorio.

13. Derecho a la explicación regional

La experiencia regional ha formalizado el derecho a la explicación como un imperativo ético. El Código Ético Iberoamericano para el uso de la inteligencia artificial exige que su aplicación sea transparente y respetuosa de los derechos humanos (Cumbre Judicial Iberoamericana, 2025).

Este principio se traduce en derechos procesales concretos en propuestas regulatorias avanzadas, como la de Brasil, que exige una explicación detallada de cualquier decisión o predicción hecha por la inteligencia artificial, la cual debe incluir las características operacionales, el grado de contribución del sistema, los datos procesados, los criterios de decisión y los mecanismos para impugnarla (FPF, 2025).

14. Conclusión

El sistema judicial paraguayo se encuentra en una fase de adopción acelerada de la inteligencia artificial, impulsada por la búsqueda de eficiencia heredada de la descentralización del Expediente Judicial Electrónico (Díaz et al., 2022). Sin embargo, esta se produce en un vacío normativo que la doctrina local ha identificado, de manera consistente, como el principal riesgo para la seguridad jurídica (Calvo de Mora Rivas & Factor Santoveña, 2024).

La aportación crítica más significativa de los juristas paraguayos reside en la contextualización del riesgo de desigualdad tecnológica. Los académicos han demostrado cómo la existencia de herramientas de inteligencia artificial privadas genera una asimetría procesal que amenaza directamente el principio de igualdad de herramientas o recursos digitales y el debido proceso, transformando el acceso a la justicia avanzada en una cuestión de poder económico (Allen, s.f.) lo que representa la privatización del estándar de diligencia judicial.

Para mitigar estos desafíos, la experiencia paraguaya pondera la primacía de la ética institucional, a través de la capacitación de jueces (UNESCO, 2024) y la necesidad de proteger la soberanía tecnológica, garantizando que el Poder Judicial mantenga el control sobre sus datos y los modelos algorítmicos utilizados en la función jurisdiccional.

De allí que algunas estrategias proclives y necesarias de implementarse por los actores del sistema parten primeramente del Poder Legislativo del Paraguay, en la adopción de un marco regulatorio que clasifique explícitamente los sistemas de inteligencia artificial utilizados en la administración de justicia y en la interpretación constitucional como sistemas de alto riesgo, sujetos a la mayor exigencia de transparencia y auditabilidad.

De igual modo, establecer como requisito obligatorio una Evaluación de Impacto sobre los Derechos Humanos (DPIA) y auditorías de sesgo algorítmico previo a la implementación de cualquier sistema de inteligencia artificial en los tribunales del país.

Y que, a través de la misma normativa, se pueda establecer un régimen de responsabilidad civil y profesional específico para el daño algorítmico, que defina con claridad la imputación de daños derivados de errores o fallas en el uso de sistemas automatizados (Calvo de Mora Rivas & Factor Santoveña, 2024).

A su vez, por parte de la Corte Suprema de Justicia, la emisión de acordadas o resoluciones que formalicen los protocolos éticos para el uso de la inteligencia artificial, estableciendo la obligatoriedad de que la justificación de cualquier decisión judicial asistida sea siempre humana, deliberada y explicable, estando este principio alineado con la jurisprudencia regional que prohíbe la sustitución de la racionalidad humana (UNESCO, 2024).

De igual manera, por el mismo ente rector, la creación de una unidad técnica especializada encargada de la curación de los *datasets* judiciales, con el mandato de garantizar la anonimización, la calidad de los datos y la realización de auditorías internas para mitigar los sesgos históricos.

Y no menos importante, gestionar la provisión de herramientas de inteligencia artificial asistencial de alta calidad para la defensa pública y otros operadores de justicia gratuita, contrarrestando activamente la desigualdad de recursos digitales generada por el sector privado (Allen, s.f.).

Por último, por el Consejo de la Magistratura y Escuela Judicial, la promoción de una alfabetización digital ética que integre la comprensión profunda de los riesgos de la inteligencia artificial, incluyendo el sesgo, la opacidad y la explicabilidad, como un módulo obligatorio y evaluable en la formación inicial y continua de jueces, fiscales y defensores públicos (UNESCO, 2024).

Y el financiamiento de proyectos de investigación que estudien la brecha digital y el sesgo algorítmico en el contexto sociodemográfico paraguayo, utilizando una metodología de diseño sensible al valor (*Value Sensitive Design*) para garantizar que el diseño tecnológico esté orientado a fortalecer los valores del Estado de derecho (LATAM DIGITAL, 2022).

Bibliografía

Allen, V. (s.f.). *La asimetría tecnológica en el litigio* [Monografía]. Poder Judicial/UNA.

Argentina.gob.ar. (2025). *El programa piloto de IA Generativa en la Justicia arrojó resultados positivos*.

AuroraLex. (s.f.). [Página oficial de la plataforma legal].

Banco Interamericano de Desarrollo (BID). (s.f.). *PROMETEA: Transformando la administración de justicia con herramientas de inteligencia artificial*.

Caballero Galeano, J. M. (2024). La Inteligencia Artificial en el Derecho: desafíos y oportunidades para el futuro Jurídico de Paraguay. *Revista Jurídica, 9*(1), 180-192.

Calvo de Mora Rivas, A., & Factor Santoveña, E. (2024). La integración de la Inteligencia Artificial (IA) en la abogacía y el sistema judicial: desafíos y oportunidades. *Revista Jurídica, 9*(1), 159-178.

Cumbre Judicial Iberoamericana. (2025). *Código Ético Iberoamericano para el Uso de la Inteligencia Artificial*.

Díaz, C., Franco, S., & Mendieta, M. (2022). *ESTADO DE DERECHO, SISTEMA JUDICIAL Y COVID-19: UNA VENTANA DE OPORTUNIDAD PARA EL SISTEMA JUDICIAL EN PARAGUAY*. Centro de Análisis y Difusión de la Economía Paraguaya (CADEP).

DPL News. (2025). *Paraguay: MITIC y CSJ trabajan en la digitalización del sistema judicial (Sistema Integral de Tramitación Judicial Electrónica - SITJE)*.

El Observador. (2025). *La Cámara avala el "poder de disciplina" de juez y ratifica sanción a abogados por litigar con mala fe*.

FPF. (2025). *AI Regulation in Latin America: Overview and Emerging Trends in Key Proposals*.

Gauna, G., Palumbo, P., & Baudino, C. (2021). Prometea: La inteligencia artificial al servicio de la Justicia. *Revista Magistrados, 20*(3), 50-61.

Judiciales.net. (s.f.). *Esto piensa una inteligencia artificial sobre la justicia paraguaya*.

LATAM DIGITAL. (2022). *El sesgo algorítmico y el enfoque del diseño sensible al valor*.

Mendieta, M., Franco, S., & Ciciolli, H. (2023). Brecha digital y personas mayores en Paraguay. *Revista Umbral*, (18).

Ministerio de la Defensa Pública. (2024). *Defensa Pública incluye herramientas tecnológicas para agilizar acceso a la justicia (Proyecto AuroraLex)*.

Ministerio de Tecnologías de la Información y Comunicación. (2025). *MITIC acompañó audiencia pública sobre proyecto que promueve el uso de la Inteligencia Artificial para el desarrollo del país*.

Ortiz Cristaldo, E. R. (2022). *Análisis del alcance de la facultad disciplinaria del juez en el ordenamiento jurídico paraguayo*. Universidad Tecnológica Intercontinental.

Poder Judicial. (s.f.). *Independencia del Poder Judicial*.

Poder Judicial. (2025a). *Jornada de capacitación sobre uso de IA en el servicio de justicia*.

Poder Judicial. (2025b). *Ministro Martínez Simón expuso avances del Poder Judicial en digitalización y calidad durante la Cumbre Iberoamericana*.

Revista Científica UNIDA. (2024). Inteligencia artificial y propiedad intelectual en Paraguay. *Revista Científica UNIDA, 2*(1).

Revistas.pj.gob.pe. (s.f.). *Inteligencia artificial, ChatGPT y jueces. Nuevos retos para la función jurisdiccional*.

Reyes Torres, M. (2025). *Magistrado Reyes Torres advierte mal uso IA en sistema judicial afecta derechos de las personas*. Tribunal Constitucional.

Sequera Buzarquis, M. (2021). Inteligencia Artificial en la sala constitucional en Paraguay. *TEDIC*.

TEDIC. (2024). *Protección de datos personales en el sector privado en Paraguay: Un estudio exploratorio*.

UPE. (s.f.). *¿Puede una inteligencia artificial ser considerada autora de una obra creativa?*.

Unidos en Red. (s.f.). *Uso estratégico de datos e inteligencia artificial en la justicia (Modelo Brasil)*.

UNESCO. (2024). *Paraguay capacita a jueces sobre el uso ético de la Inteligencia Artificial en el sistema judicial*.

Urueña, R. (2021). ¿Máquinas de justicia?: Inteligencia artificial y sistema judicial en América Latina. *Agenda Estado de Derecho*.

O mito de Midas e o devir criativo na era das plataformas digitais. Uma análise do relatório A/80/278 da ONU sobre inteligência artificial e criatividade

Por Marcílio Toscano Franca Filho [(*)] *y Jadgleison Rocha Alves* [(**)]

Resumo: Para os antigos gregos, os mitos não eram simples narrativas, mas programas de verdade, como lembra Paul Veyne (2014), com função pragmática e programática na vida cotidiana. O mito do rei Midas ilustra o paradoxo da abundância enganosa: ao transformar em ouro tudo o que tocava, o rei terminou condenado à fome e à perda dos vínculos humanos. No contexto contemporâneo, em especial na economia das plataformas digitais, essa metáfora assume novos contornos. A criatividade converte-se em metal precioso que alimenta a lógica lucrativa das empresas inseridas nas indústrias culturais e criativas, as quais, por meio da inteligência artificial (IA), redefinem os modos de produção, circulação e remuneração das obras culturais. A partir da análise da publicação do Relatório A/80/278 sobre Inteligência Artificial e Criatividade, elaborado em 2025 pela Relatora Especial das Nações Unidas em Direitos Culturais, Alexandra Xanthaki, o presente estudo examina as constatações e recomendações contidas no documento, a partir de três dimensões que se destacam: (i) os deslocamentos invertidos entre IA e criatividade (ii) o colonialismo de dados e a precarização do

---

(*) Árbitro da Organização Mundial da Propriedade Intelectual (OMPI, Genebra), da Corte de Arbitragem para a Arte (CAfA, Rotterdam) e do Centro Latinoamericano de Arbitragem e Mediação (LatCam, Assunção). Professor do Centro de Ciências Jurídicas da Universidade Federal da Paraíba e Procurador-Geral do Ministério Público de Contas do Estado da Paraíba, onde também chefia a Força-Tarefa do Patrimônio Cultural (FTPC). Membro permanente do Grupo de Trabalho da UNIDROIT sobre objetos culturais órfãos. Relator brasileiro para o tema Arte e Direito da Academia Internacional de Direito Comparado (AIDC, Haia). Professor visitante nas Universidades de Turim, Pisa e Ghent. Foi Calouste Gulbenkian Fellow no Instituto Universitário Europeu (EUI), árbitro suplente do Tribunal Permanente de Revisão do Mercosul (TPR), aluno (Gasthörer) da Universidade Livre de Berlim, estagiário-visitante do Tribunal de Justiça da União Europeia (Luxemburgo), Consultor Jurídico da Missão da ONU em Timor-Leste (UNOTIL) e Senior Legal Advisor do Ministério das Finanças de Timor Leste e do Banco Mundial. Membro da International Law Association (ILA). E-mail: mfilho@tce.pb.gov.br. E-mail: mfilho@tce.pb.gov.br

(**) Doutorando em Ciências Jurídicas na Universidade Federal da Paraíba (UFPB) e na *Università degli Studi Firenzi* (UNIFI); Pesquisador do Laboratório Internacional de Investigação da Transjuridicidade (LABIRINT/UFPB); Membro da European Society of International Law (ESIL). E-mail: jadgleison@gmail.com; jadgleison.rochaalves@unifi.it

trabalho criativo; e (iii) as insuficiências normativas no plano internacional. Conclui-se, reafirmando a tese do Relatório A/80/278, pela necessidade de reconhecer a centralidade explícita dos direitos culturais como núcleo da regulação internacional da IA – único antídoto à versão tecnológica da "maldição de Midas", que ameaça destinar a diversidade cultural a padecer de inanição.

Palavras-chave: Inteligência Artificial; direitos culturais; plataformas digitais; Direito Internacional dos Direitos Humanos; América Latina.

El mito de Midas y el devenir creativo en la era de las plataformas digitales. Un análisis del informe a/80/278 de la ONU sobre inteligencia artificial y creatividad

Resumen: Para los antiguos griegos, los mitos no eran simples narraciones sino programas de verdad, como recuerda Paul Veyne (2014), con una función pragmática y programática en la vida cotidiana. El mito del rey Midas ilustra la paradoja de la abundancia engañosa: al transformar en oro todo lo que tocaba, el rey terminó condenado al hambre y a la pérdida de los vínculos humanos. En el contexto contemporáneo, especialmente en la economía de las plataformas digitales, esta metáfora adquiere nuevos contornos. La creatividad se convierte en un metal precioso que alimenta la lógica lucrativa de las empresas insertas en las industrias culturales y creativas, las cuales, a través de la inteligencia artificial (IA), redefinen los modos de producción, circulación y remuneración de las obras culturales. A partir del análisis de la publicación del Informe A/80/278 sobre Inteligencia Artificial y Creatividad, elaborado en 2025 por la relatora especial de las Naciones Unidas en Derechos Culturales, Alexandra Xanthaki, el presente estudio examina las constataciones y recomendaciones contenidas en el documento, desde tres dimensiones que destacan: (i) los desplazamientos invertidos entre IA y creatividad; (ii) el colonialismo de datos y la precarización del trabajo creativo; y (iii) las insuficiencias normativas en el ámbito internacional. Se concluye, reafirmando la tesis del Informe A/80/278, en la necesidad de reconocer la centralidad explícita de los derechos culturales como núcleo de la regulación internacional de la IA —único antídoto frente a la versión tecnológica de la "maldición de Midas", que amenaza con condenar a la diversidad cultural a padecer inanición—.

Palabras clave: Inteligencia Artificial; derechos culturales; plataformas digitales; Derecho Internacional de los Derechos Humanos; América Latina.

The myth of Midas and the creative becoming in the age of digital platforms. An analysis of UN report a/80/278 on artificial intelligence and creativity

Abstract: For the ancient Greeks, myths were not mere narratives but programs of truth, as Paul Veyne (2014) reminds us, with a pragmatic function in everyday life. The myth of King Midas illustrates the paradox

of deceptive abundance: by turning everything he touched into gold, the king ultimately condemned himself to hunger and the loss of human connection. In the contemporary context particularly within the digital platform economy this metaphor acquires new meanings. Creativity becomes a precious metal fueling the profit-oriented logic of companies within the cultural and creative industries, which, through artificial intelligence (AI), are redefining the modes of production, circulation, and remuneration of cultural works. Based on the United Nations Special Rapporteur in the Field of Cultural Rights, Alexandra Xanthaki's 2025 Report A/80/278 on Artificial Intelligence and Creativity, this study examines its key findings and recommendations across three interrelated dimensions: (i) the inverted relationship between AI and creativity; (ii) data colonialism and the resulting precarization of creative labor; and (iii) the regulatory shortcomings of the international framework. The study concludes, reaffirming the central thesis of Report A/80/278, that it is necessary to recognize the explicit centrality of cultural rights as the normative core of international AI regulation – the only antidote to the technological version of the "Midas curse," which threatens to destine cultural diversity to suffer from starvation.

Keywords: Artificial Intelligence; cultural rights; digital platforms; International Human Rights Law; Latin America.

1. Introdução

Conforme narrado pela mitologia clássica, Midas, lendário rei da Frígia, recebeu do deus do vinho Dioniso, um dom em retribuição à hospitalidade com que acolheu o velho sátiro Sileno, companheiro de peregrinações do deus. Movido pela ambição, solicitou que tudo o que tocasse fosse transformado em ouro. Sem demora, logo percebeu o caráter trágico do dom recebido: os alimentos, a água e até o vinho convertiam-se em metal inerte, privando-o de alimento e prazer. Dominado pela fome e pelo desespero, Midas suplicou a Dioniso que o libertasse do fardo concedido. O deus, compadecido, ordenou-lhe que se purificasse nas águas do rio, que absorveram o ouro devolvendo sua humanidade perdida. A narrativa, assim, simboliza a maldição de uma riqueza ilusória que, ao negar o essencial da vida, condena o ser humano a um destino terrível. (Maraschin, 2014). [(1)] Para os antigos gregos, como lembra Paul Veyne (2014), os mitos não eram simples narrativas, mas autênticos programas de verdade, dotados de função pragmática na vida cotidiana.

No cenário contemporâneo, marcado pela emergência da inteligência artificial (IA) como elemento estruturante da economia das plataformas

(1) Ver também OVIDIO. *Metamorfosis*. XI, 85-193. Tradução de Ana P. 3. Ed. Sevilla, 2002-2008. Texto bilíngue.

digitais, a metáfora do Mito de Midas é então mobilizada como chave interpretativa para ilustrar novos contornos do paradoxo da abundância enganosa no âmbito das Indústrias Culturais e Criativas (ICCs), [2] onde a criatividade é deslocada e transformada em "metal precioso" através do fascínio pela eficiência algorítmica e a promessa de produtividade ilimitada que ao redefinirem os modos de produção, circulação e remuneração das obras culturais, obscurecem as ameaças estruturais, gerando implicações mais profundas para os direitos humanos, em especial os direitos culturais.

A pandemia de COVID-19 intensificou essa inflexão histórica. Ao mesmo tempo em que revelou a vulnerabilidade humana e o papel essencial da criatividade como forma de resiliência, consolidou um mercado digital massivo de conteúdos culturais mediado por plataformas digitais e intensificado pelo uso de IA. Se por um lado a digitalização expandiu o acesso em tempos de isolamento, por outro, acentuou igualmente uma lógica [3] estruturada e orientada pelo lucro, trazendo riscos significativos para efetivação plena dos direitos culturais.

A partir de 2021, diversas agências das Nações Unidas reconheceram a urgência de compreender e alertar os impactos dessa aceleração tecnológica [4] –caracterizada por um crescimento exponencial no período pós-pandêmico– e seus reflexos sobre populações e áreas diversas. No âmbito do sistema internacional de direitos humanos, emergiram novas agendas normativas voltadas à necessidade de estabelecer marcos de governança ética e jurídica do ambiente digital, com ênfase na proteção de dados e na privacidade. Em paralelo, os procedimentos especiais do Conselho de Direitos Humanos passaram a analisar a interface entre IA e direitos humanos fundamentais, produzindo uma série de relatórios temáticos sobre liberdade de opinião e expressão, privacidade, combate ao terrorismo, enfrentamento da pobreza extrema, políticas migratórias, bem como os direitos das pessoas idosas e pessoas com deficiência. [5]

---

(2) O desenvolvimento de serviços digitais baseados em novos modelos de negócios como resposta ao desafio digital impactou consideralmente setores das Indústrias Culturais e Criativas (ICCs) como a música e o audiovisual. Ver Octavio Kulesz, *La cultura en la era digital: Evaluar el impacto en América latina y en España*. Paris: UNESCO, 2017.

(3) A Nota Informativa da OIT sobre COVID-19 e o setor cultural (2020) aprofunda essa leitura ao mostrar como a pandemia acelerou a transição digital, ampliando a dependência de plataformas de *streaming* e evidenciando vulnerabilidades estruturais. Durante o confinamento, setores inteiros –música, audiovisual, espetáculos ao vivo– sofreram um colapso de receitas: só a indústria musical perdeu 10 bilhões de dólares em patrocínios. (OIT, 2020, p. 2). O relatório documenta perdas de emprego massivas: no Reino Unido, 75,9% dos trabalhadores das artes foram demitidos; nas Filipinas, a ocupação caiu 55% em abril de 2020; no Brasil, as receitas audiovisuais caíram 14,8% (OIT, 2020, p. 2).

(4) Estados e organizações internacionais vêm produzindo resoluções e recomendações, como a Recomendação da UNESCO sobre a Ética da Inteligencia Artificial (2021), que se consolidou como referência normativa global.

(5) Ver os Relatórios A/73/348, A/76/380, A/HRC/50/32, A/HRC/53/65, A/HRC/36/48, A/HRC/45/14, A/HRC/49/52, A/HRC/52/39, A/HRC/42/59, A/HRC/44/57, A/75/590, A/HRC/48/76 e A/HRC/47/52.

É nesse contexto que se insere o Relatório A/80/278 sobre Inteligência Artificial e Criatividade (2025), apresentado pela Relatora Especial das Nações Unidas para os Direitos Culturais, Alexandra Xanthaki [6]. Fruto de amplas consultas globais envolvendo Estados, organizações intergovernamentais e sociedade civil [7], o documento constitui um marco analítico essencial ao deslocar a discussão sobre inovação tecnológica para o eixo dos direitos culturais universais. O Relatório adverte que a IA, longe de ampliar os horizontes criativos, frequentemente os restringe, reduzindo a produção cultural a combinações probabilísticas que esvaziam a autenticidade da experiência artística (Xanthaki, 2025).

Segundo o Relatório em grande medida, têm sido amplamente negligenciadas as respercussões da IA sobre a criatividade posto que "como não há tempo para reflexões sobre essas questões, se impõe o caráter 'inevitável' dos avanços tecnológicos, junto com a insistência de abraçar a IA no lugar de se opor" (Xanthaki, 2025, p. 4). Entre os riscos identificados no relatório destacam-se o colonialismo de dados, o treinamento de sistemas de IA com dados culturais dominantes que reforçam estéticas hegemônicas, a reprodução de estereótipos e conteúdos discriminatórios por meio de algoritmos enviesados.

A importância do Relatório reside, portanto, em evidenciar que as repercussões da IA sobre a criatividade e os direitos culturais, têm sido amplamente negligenciadas, com ausência de reflexão crítica e ação regulatória, permitindo que a IA se desenvolva exclusivamente sob lógicas corporativas de concentração de poder, ameaçando a diversidade cultural, a liberdade artística e o próprio futuro do trabalho criativo.

2. Criatividade como humanidade: os perigos da redução auto-matizada do ser criativo advertidos no relatório

> "– Não. Eu não sei por que escrevo. A gente escreve, como quem ama, ninguém sabe por que ama, a gente não sabe por que escreve. Escrever é um ato solitário, solitário de um modo diferente de solidão. Escrevo com amor e atenção e ternura e dor e pesquisa, e queria de volta, como mínimo, uma atenção e interesse".
>
> Clarice Lispector

A reflexão sobre a artificialidade acompanha a história do pensamento ocidental como uma tentativa de compreender os limites da criação humana e os riscos de suas próprias invenções. Desde Aristóteles, que distinguiu

(6) Alexandra Xanthaki é professora de Direito na Brunel University, Londres, foi nomeada Relatora Especial da ONU na área de direitos culturais em outubro de 2021. Ela é grega e concluiu seu curso de Direito na Faculdade de Direito de Atenas.

(7) As contribuições podem ser acessadas em www.ohchr.org/en/calls-for-input/2025/call-contributions-artificial-intelligence-and-creativity.

o natural —aquilo que possui dentro de si mesmo um princípio de movimento— do artificial —dependente da ação externa do homem [8]—, até os diagnósticos críticos contemporâneos de Byung-Chul Han, que denuncia a colonização algorítmica da vida [9], o pensamento filosófico tem assinalado que a questão da artificialidade não é apenas uma mera técnica, mas também uma simbologia para questões sociais e políticas.

É justamente nesse horizonte que se insere a primeira advertência do Relatório A/80/278 sobre Inteligência Artificial e Criatividade (2025), apresentado pela Relatora especial das Nações Unidas para os Direitos Culturais, Alexandra Xanthaki.

O Relatório A/80/278 inaugura sua análise com a sistematização das principais concepções elaboradas em diversos documentos internacionais sobre inteligência artificial (IA), definido como um conjunto de sistemas "baseados em máquinas que, para objetivos explícitos ou implícitos, deduzem, da informação que recebe, como gerar resultados, como previsões, conteúdos, recomendações ou decisões que possam influenciar os entornos físicos ou virtuais" (Xanthaki, 2025, p.5). A partir dessa definição, se evidencia o duplo eixo sobre o qual se estrutura a IA: sua autonomia e adaptabilidade, que quando se relacionam com a criação humana, provocam deslocamentos de natureza conceitual com consequências jurídicas para o gozo dos direitos culturais relacionados à criatividade. [10]

É dessa forma que o Relatório A/80/278 enfatiza a necessidade de diferenciar de forma "clara e explícita", segundo Xantthaki (2025, p.6), o conteúdo ou o material "em cuja criação participa um ser humano e o que é puramente sintético", reconhecendo que este último "não é criativo, senão apenas uma adição automática, incluída [...]" (Xanthaki, 2025, p. 6). A própria Relatora recorda que a IA é também conhecida pelo nome de "Internet média", precisamente porque não é "criativa em si", limitando-se apenas a recombinar o já existente (Xanthaki, 2025, p. 6).

Alguns dos fundamentos da criatividade são desenvolvidos em seguida, asseverando Xanthaki (2025), que a criatividade não se restringe a um mero processo criativo ou apenas a um conceito alusivo a criação artística, mas antes de tudo que:

> "É parte da vida diária, e consiste em combinar o uso da imaginação, as experiências anteriores e a lógica para conceber novas ideias ou criar elementos novos. [...] inclui o eco emocional, a percepção, o talento, a narração

(8) ARISTÓTELES. *Física* II, 192b 13-16.

(9) HAN, Byung-Chul. Infocracia: digitalização e a crise da democracia. 2022.

(10) Em particular o direito de toda pessoa de participar da vida cultural e o direito de se beneficiar da proteção dos interesses morais e materiais resultantes de suas criações científicas, literárias e artísticas, previsto no artigo 15 do Pacto Internacional sobre Direitos Econômicos, Sociais e Culturais (PIDESC).

> e a possibilidade de romper padrões e jogar com a incerteza, destrezas que podem aplicar-se em todas as esferas da vida e em todos os tipos de atividade [...] é o que dá sentido a todas as atividades humanas". (Xanthaki, 2025, p. 6).

Criar, nesse contexto, significa articular experiências prévias, percepções sensíveis que são inseparáveis da memória e da razão humana na concepção do novo. Sob a ótica dos direitos culturais, essa formulação reforça que a criatividade é um elemento indispensável da experiência humana, portadora de densidade subjetiva, narrativa simbólica e juízo ético, e como sublinhado pela Relatora, tais elementos que "as máquinas não podem reproduzir, por mais avançados que sejam os algoritmos" (Xanthaki, 2025, p. 6).

A possibilidade de uma inversão conceitual —que atribua à máquina o protagonismo criativo— coloca em risco a própria integridade da expressão cultural. Ainda que a IA possa ser utilizada como instrumento auxiliar, não deve ser confundida como atividade criativa autêntica, pois lhe faltam consciência narrativa e vínculo comunitário. Em vez de expandir horizontes criativos, os sistemas de IA frequentemente reduzem a experiência humana a meros cálculos probabilísticos, capazes apenas de "inferir o que seguirá a continuação segundo o que seja mais provável (o mais comum)" (Xanthaki, 2025, p.6).

Dessa forma a criação artística, científica ou cultural é intrinsecamente relacional e histórica, construída na interação entre indivíduo, memória coletiva e contexto social. Como observa a Relatora:

> "As atividades de escrever um poema, compor uma peça musical ou preparar um prato não giram em torno de produtos apenas, mas sim aos processos que envolvem parte do próprio passado, das próprias ideias e das próprias preferências da família. Esses processos dizem ao mundo algo específico dos seres humanos como nunca poderão um livro ou melodia gerados por IA". (Xanthaki, 2025, p. 6-7).

A história da arte oferece exemplos eloquentes dessa irredutabilidade que a criação autêntica através das subjetividades —pessoal e coletiva— subverteram cânones literários, instaurando e reinventando novas linguagens e dimensões impossíveis de serem reproduzidas por algoritmos. Vincent van Gogh (11), Frida Kahlo (12), Cardosinho (13), Heitor Villa-Lobos (14),

---

(11) Em sua obra "*Noite Estrelada"* (1889), sublimou sua angústia.

(12) Na Obra "A Coluna Partida" (1944), projetou sua dor física e íntima como testemunho de resistência e identidade.

(13) Como era conhecido o pintor brasileiro José Bernardo Cardoso Júnior, que retratou experiências populares invisibilizadas nos salões de arte europeus na virada do século XIX para o XX. Exemplo de como a arte emerge da vivência concreta, da memória coletiva e da subjetividade do olhar humano.

(14) Através da Obra "*Bachianas Brasileiras nº 5"* (1938-1945), possibilitou a união da tradição erudita e popular para responder às tensões de identidade cultural da época.

Violeta Parra [15], Cândido Portinari [16], Jorge Luís Borges [17] e Manuel de Barros [18], entre tantos outros, demonstraram a essência criativa de transgredir padrões, assumindo o risco inerente da incerteza que reside o impulso para o inédito e o diverso, valores caros ao corpus normativo dos Direitos Culturais Universais.

Diferentemente da IA, que apenas recombina dados preexistentes, a criação humana envolve corpo, memória, subjetividade e ética. É nesse vínculo intransponível entre experiência e criação que reside a dinâmica insubstituível da criatividade humana diante do avanço da IA.

O problema central, portanto, não reside apenas na utilização da IA enquanto ferramenta [19], mas na tendência de delegar processos criativos fundamentais às máquinas, como alerta Xanthaki (2025, p. 7), o uso generalizado de ferramentas de IA "equivale a uma delegação do pensamento" que pode gerar desresponsabilização e atrofia das habilidades críticas, ameaçando a continuidade das práticas culturais e artísticas. Nesse sentido, o uso acrítico de IA ameaça a liberdade artística e o direito cultural de participar na vida cultural, já que a criatividade humana é reduzida a outputs automatizados.

Assim, a primeira advertência apresentada no Relatório A/80/278 revela que a artificialidade, quando confundida com criatividade, produz um deslocamento ontológico e axiológico que ameaça a própria integridade da criatividade humana. Ao converter o gesto criador em mera operação algorítmica, essas manobras tecnológicas promovem uma inversão de fundamentos, na qual a técnica deixa de servir à criatividade para torná-la servil aos parâmetros de cálculo e previsibilidade, substituindo toda uma dimensão subjetiva, processual e coletiva protegida pelos direitos culturais. Tal processo de instrumentalização das expressões culturais reduz a criação a outputs padronizados, desprovidos de intencionalidade, afetividade e historicidade —atributos que constituem o cerne do ato criativo enquanto ma-

---

(15) Por meio da Obra "*Gracias a la vida"* (1966), converteu uma despedida marcada pela dor em canção universal de gratidão.

(16) Ao pintar "*Guerra e Paz"* (1952–1956), transfigurou os horrores da guerra em denúncia estética contra a violência e a injustiça.

(17) Ao investigar em sua obra "Labirintos" (1962), os paradoxos do tempo, da autoria, do inconsciente e da memória.

(18) Com sua poesia que subverteu hierarquias linguísticas, valorizando a simplicidade e a estética do "desaprendizado" como forma de revelação criativa. Ver Manoel de Barros. *Livro das ignorãças*. Rio de Janeiro: Record. 1993.

(19) O próprio ato de utilização de uma máquina como ferramenta para elaboração desse estudo exemplifica esse paradoxo contemporâneo, em que a ferramenta tecnológica participa do processo, mas sem jamais substituí-lo. Essa relação é igualmente visível no campo das imagens, ao utilizarmos máquinas para a produção de obras fotográficas onde estas foram consideradas, por muito tempo, como uma forma de "não" arte ou arte secundária, exemplificando essa tensão no direito italiano, que só elevou a fotografia à categoria de obras em 1979, quando foi introduzida uma distinção entre obras fotográficas, consideradas produto da engenhosidade humana, e meras fotografias. (ICOM Itália, 2025, p. 26).

nifestação da liberdade humana—. A Relatora reafirma que a criatividade, enquanto expressão da dignidade e da pluralidade humanas, é dimensão inalienável dos direitos culturais, e sua proteção representa condição essencial para preservar a humanidade criativa frente às tentações homogeneizadoras da inteligência artificial. (Xanthaki, 2025).

É precisamente nesse ponto que se abre o debate sobre o colonialismo de dados, fenômeno que, ao instrumentalizar a informação e os repertórios culturais como matéria-prima algorítmica, instaura dependências sutis, mas profundas, entre centros de produção tecnológica e periferias culturais. A próxima seção, examina essas dinâmicas, revelando como a promessa de neutralidade da IA encobre um projeto de poder que ameaça a diversidade e a autodeterminação cultural dos povos.

3. Criatividade para além da produtividade: o mundo algoritxmico das quantidades e velocidades como um banho ácido para a vida criativa

Se a primeira advertência do Relatório A/80/278 refere-se ao risco de erosão da criatividade humana e da diversidade cultural por conta de deslocamentos de fundamentos da relação com a Inteligência Artificial (IA), a segunda advertência, não menos importante, refere-se, dentre outras temáticas, ao chamado "colonialismo de dados", a questão do agravamento da "brecha criativa", da invisibilidade de criadores e das condições do espaço de trabalho reservado a criatividade no ambiente da lógica de exploração algorítmica. (20)

Conforme destaca Xanthaki (2025), a "nova forma de colonialismo" levado a cabo pelas "poderosas empresas privadas", se configura pela extração "não apenas de recursos naturais, mas de recursos em forma de dados" (Xanthaki, 2025, p. 8). Trata-se de uma forma sofisticada de colonialismo cultural, que substitui a dominação territorial por mecanismos de dependência tecnológica. A concentração de poder no Norte Global, observa a Relatora, faz com que as "cosmovisões e interesses econômicos de um

(20) O relatório global da Organização Internacional do Trabalho (OIT) sobre mudanças tecnológicas e o futuro do trabalho amplia de forma significativa o diagnóstico apresentado pela Relatora Especial da ONU em Direitos Culturais. O documento sublinha que a automação e a inteligência artificial (IA), quando orientadas unicamente por critérios de eficiência econômica, "reforçam desigualdades estruturais e concentram poder em poucas plataformas digitais" (OIT, 2023a). As ocupações mais vulneráveis à substituição por máquinas são justamente aquelas marcadas pela informalidade, baixos salários e ausência de proteção social -características amplamente presentes nas ICCs. Em consequência, a precariedade histórica dos artistas torna-os terreno fértil para a exploração digital através de ferramentas de IA. Além disso, a OIT alerta que os algoritmos tendem a privilegiar conteúdos dominantes e comercialmente rentáveis, marginalizando expressões locais, indígenas e minoritárias. Sem regulação específica, esse processo ameaça a diversidade cultural e a liberdade de expressão artística – pilares indispensáveis da democracia e do trabalho decente. (OIT, 2023a).

pequeno número de atores determinem o desenho dos sistemas" reforçando discriminações e desigualdades estruturais existentes (Xanthaki, 2025, p. 8).

A Relatora Especial chama atenção para o fenômemo que denomina de "brecha criativa" [21], onde "quem detém o poder informático, a infraestrutura de dados e o domínio das línguas hegemônicas determina, de maneira desproporcionada, os produtos e a estética da IA generativa" (Xanthaki, 2025, p. 13), ocasionado a falta de acesso por uma grande parte de criadores —principalmente os localizados no Sul Global— aos espaços, conhecimentos e benefícios tecnolóicos ligados ao fazer criativo, amplificando as assimetrias e violações já existentes no ambiente "analógico" da criatividade. [22]

Os sistemas de inteligência artificial são alimentados por vastos bancos de dados, usualmente construídos de forma unilateral, sem consentimento informado nem transparência [23]. Esses "inventários de dados", que incluem material cultural, dificilmente são elaborados de modo participativo, multidisciplinar ou guiados por marcos de direitos humanos (Xanthaki, 2025, p. 8). As comunidades, lembra a Relatora, "nunca haviam imaginado que suas vidas se converteriam em dados utilizados de maneira tão ilimitada e opaca" (p. 8). O resultado é a consolidação de um modelo cultural que privilegia conteúdos dominantes e homogeneizadores, reduzindo a diversidade estética e simbólica.

A lógica algorítmica reforça normas culturais dominantes, promovendo "apenas as formas criativas que vendem, restringindo a diversidade cultural e fomentando a homogeneização" (Xanthaki, 2025, p. 11). Esse processo, conforme destacado pela contribuição do Brasil ao Relatório, impacta especialmente criadores de baixa renda e os do Sul Global, que sofrem "de maneira desproporcionada os efeitos negativos da disrupção provocada pela IA nas economias culturais", dada a ausência de marcos legais protetivos (Xanthaki, 2025, p. 12).

---

(21) Em comparação ao conceito de "brecha digital" que se refere ao "acesso material a dispositivos digitais e conectividade digital, conhecimentos e habilidades necessárias para utilização desses dispositivos e o benefício que se obtem das tecnologias digitais" (Xanthaki, p. 12).

(22) A OIT (2020) observa esse contexto de crise revelou também a desigualdade na apropriação tecnológica: grandes empresas sobreviveram e até expandiram suas margens via plataformas de *streaming*, enquanto pequenos espaços culturais, teatros e companhias independentes fecharam ou se endividaram. (OIT, 2020, p. 4).

(23) Justamente por conta disso, o jornal brasileiro Folha de São Paulo ingressou com uma ação judicial contra a OpenAI, criadora do ChatGPT, acusando a empresa de usar, sem autorização, o conteúdo jornalístico do veículo para treinar seus modelos de inteligência artificial. O jornal pede indenização e interrupção imediata do uso de seu material, além da destruição dos modelos treinados com seus dados. Conforme a petição inicial dessa ação, não se trata de um caso isolado. Veículos de mídia de renome global estão reagindo: o New York Times processou a OpenAI e a Microsoft para impedir o uso não autorizado de seus conteúdos, e assinou contrato (para remuneração pelo uso de seus conteúdos) com outra empresa de IA, a Amazon (assim como o fizeram também a Conde Nast e a Hearst); a BBC e a News Corp estão acionando judicialmente a empresa de IA Perplexity, que também é processada pelo maior periódico japonês, o Yomiuri-shimbun.

Além da concentração econômica, emerge o fenômeno da autopromoção algorítmica. Plataformas que também produzem conteúdo, como Netflix, tendem a privilegiar suas próprias obras, reduzindo a visibilidade de produções independentes e locais (Xanthaki, 2025, p. 16). Esse mecanismo cria um círculo vicioso de exclusão: artistas periféricos e grupos minoritários permanecem invisíveis, enquanto algoritmos calibram recomendações para maximizar engajamento e lucro. Tais práticas equivalem a uma censura indireta, contradizendo o direito à liberdade artística e à diversidade cultural.

O uso da IA para o desenvolvimento algorítmico acentua as deficiências dos algorítmicos em relação com a criatividade. O desenvolvimento algorítmico permite seletivamente que alguns criadores culturais tenham exposição ao público e que outros não. É possível que se mostre ao público menos diversidade de obras. [...] refletem o interesse econômico estratégico em aumentar ao máximo a atividade do usuário e consolidar a dominação do mercado. Por exemplo, uma plataforma de transmissão em direto (streaming) que também produz conteúdos em linha (como Netflix) tenderá a ressaltar as obras de produção própria, o que se conhece como "autopromoção algorítmica" ou "favoritismo algorítmico". (p. 16).

Gilles Lipovetsky (2014, p. 468) já advertia que "não faltam as análises que mostraram de que maneira o mundo virtual das grandes velocidades representava um banho ácido para a vida estética". A advertência se concretiza na era das plataformas digitais quando a aceleração das quantidades e das métricas corrói o tempo de reflexão, de elaboração, de experiências e de maturação criativa. A produtividade torna-se valor supremo, e a arte —reduzida a "metal precioso"— passa a existir em função da atenção instantânea e da rentabilidade algorítmica.

Além disso, as plataformas e multinacionais tecnológicas passaram a investir maciçamente em conteúdos gerados por IA, substituindo produções humanas por versões automatizadas mais baratas e escaláveis (Xanthaki, 2025, p. 17). Isso ameaça a sustentabilidade econômica dos artistas, reduz fontes de renda, [24] intensifica a precarização e promove contratos desiguais. A Relatora alerta que muitos criadores recebem "pagamentos únicos" por usos ilimitados de sua voz, imagem ou dados biométricos, re-

(24) Estudo citado pela Relatora Especial –elaborado pela Confederação Internacional de Sociedades de Autores e Compositores (CISAC)– estima que 24% das receitas de criadores de música poderão ser substituídas por produtos de IA até 2028 (Xanthaki, 2025, p.17). Esse dado revela que a IA não apenas compete com os artistas, mas pode "canibalizar" setores inteiros das ICCs, transferindo o valor simbólico e econômico da criatividade humana para algoritmos controlados por empresas globais.

nunciando, na prática, ao controle e à remuneração futura de seus próprios atributos pessoais (Xanthaki, 2025, p. 18). [25]

O Relatório da OIT sobre o futuro do trabalho nas artes e no entretenimento complementa esse quadro, observando que a digitalização e a IA ampliam a informalidade, a ausência de proteção social e a desigualdade de gênero no setor cultural (OIT, 2023b). A conjunção entre colonialismo de dados, homogeneização estética e precarização laboral constitui, assim, um sistema integrado de exploração do trabalho criativo. [26]

A Relatora Especial para os Direitos Culturais enfatiza que "a criatividade é mais que produtividade: exige tempo, pensamento e intercâmbio" (Xanthaki, 2025, p. 6). Ao sobrepor a criatividade da produtividade, a Relatora tem como objetivo proteger e resguardar o espaço de trabalho destinado à criatividade, como o "oposto ao imperativo de produzir mais e mais rápido [...] o que implicaria no mundo atual uma inteira dependência das máquinas para escrever e ler em uma quantidade assustadora de obras de duvidosa qualidade. A criatividade exige tempo, pensamento e intercâmbio". (Xanthaki, 2025, p. 6). A criatividade, é essencialmente profética e contestadora – algo que a IA jamais poderá emular.

Assim, o avanço da inteligência artificial, quando orientado por métricas quantitativas e lógicas de mercado, compromete o direito à liberdade artística, a diversidade cultural e as condições dignas do trabalho criativo. O desafio contemporâneo não é rejeitar a IA, mas submeter sua governança a parâmetros críticos baseados nos direitos culturais, assegurando que a tecnologia sirva à vida e não à sua mercantilização.

---

(25) O relatório da Organização Mundial da Propriedade Intelectual (OMPI, 2021) sobre práticas contratuais no setor audiovisual latino-americano evidencia a persistência de assimetrias estruturais nas relações entre produtores independentes e grandes plataformas de distribuição. O estudo aponta contratos que exigem a cessão de "todos os direitos disponíveis" –cinema, TV, SVOD, TVOD e AVOD– em troca de pagamentos fixos limitados, restringindo a autonomia futura dos produtores e bloqueando a exploração de múltiplas janelas de exibição. (OMPI, 2021, p. 7). Outro problema recorrente é a ausência de *mínimos garantidos* (MG): diferentemente dos mercados europeu e norte-americano, na América Latina predominam contratos de *revenue sharing* ou taxas fixas baixas, expondo os produtores a riscos desproporcionais. Em muitos casos, as receitas não cobrem sequer custos técnicos de legendagem e dublagem. Um exemplo emblemático é o de uma produtora argentina que, entre 2019 e 2020, produziu quatro episódios de uma série por conta própria e buscou financiamento posterior. A plataforma que financiou os episódios restantes impôs licença exclusiva sobre todos os oito episódios em toda a América Latina, limitando a participação do produtor a 50% das vendas secundárias fora da região. (OMPI, 2021, p. 10). O relatório ainda destaca a atuação crescente de agregadores de vídeo sob demanda (VOD), como *Sofa Digital* (Brasil) e *Promovere* (Argentina), que, embora ampliem o alcance das obras, cobram taxas elevadas de *onboarding* técnico e percentuais sobre receitas, criando novas barreiras de entrada para pequenos produtores. (OMPI, 2021, p. 18).

(26) A OIT (2023b) apresenta um diagnóstico abrangente sobre o futuro do trabalho nas artes e no entretenimento, destacando que, embora o setor possua enorme potencial de coesão social e dinamização econômica, enfrenta níveis alarmantes de informalidade, fragilidade estrutural e déficits de trabalho decente. Tais fragilidades foram agravadas pela digitalização e pelo avanço da inteligência artificial. (OIT, 2023b).

4. Insuficiências normativas no cenário internacional e as in-flexões do Sul Global: práticas emergentes

Embora diversas organizações internacionais tenham produzido recomendações e estratégias sobre o uso da inteligência artificial (IA), a Relatora Especial das Nações Unidas para os Direitos Culturais identifica que os marcos existentes permanecem insuficientes e juridicamente não vinculantes. Conforme observa Xanthaki (2025), "a maioria desses instrumentos não tem eficácia vinculante e, embora se refiram aos direitos e liberdades fundamentais como base para estabelecer objetivos, princípios ou compromissos, não fazem menção explícita dos direitos culturais" (p. 18). Documentos recentes, como a Recomendação da UNESCO sobre a Ética da Inteligência Artificial (2021) [27] e diversas outras estratégias nacionais, enfatizam princípios éticos revisados, mas omitem frequentemente referências explícitas aos direitos culturais e à criatividade humana como bens jurídicos a serem protegidos (Xanthaki, 2025, p. 18).

A ausência de salvaguardas concretas e mecanismos vinculantes permite que plataformas digitais utilizem milhões de dados e obras culturais para o treinamento de modelos algorítmicos sem autorização, compensação ou mecanismos de supervisão transparentes. Como adverte a Relatora, "a transparência e o direito de autor não estão incluídos nas iniciativas legislativas recentes, como o Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito [28]; ou, quando estão, como no Regulamento de Inteligência Artificial da União europeia, faltam garantias de implementação" (Xanthaki, 2025, pp. 18-19).

Paradoxalmente, essa insuficiência normativa abre janelas de oportunidade para que países do Sul Global avancem na formulação de novos arranjos regulatórios vinculantes, construídos a partir de sua condição de exportadores de bens e serviços culturais e, simultaneamente, consumidores de conteúdos de plataformas estrangeiras. Nesse contexto, além do desafio de integrar os direitos culturais de forma explícita, torna-se igualmente estratégico assegurar a aplicação efetiva do princípio do tratamento preferencial, consagrado no artigo 16 da Convenção da UNESCO de 2005 sobre a Proteção e Promoção da Diversidade das Expressões Culturais, como instrumento jurídico de proteção da diversidade cultural e da criatividade numa perspectiva voltada ao fortalecimento no ambiente digital das Indústrias Culturais e Criativas (ICCs) dos países em desenvolvimento. Essa perspectiva pode consolidar-se como eixo central de novos

(27) Disponível em: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

(28) Disponível em: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence.

acordos de cooperação internacional, em especial, no Acordo Mercosul-União Europeia (UE) concluído recentemente, superando o desafio de modernização e/ou adaptação regulatória, principalmente em áreas ainda não reguladas —ou reguladas de forma deficitárias— nas regiões contempladas pelo acordo.

O Acordo Mercosul-UE surge como instrumento estratégico não apenas para o bloco sul-americano, mas em matéria de Indústrias Culturais e Criativas (ICCs), oportuniza tanto para os países do bloco europeu, como para a comunidade internacional no que diz respeito a uma atualização dos marcos jurídicos regionais e internacionais aplicados à diversidade cultural e à criatividade, promovendo a introdução de novas disciplinas como os direitos culturais.

Esse olhar já pode ser lançado a partir de alguns mecanismos e práticas já existentes no âmbito do Mercosul, como o espaço de governança da cultura voltada ao setor audiovisual através da criação do RECAM (Reunião Especializada das Autoridades Cinematográficas e Audiovisuais do Mercosul) criado em 2003 [29], que estabelece um espaço de governança cooperativa entre Argentina, Brasil, Paraguai, Uruguai e Bolívia. Desde então, o RECAM tem consolidado praticas regionais voltadas à circulação de obras, à formação de profissionais e à harmonização regulatória no setor audiovisual.

Entre suas principais realizações destacam-se: (i) a Plataforma MERCOSUL Audiovisual, que promove coproduções e difusão de conteúdos regionais; (ii) a criação do Registro Único do Audiovisual do Mercosul (RUAM), que visa simplificar procedimentos jurídicos e fiscal entre os países do bloco; (iii) a adoção de acordos de reciprocidade em certificação de obras e coprodução, antecipando práticas hoje aplicáveis às plataformas digitais; (iv) o desenvolvimento de linhas regionais de financiamento e digitalização, apoiadas pela UNESCO e pela OEI, que servem de modelo para políticas de transição digital centradas em valores culturais. (MERCOSUL/RECAM, 2025).

A trajetória acumulada pelo RECAM, no interior do Mercosul, traz esse possível caminho epistemológico do Sul Global, onde, descentralizado de culturas dominantes, seja capaz de inspirar novos marcos regulatórios decoloniais na agenda birregional do Acordo Mercosul-UE, principalmente, quanto a importância da centralidade dos direitos culturais na política regulatória dos sistemas de IA na ordem internacional.

---

(29) Disponível em: https://www.recam.org/?set_lang=pt.

5. Conclusão: recomendações para uma política de regulação internacional de IA centralizada nos direitos culturais

Em seu Relatório, Xanthaki (2025), por meio do trabalho do Órgão Assessor de Alto Nível sobre Inteligência Artificial, assinala que o futuro da IA dependerá menos do avanço tecnológico e mais da resposta ética e política da humanidade. A questão não é a velocidade da inovação, mas a capacidade de orientar seu sentido social e cultural.

Quando olhamos para atrás num período de cinco anos, o panorama tecnológico poderia ser radicalmente distinto do atual. No entanto, se mantemos o rumo e superamos as vacilações e as dúvidas, dentro de cinco anos podemos contemplar um panorama de governança da IA inclusivo e empoderador para as pessoas, as comunidades e os países do todo o mundo. Em última instância, o que importa não é a mudança tecnológica, mas sim, senão como responde a ela a humanidade (ONU, 2024).

A análise desenvolvida neste estudo permite afirmar que a inteligência artificial (IA), se regulada apenas sob critérios técnicos ou éticos genéricos, representa um risco estrutural à criatividade humana, à diversidade cultural e ao trabalho decente no setor cultural. As evidências recolhidas convergem para uma constatação inequívoca: a centralidade dos direitos culturais deve orientar a governança global da IA.

O Relatório A/80/278 da Relatora Especial da ONU para os Direitos Culturais (Xanthaki, 2025) destaca que a regulação não pode restringir-se a parâmetros de segurança ou eficiência, mas deve estar ancorada em uma agenda internacional de direitos culturais e trabalhistas, capaz de proteger a criatividade e redistribuir benefícios econômicos.

A IA não substitui a criação humana: a cultura envolve subjetividade, memória e juízo ético —dimensões que nenhum algoritmo pode replicar—. Em outras palavras, embora a inteligência artificial generativa represente uma forma inédita de originalidade técnica (combinatória e exploratória), isso não se confunde com criatividade autêntica. Embora seja capaz de recombinar padrões e gerar produções novas e impressionantes —textos, imagens, melodias, códigos—, sua atuação permanece desprovida de consciência, intenção, emoção e sentido simbólico. O humano cria para expressar sentido, sentimento e sensações; a IA cria porque foi programada para combinar probabilidades.

A coleta massiva de dados e o colonialismo informacional concentram poder em poucas plataformas, reforçando desigualdades históricas. No campo laboral, a expansão da IA acentua precarização, contratos desiguais e dependência de intermediários digitais, especialmente no setor audiovi-

sual latino-americano. Sem regulação, a IA consolida um mercado desigual em que artistas e produtores independentes assumem riscos, enquanto corporações capturam os ganhos [30].

Diante disso, a regulação da IA deve ter caráter normativo robusto, fundado no Direito Internacional dos Direitos Humanos. O futuro da criatividade e do trabalho digno depende da construção de normas que coloquem a dignidade humana no centro da inovação tecnológica. A IA deve ser compreendida como desafio regulatório civilizatório, não como ameaça inevitável ou promessa redentora.

Com base nas evidências e instrumentos internacionais, destacam-se algumas recomendações convergentes: (a) Reconhecimento normativo explicitamente os direitos culturais nos tratados e estratégias globais de IA; (b) Proteção do trabalho criativo e visibilidade para os criadores; (c) Formação, integração e intervenção de equipes multidisciplinares e culturalmente diversa em direitos culturais no treinamento dos sistemas de IA; (d) respeitar o direito a descobrir expressões criativas diversas, inclusive as expressões de grupos subrepresentados dando acesso a essas expressões;

A IA será legítima no campo cultural apenas se subordinada aos direitos culturais universais. Como adverte Xanthaki (2025, p. 70), é urgente reconhecer "as múltiplas formas pelas quais a IA pode erosionar irreversivelmente a criatividade humana". O vazio regulatório atual favorece a concentração de poder e a precarização, mas também oferece uma oportunidade para o Sul Global, sobretudo a América Latina, propor modelos de regulação epistemologicamente inspirados na diversidade cultural de países do Sul.

REFERÊNCIAS

Aristóteles. (2002). "Física I–II" (J. Barnes, Org.; L. Angioni, Trad.). Campinas, SP: Editora da Unicamp. (Obra original séc. IV a.C.).

Barros, M. de. (1993). "Livro das ignorãças". Rio de Janeiro: Record.

Borges, J. L. (1962). "Labirintos" (H. M. de Campos, A. C. de Campos & D. A. Filho, Trads.). São Paulo: Companhia das Letras. (Obra original publicada em 1960 como Labyrinths: Selected Stories & Other Writings.).

Clarice Lispector apud Borrelli, O. (1981). "Clarice Lispector: Esboço para um possível retrato". Rio de Janeiro: Nova Fronteira.

(30) O Relatório Racine (2020), do Ministério da Cultural da França, já alertava para a "uberização" da economia criativa. A proliferação de conteúdos online, aliada à dificuldade de garantir o respeito ao direito autoral, ameaça a sustentabilidade das profissões criativas e exige novas formas de regulação pública para equilibrar inovação tecnológica e proteção cultural.

Council of Europe. (2024). "Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law". Strasbourg: Council of Europe. https://www.coe.int/en/web/artificial-intelligence/convention.

Empresa Folha da Manhã S/A v. OpenAI INC, OpenAI Global LLC e OpenAI GP LLC (2025). Petição inicial no processo nº 1107237-96.2025.8.26.0100-1, 3ª Vara Empresarial e de Conflitos de Arbitragem da Comarca de São Paulo, Tribunal de Justiça de São Paulo, São Paulo, SP.

European Union. (2024). "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts". Official Journal of the European Union, L 202, 12.7.2024, 1–156. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689

Han, B.-C. (2022). "Infocracia: Digitalização e a crise da democracia". Petrópolis: Editora Vozes.

ICOM Italia. (2025). "FAQ: Intelligenza artificiale e diritti culturali umani". Roma: ICOM Italia. https://www.icom-italia.org/pubblicazioni/

International Labour Organization. (2020). "La pandemia de la COVID-19 y el sector de los medios de comunicación y de la cultura: Nota informativa sectorial". Geneva: ILO. https://www.ilo.org/es/resource/brief/la-pandemia-de-la-covid-19-y-el-sector-de-los-medios-de-comunicaci%C3%B3n-y-de

International Labour Organization. (2023a). "Technological changes and the future of work". Geneva: ILO. https://www.ilo.org/meetings-and-events/technology-and-future-work

International Labour Organization. (2023b). "Meeting report: The future of work in the arts and entertainment sector". Geneva: ILO. https://www.ilo.org/publications/future-work-arts-and-entertainment-sector

Lipovetsky, G. (2014). "O capitalismo estético na era da globalização". Lisboa: Edições 70.

Maduro, L., Veiga, P. da M., & Rios, S. P. (2020). "Acordo Mercosul–União Europeia: Impactos normativos e regulatórios no Mercosul". Banco Interamericano de Desenvolvimento (BID) / Instituto para a Integração da América Latina e do Caribe (INTAL). https://publications.iadb.org/handle/11319/10275

Maraschin, L. T. (2014). "O esplêndido e mísero rei: Reflexões sobre as metamorfoses de Midas". Fênix – Revista de História e Estudos Culturais, 11(2), Universidade Federal de Santa Maria. https://www.revistafenix.pro.br/revistafenix/article/view/602

MERCOSUL/RECAM. (2025, 24 de junho). "Acta n.º 01/25: Reunión Especializada de Autoridades Cinematográficas y Audiovisuales del MERCOSUR".

Buenos Aires: MERCOSUR/RECAM. https://www.recam.org/?do=downloads&idCategory=34c8e803cc87e5e1da56241fa5059609

Racine, B., avec le concours de N. Corbin, C. Roux, & B. Saint-Etienne. (2020). L'auteur et l'acte de création: Rapport au ministre de la Culture. Ministère de la Culture. França.

Santa Sé. (2023). Antiqua et Nova: Notas sobre a inteligência artificial para uma comunicação plena e verdadeiramente humana. Dicastério para a Comunicação, Cidade do Vaticano.

UNESCO. (2005). "Convenção sobre a Proteção e Promoção das Expressões da Diversidade Cultural". Paris: UNESCO. https://www.unesco.org/en/legal-affairs/convention-protection-and-promotion-diversity-cultural-expressions

UNESCO. (2021). "Recomendación sobre la Ética de la Inteligencia Artificial". Paris: UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000376713_spa

Vercelli, A. (2024). "United Nations, artificial intelligences and regulations: Analysis of the General Assembly AI resolutions and the final report of the Advisory Body on AI". CEUR Workshop Proceedings, 3881. https://ceur-ws.org/Vol-3881/paper11.pdf

Veyne, P. (2014). "Os gregos acreditavam em seus mitos? Ensaio sobre a imaginação constituinte". Tradução Mariana Echalar . 1 ed. São Paulo: Editora Unesp.

Xanthaki, A. (2025). "Informe de la Relatora Especial sobre los derechos culturales: Inteligencia artificial y creatividad (A/80/278)". Naciones Unidas, Asamblea General. https://docs.un.org/es/A/80/278

World Intellectual Property Organization (WIPO). (2021). "Study on the audiovisual legal framework in Latin America: Part 4 – Contractual practices in the Latin American audiovisual sector in the digital environment". Geneva: WIPO. https://dacatalogue.wipo.int/projects/DA_1_3_4_10_11_16_25_35_01

Inteligencia artificial y persecución penal en Chile: análisis y desafíos del proyecto de ley del subsistema de inteligencia económica (boletín N.º 15.975-25) [*]

Por Francisco Narváez Gallo [**] *y David Vallespín Pérez* [***]

Resumen: El presente trabajo analiza el Boletín N.° 15.975-25. La iniciativa busca integrar la labor de diversas agencias estatales, como la Unidad de Análisis Financiero (UAF), el Servicio de Impuestos Internos (SII) y el Servicio Nacional de Aduanas (SNA), entre otras, fortaleciendo sus capacidades de inteligencia. Durante el debate legislativo se presentó la idea de utilizar inteligencia artificial (IA) para el análisis masivo de datos y la detección de operaciones sospechosas. A partir de esta premisa, advertimos los desafíos jurídicos que esto implica, con relevancia en la tensión entre la eficiencia persecutoria y el respeto a los derechos fundamentales. Se abordan riesgos críticos como los sesgos algorítmicos, la opacidad de los sistemas de IA (*black box*) y la necesidad de establecer responsabilidades. Subrayamos la necesidad de un marco regulatorio robusto, que garantice la transparencia y el debido proceso. A partir de este análisis, se presentan algunas inquietudes sobre el producto del informe de inteligencia generados por IA: su tratamiento en juicio, sugiriendo su configuración como una prueba pericial especializada y no vinculante, sometida a la libre valoración judicial.

Palabras claves: Inteligencia artificial; inteligencia económica; crimen organizado; persecución penal; derechos fundamentales.

(*) Este artículo fue realizado en el marco de la beca doctoral otorgada por la Fundación Carolina, España, período 2025-2026.

(**) Magister en Derecho, mención Derecho Procesal; doctorando en Derecho y Cs. Políticas de la Universitat de Barcelona, España; académico asistente de la Facultad de Derecho, Universidad de Santiago de Chile. Código Postal 9170020, Santiago, Chile; francisco.narvaez@usach.cl, ORCID https://orcid.org/0009-0008-0125-2337.

(***) Doctor en Derecho; catedrático de Derecho Procesal y coordinador del Máster de Abogacía y Procura de la Universitat de Barcelona. España. Código Postal 08034, Barcelona, España; dvallespin@ub.edu, ORCID https://orcid.org/0000-0002-5435-0482.

Artificial intelligence and criminal prosecution in Chile: analysis and challenges of the economic intelligence subsystem bill (bulletin N.° 15.975-25)

Abstract: This paper analyses Bulletin No. 15.975-25. The initiative seeks to integrate the work of various state agencies, such as the Financial Analysis Unit (UAF), the Internal Revenue Service (SII) and the National Customs Service (SNA), strengthening their intelligence capabilities. During the legislative debate, the idea of using Artificial Intelligence (AI) for mass data analysis and the detection of suspicious transactions was presented. Based on this premise, we note the legal challenges that this implies, with relevance to the tension between prosecutorial efficiency and respect for fundamental rights. Critical risks such as algorithmic biases, the opacity of AI systems (black-box) and the need to establish responsibilities are addressed. We emphasise the need for a robust regulatory framework that guarantees transparency and due process. Based on this analysis, we present some concerns about the product of AI-generated intelligence reports: their treatment in court, suggesting their configuration as specialised and non-binding expert evidence, subject to free judicial assessment.

Keywords: Artificial intelligence; economic intelligence; organised crime; criminal prosecution; fundamental rights.

1. INTRODUCCIÓN

El estreno de la película *Minority Report* (traducida a *Sentencia previa* en Hispanoamérica) en el 2002 presenta *in extremis* una de las funciones del derecho penal contemporáneo: la preventiva. Si bien la posibilidad dada por la mencionada obra aún está lejos de concretarse, no podemos desconocer que los actuales avances y desarrollos tecnológicos en inteligencia artificial (IA) ya están abriendo el camino para que esta sea empleada en orden a prevenir la perpetración de delitos. ¿Y si el escenario no fuera tan irreal, para aquellos delitos de asociación y organización? Lo relevante de este planteamiento radica en que ya no hablamos de instrumentos jurídicos propios de la prevención general y de la legitimidad del castigo estatal (advertencias, citaciones judiciales, medidas alternativas a la pena, sentencias ejemplificadoras, entre otros), sino del empleo de instrumentos *extrajurídicos* para la predicción del comportamiento probable de un individuo o grupo. Esto se ha entendido como parte de una prevención especial de concepción gerencial de evaluación y prevención de los riesgos, la que en el último tiempo busca exacerbar el empleo de la IA (Silva Sánchez, 2025, p. 287) [(1)].

(1) El abanico de posibilidades que presenta la IA en la prevención general y en la legitimidad de la actividad punitiva estatal, nos lleva a plantearos que la distopía se dirige hacia la catalogación de comportamientos sospechosos, en especial, aquellos que atenten contra la seguridad interior.

Por su parte, la doctrina ha descrito la necesidad de hacer frente a los delitos de mayor peligrosidad para el Estado. Así, se ha señalado que el derecho penal del enemigo es un derecho penal de tercera velocidad, en el que ciertas figuras o tipos amenazan con socavar los fundamentos esenciales de una sociedad organizada en Estado. Opera sobre un ámbito donde su existencia debe ser reducida a la mínima expresión: un derecho penal aplicable a imputables peligrosos (Silva Sánchez, 2001, pp. 163–164). Este derecho penal de la excepción, con un contenido dogmático sistemático, impacta la política criminal al pretender anticipar la punibilidad sin reducir penas y al debilitar garantías procesales. Este conjunto de cuerpos normativos, enfocados en la amenaza y no en el acto consumado, es lo que doctrina denomina "leyes de lucha" para combatir amenazas específicas (Kindhäuser, 2011, pp. 172–174).

Por las consideraciones planteadas, no debiese extrañarnos que, para penalistas, procesalistas, psicólogos, criminólogos, y tantos otros operadores del sistema penal, uno de sus mayores intereses de investigación y análisis consista en idear y debatir acerca de la prevención de la ocurrencia de eventos delictuales. Siguiendo con ese tópico, el debate se ha centrado en el comportamiento humano y en la fiabilidad metodológica (Romeo Casabona, 2020, p. 185). Estas inquietudes no deben ser desatendidas, ya que han abierto paso a un proceso de reconfiguración del derecho penal; desde su tradicional modelo *ex post* a un enfoque *ex ante*, que pretende reaccionar ante riesgos y amenazas (Barona Vilar, 2025, p. 168).

El presente artículo tiene por objeto dar a conocer a la comunidad científica y académica el estado de tramitación de un proyecto de ley en Chile denominado N.° 15.975-25 que "crea el Subsistema de Inteligencia Económica y establece otras medidas para la prevención y alerta de actividades que digan relación con el crimen organizado" (Boletín). Durante el debate legislativo se ha presentado la idea del empleo de IA, para el tratamiento de operaciones bancarias y financieras sospechosas que sean conducentes para actividades terroristas y de criminalidad organizada.

Nuestro principal objetivo será determinar cuáles debieran ser los elementos mínimos que debiere incorporar la regulación del empleo de IA en orden a la denominada "inteligencia económica". A tal efecto, no solo procederemos a describir el proyecto de ley de referencia y los problemas que le acompañan, sino también a proponer algunos lineamientos y puntos críticos acerca de los caminos que debiera seguir su regulación, en un contexto en el que resulta patente el complejo y tenso equilibrio entre garantías y eficiencia. Para la finalidad de este trabajo, no abordaremos la prevención desde la *psiquis* o fase interna de la autoría, sino más bien desde la óptica de las oportunidades (no exentas de algunos riesgos) que presenta el empleo de IA en la trazabilidad de activos.

2. EL PROYECTO DE LEY DE CHILE (15.975-25) POR EL QUE SE CREA EL SUBSISTEMA DE INTELIGENCIA ECONÓMICA Y SE ESTABLECEN MEDIDAS DE PREVENCIÓN Y ALERTA ANTE ACTIVIDADES RELACIONADAS CON EL CRIMEN ORGANIZADO: PLANTEAMIENTO GENERAL

Con fecha 31 de mayo de 2023, el Poder Ejecutivo ingresó este proyecto de ley, iniciando así la tramitación del Boletín en el Congreso Nacional. El proyecto propone una serie de medidas de prevención y alerta para actividades vinculadas con al terrorismo y el crimen organizado. Esta iniciativa es parte de una agenda legislativa que se está implementando desde el 2022, con orientación y prioridad en seguridad y que, por extensión, persigue ajustar la legislación penal y procesal penal al fenómeno del crimen organizado en Chile.

Concretamente, el Boletín se enmarca al Plan Nacional de Seguridad Pública y Prevención del Delito para el período 2022-2026. A su vez, dicho plan hace expresa mención a la elaboración de la Política Nacional contra el Crimen Organizado. La aprobación de este Boletín es parte de las líneas de acción que buscan mejorar la persecución del dinero proveniente de lavado de activos, financiamiento del terrorismo y otros delitos vinculados al crimen organizado. También se pretenden crear e implementar mecanismos, herramientas y unidades especializadas orientados en el análisis de datos, la trazabilidad de operaciones, el levantamiento del secreto bancario, la ampliación de rubros económicos obligados a informar y la institucionalidad.

Este Boletín consta de 25 artículos. El primero de ellos crea el Subsistema de Inteligencia Económica (el subsistema). Todas las demás propuestas inciden directamente en los organismos con competencia bancaria, financiera, económica, tributaria y comercio: la Unidad de Análisis Financiero (UAF), el Servicio de Impuestos Internos (SII), el Servicio Nacional de Aduanas (SNA), la Comisión para el Mercado Financiero (CMF), la Tesorería General de la República (TGR) y Superintendencia de Casinos (SdC). El proyecto se articula sobre la base de tres grandes ejes: el fortalecimiento de los ecosistemas de inteligencia económica, la prevención y detección temprana de operaciones económicas sospechosas, y la toma en consideración de facultades intrusivas y sancionadoras.

El primero de estos ejes, guarda relación con el fortalecimiento del ecosistema de inteligencia económica. En esta dirección, se creará un Subsistema de Inteligencia Económica, integrando tres instituciones ya existentes: la UAF, el SII y el SNA. Integración que se materializará mediante las respectivas Unidades de Inteligencia que el Boletín creará para cada uno de esos servicios. El subsistema y sus integrantes funcionalmente coordinados, dirigirán y ejecutarán labores de inteligencia consistentes en la recolección, evaluación y análisis de información sobre actividades relacionadas con ciertos delitos. El fin de esta integración coordinada es realizar alertas y/o colaborar con el Ministerio Público y el Sistema de Inteligencia del Estado, según corresponda.

De igual modo, los integrantes del Subsistema podrán intercambiar información y requerir antecedentes entre sí y a otros organismos públicos; y para el caso de que dicha información sea secreta o reservadat esta deberá ser compartida con los organismos integrantes del subsistema. Se amplían también las competencias de la UAF, las que, actualmente, consiste en prevenir, detectar y requerir información, con miras a evitar el lavado de activos y la financiación del terrorismo en Chile; lo cierto es que el Boletín también persigue que puedan levantarse alertas asociadas, con agregación del fenómeno de la delincuencia organizada. En esta línea de acción, se habilitará a la UAF para que acceda a la información protegida por el secreto o reserva bancaria legal sin previa autorización judicial, cuando dichos antecedentes sean necesarios y conducentes a desarrollar o complementar el análisis de una operación sospechosa reportada o detectada.

El segundo de los ejes referidos, referido a la prevención y detección temprana de operaciones económicas sospechosas, se hace descansar en la actualización de la normativa aplicable a los servicios públicos que integran el Subsistema de Inteligencia Económica y los demás organismos que ejercen supervisión en materia económica para la prevención y alerta temprana.

Y el tercer eje, relativo a las facultades intrusivas y sancionatorias, se centra en perfeccionar la legislación aplicable respecto de tales facultades, desde la perspectiva de los órganos que cumplen funciones de supervisión y fiscalización económica y financiera. En este punto, el Proyecto pretende, en concreto, flexibilizar las formalidades legales para la procedencia de algunas diligencias, al punto de omitir el control judicial.

Por lo que se refiere a las iniciativas presentadas cabe decir, en primer lugar, por lo que respecta al secreto bancario, que se pretende otorgar a la CMF la facultad administrativa de acceder a información resguardada por secreto y reserva bancaria, sin la intermediación de un tribunal que lo autorice. Esta medida se justifica para dar consistencia a la CMF con las facultades generales para acceder a la información de sus supervisados bancos. Para dar cuenta de esta flexibilización intrusiva, el Boletín pretende eliminar el control judicial que la persona afectada pueda realizar ante un ministro de la Corte de Apelaciones (desarrollo en una única audiencia no solemne).

En segundo lugar, el Boletín propone modificar los requisitos de procedencia para las diligencias intrusivas con las que cuenta la CMF. De hecho, la falta de formalidad y de cumplimiento oportuno de dichas medidas, en el marco de una investigación o procedimiento sancionatorio, no afectará a que dichos antecedentes puedan ser utilizados en una investigación penal. Y tampoco podrán ser objeto de debate de exclusión de prueba durante la audiencia preparatoria de juicio oral en lo penal.

Por último, también resulta obligado señalar que en orden a las facultades de levantamiento de secreto y reserva bancaria para la UAF y la CMF, se apuesta por la potenciación y de la eventual aplicación de sistemas de inteligencia artificial, en particular por lo que concierne al análisis masivo de datos.

Expuesto lo anterior, con su ingreso de 31 de mayo de 2023, ante el Congreso Nacional; el Boletín inició su Primer Trámite Constitucional ante el Senado. En dicha fase, y en el uso de sus atribuciones constitucionales, el Senado puede solicitar consultas a otros órganos del Estado, para conocer de su parecer sobre el Boletín. La Corte Suprema ya expresó que "cualquier esfuerzo por mejorar la efectividad y la eficiencia de esta clase de tarea conlleva la ponderación de ciertos intereses legítimos que pueden encontrarse en tensión", al referirse a los riesgos de la integridad de instituciones relevantes, como son el secreto y la reserva bancaria. Destacamos, de los comentarios presentados por el máximo Tribunal, aquel que tiene que ver con el correcto balance normativo para dicha facultad, ya que entiende que solo es procedente para el ámbito de competencias propias de la CMF: "[E]l supuesto legal que da lugar a esta potestad es altamente restrictivo: ellaosolo se refiere al análisis financiero acotado al mandato legal que rige ambas instituciones, y no incluye una permisión general para realizar diligencias investigativas intrusivas sin autorización judicia." (*sic*) [(2)].

En otro orden de cosas, con fecha 18 de diciembre de 2023, la Comisión de Seguridad Pública del Senado evacuó su Primer Informe respecto al Boletín. En esta instancia, el otrora ministro de Hacienda, Mario Marcell Cullell, señaló que es crucial que las unidades de inteligencia apliquen métodos de IA, para procesar volúmenes grandes de información tributaria e identificar movimientos anómalos. Citó dos situaciones que suelen observarse: 1) un contribuyente declara pérdidas por mucho tiempo, o 2) un contribuyente formaliza su negocio en Chile, hace una gran transacción y luego desaparece.

Posteriormente, con fecha 17 de enero de 2025, evacuó su informe la Comisión de Hacienda del Senado. En esa oportunidad, el senador Felipe Kast Sommerhoff destacó y celebró que la UAF cuente con facultades acotadas para proceder al levantamiento del secreto bancario, sin previa autorización judicial, haciendo expresa referencia a la indicación 2H b) [(3)]. Para el representante, esto no hace más que dar mayor agilidad al seguimiento de

---

(2) Chile, Corte Suprema. Oficio N° 158- 2023. *Informe de proyecto de ley que "crea el subsistema de inteligencia económica y establece otras medidas para la prevención y alerta de actividades que digan relación con el crimen organizado"*. 5 de junio de 2023.

(3) En síntesis, señala la indicación que la UAF tendrá la facultad de levantar el secreto bancario para analizar operaciones sospechosas vinculadas al crimen organizado. De forma excepcional, también podrá requerir esta información directamente a los bancos, sin autorización judicial previa, cuando el análisis involucra a personas jurídicas, funcionarios públicos o reportes ya iniciados por una entidad bancaria. Este poder es ejercido por el director de la UAF mediante una resolución secreta y fundada, la cual es sometida a control judicial *ex post* y aleatoria por la Corte de Apelaciones. La información obtenida es estrictamente confidencial y su divulgación se sanciona con la destitución.

la ruta del dinero. Pero además de eso y sin que sea objeto del Mensaje– el senador entiende que esto abre la posibilidad a que la UAF cuente con sistemas informáticos sofisticados, que permitan a dicha entidad tratar datos bancarios, económicos y financieros con mayor agilidad.

Así las cosas, no cabe duda de que la IA abre el camino, por ejemplo, para reducir los tiempos en el tratamiento de datos o, incluso, para detectar patrones sospechosos. Este trámite constitucional concluyó el 25 de marzo 2025, aprobándose parcialmente el Boletín, que se derivó a la Cámara de Diputados y Diputadas, iniciando su segundo trámite constitucional.

2.1. Implicaciones jurídicas del uso de la IA en el Subsistema. ¿Qué podemos esperar?

Dada la iniciativa esbozada, en la tramitación del Boletín, entramos ahora en el espacio de la especulación para tratar, de una parte, de determinar cuál pueda ser la utilización de la IA en la detección preventiva de operaciones sospechosas, bancarias y financieras; y de otra, sugerir ciertas indicaciones normativas y regulatorias, que debiesen ser consideradas en la tramitación del proyecto.

Al respecto, se ha señalado que ciertos sistemas de IA emplean una metodología de base actuarial, convirtiendo datos en algoritmos, para realizar la evaluación de estos, su procesamiento y el tratamiento por procedimientos de IA (Romeo Casabona, 2020, pp. 185–187). Instrumentos que, en verdad, constituyen herramientas para la asistencia de los operadores del sistema penal (instrumentalidad), en la toma de decisiones relevantes. En complemento de lo anterior, conviene recordar que cada sistema o programa policial de inteligencia predictiva converge en un punto crucial: el análisis en base a la utilización de *big data* (Souza de Menezes & Agustina, 2020, pp. 157–158).

Una IA que resulta tan fascinante como abierta a notables desafíos éticos y legales. No debemos olvidar que, a ese instrumento, se le deben proporcionar fuentes de información y datos, para que cumpla con sus fines. Pero ¿Qué ocurre si la fuente ya está viciada por los sesgos del autor de la fuente? La experiencia ha demostrado que, en materia de delincuencia, las fuentes de información suelen provenir de la observación de un denunciante, un agente policial o testigos, quienes pueden aportar datos incompletos, sospechosos, manipulables y plegados de prejuicios (Miró Llinares, 2020, p. 9). De ahí, precisamente, que apostemos por su uso, con sentido común, es decir, toda vez que se haya valorado su fiabilidad y ajuste con el obligado respeto a los derechos fundamentales, (en particular por lo que se refiere a la calidad de los datos utilizados, la forma en que han sido recopilados y los sesgos que puedan afectarles en el caso concreto) (Miró Llinares, 2020, p. 12).

Por lo que respecta al Boletín objeto de análisis, su empleo principal iría en la detección de patrones u operaciones bancarias y financieras sospechosas, destinadas a la financiación de actos terroristas o propios de la criminalidad organizada. Ahora bien, dado que la UAF podría levantar el secreto bancario sin control judicial *ex ante*, empleando IA. El riesgo evidente es que el *machine learning* permita que la IA perpetue y amplifique los sesgos discriminatorios al nutrirse de datos históricos. Estas herramientas, al basarse en el pasado, refuerzan prejuicios existentes en lugar de superarlos, creando un sesgo de confirmación algorítmico (Nieva-Fenoll, 2022, pp. 432–433).

2.2. Algunos casos que nos invitan a la reflexión

Cabe considerar, en primer lugar, el uso y empleo de la IA en labores de inteligencia e investigación penal, con carácter preventivo. El mundo virtual de internet y de las redes han dado paso al *derecho al entorno virtual*, que hace de su protección la intimidad sobre la información personal protegido internacionalmente, contra el Estado (Molina *et al.*, 2020, p. 1041) [4] . Este derecho, de nueva generación, ha sido reconocido en múltiples sentencias del Tribunal Constitucional (TC) y Tribunal Supremo español (TS), entre las que destacan las SSTS 823/2015 [5], 1385/2018 [6] y la STC 5928-2009 [7], todas con énfasis en determinar las líneas de actuación policial principalmente y sin perder la oportunidad de otorgarle contenido.

---

(4) Para el interés del ordenamiento jurídico chileno, el siguiente paso es incorporar esta construcción, resolviendo las siguientes preguntas ¿Cómo se conecta este concepto con los derechos garantizados en la Constitución Política de Chile, como el derecho a la vida privada y, en especial, a la protección y tratamiento de los datos personales (art. 19 N.° 4) y a la inviolabilidad de las comunicaciones privadas (art. 19 N.° 5)?

(5) Tribunal Supremo (Sala de lo Penal), sentencia de 24 de febrero de 2015, Roj 823/2015 - ECLI:ES:TS:2015:823: "(…) desde la perspectiva del derecho de exclusión del propio entorno virtual, como de las garantías constitucionales exigidas para el sacrificio de los derechos a la inviolabilidad de las comunicaciones y a la intimidad, la intervención de un ordenador para acceder a su contenido exige un acto jurisdiccional habilitante. (…)".

(6) Tribunal Supremo (Sala de lo Penal), sentencia de 11 de abril de 2018, Roj 1385/2018 - ECLI:ES:TS:2018:1385. El fallo diferencia entre *ciberpatrullaje* y una operación propia de agente encubierto. En la primera, la actividad policiaca es válida si se realizan exploraciones o indagaciones en redes sociales o canales de comunicación abiertos. Mientras que, en la segunda hipótesis, y tras la reforma a la Ley de Enjuiciamiento Criminal del 2015, si la actividad investigativa central es la intromisión en el perfil del sujeto –como lo fue con Facebook en el caso– requiere de resolución judicial que habilite la actuación.

(7) Tribunal Constitucional, sentencia de 07 de noviembre de 2011, Recurso de amparo 5928-2009. Un condenado por distribución de pornografía infantil acciona de amparo. Fue descubierto por tras haber llevado su computador portátil a un técnico encargado de repararlo, quien al advertir del material realiza la denuncia a la policía y entrega el equipo. El Tribunal desestima el intento de amparo por dos motivos: uno, el actuar del técnico no es una lesión al derecho a la intimidad, por cuanto contaba con el consentimiento del condenado para encender, acceder y explorar el equipo, sin clave de ingreso, para realizar su reparación; y dos, la pericia e intervención policiaca se justifica ya que, a pesar de no contar con autorización judicial, la injerencia supera el examen de proporcionalidad en dos dimensiones, por un lado, a la urgente necesidad de investigar y esclarecer un delito de distribución de pornografía infantil, y por otro, en evitar la destrucción de las pruebas borrando los ficheros ilícitos.

Relevante es, sin duda, el posicionamiento del TC español dado en el fallo citado, ya que asume un papel fundamental en orden a la delimitación del contenido que es propio de este derecho: proteger el conjunto de información personal almacenada en dispositivos y generada a través de la actividad en internet. Este cúmulo de datos (documentos, navegación, correos, redes sociales) no es una suma de elementos aislados, sino que configura un perfil detallado de la personalidad que revela aspectos de la esfera más íntima del individuo (ideología, salud, aficiones). Por tanto, este derecho busca garantizar la intimidad y el secreto de las comunicaciones frente a la intromisión de terceros o de los poderes públicos en este espacio digital personal.

Recientemente, en STS 902/2024 [8] se reconoce que los derechos de nueva generación –entre ellos al entorno digital– conforman una encrucijada susceptible de matices y modulaciones, en que las legislaciones nacionales presentan divergencias. Estas propician mayor permeabilidad transfronteriza. Este punto fue sostenido por el TS, para desestimar el motivo de casación intentado. En definitiva, entendió que "no existe un derecho a que el encausado pueda desvelar el contenido y alcance de las colaboraciones policiales internacionales". A su vez, "(...) el derecho de las partes personadas a conocer las pruebas materiales que estén en posesión de las autoridades competentes hace referencia al material que integra el procedimiento penal seguido ante los tribunales, sin otra exclusión que la que, de manera temporal, deriva del secreto de las actuacione." [9].

En segundo lugar, parece oportuno determinar la clase de fuentes de información (públicas o privadas). Tal como va la discusión actual sobre el Boletín, el Subsistema se nutrirá de múltiples agencias, que se alimentará de los servicios públicos que la integrarán y coordinarán. En este sentido, debemos pensar en: los *Reportes de Operación Sospechosa* (ROS) de la UAF, en las declaraciones, auditorias y reportes tributarios del SII, en las declaraciones y cargos aduaneros del SNA, en las denuncias, fiscalizaciones y medidas de control sobre instituciones bancarias, financieras y societarias bajos supervisión de la CMF, en las partidas y acciones de cobro de TGR, o en los reportes y denuncias de actividades ligadas a casinos que emanen de la SdC.

La respuesta directa parece ser, por tanto, que la información es pública, ya que emana de instituciones y agencias públicas. Sin embargo, convendría no obviar que la información con que alimentarían al subsistema es privada, y más aún, de carácter reservado y sensible, conforme a la Ley de Transpa-

(8) Tribunal Supremo (Sala de lo Penal), sentencia de 28 de octubre de 2024, Roj 3662/2024 - ECLI:ES:TS:2024:3662.

(9) Fundamento de Derecho Uno, puntos 3 y 7.

rencia [10] y a la Ley de Protección de Datos Personales [11]. Y este aspecto es del todo relevante para la validez de la actividad de vigilancia preventiva, que, no lo olvidemos, es uno de los fines del subsistema.

Por último, cabe preguntarse ¿Qué resguardos normativos debieren tomarse para evitar afectaciones a los derechos y garantías fundamentales amenazados, en especial, al derecho al entorno digital? Responder a esta cuestión no es sencillo, por cuanto los límites o contornos del ciberespacio, con una fuente abierta, cerrada o de datos públicos, va a depender, en buena medida, del medio tecnológico utilizado, así como de la propia casuística. Sobre este particular, resulta especialmente gráfica la STS 1932/2008, que distingue entre fuente abierta y cerrada [12].

Por consiguiente, el marco teórico descrito tributa a la multidimensionalidad del crimen organizado, y la necesidad de crear instituciones, metodologías y herramientas de inteligencia también multidimensionales, que permitan la integración e intercambio de información entre los diferentes agentes del Estado (Rivera V. & Sansó-Rubert Pascual, 2021, pp. 177–178).

De otra parte, es obligado tomar en consideración que, durante la discusión en la primera sesión del Comité de Seguridad Pública, se advirtió que solo dos servicios han implementado IA para el tratamiento de datos e informes bancarios, financieros y comerciales: el SII y el SNA; así como que se está tramitando, en paralelo, el Boletín N.° 12.234-03, que busca fortalecer y modernizar el actual Sistema de Inteligencia del Estado.

---

(10) Si bien la Administración pública debe mantener a disposición permanente del público ciertos antecedentes, la ley establece un límite. De acuerdo con su artículo 7°, no se incluirán en estos antecedentes los datos sensibles, esto es, los datos personales que se refieren a las características físicas o morales de las personas o a hechos o circunstancias de su vida privada o intimidad, tales como los hábitos personales, el origen social, las ideologías y opiniones políticas, las creencias o convicciones religiosas, los estados de salud físicos o psíquicos y la vida sexual.

(11) Si bien la ley 21.719 entrará en vigor a partir del 1 de diciembre del 2026, su artículo 25 ya nos presenta un panorama que debiese ser considerado en la tramitación del Boletín en cuestión. En líneas generales, el tratamiento de datos personales sobre infracciones y sanciones penales, civiles o administrativas se restringe a organismos públicos, exclusivamente para el cumplimiento de sus funciones legales y en los casos previstos por la ley. Se exige que la información sea exacta y su publicidad se limita temporalmente, estableciendo plazos específicos o un período general de cinco años. Una vez prescrita la acción o cumplida la sanción, los datos no deben ser públicos. Se prohíbe explícitamente el tratamiento masivo de estos datos, considerándose su incumplimiento una infracción gravísima, salvaguardando así los derechos individuales frente a una vigilancia generalizada.

(12) Tribunal Supremo (Sala de lo Penal), sentencia de 09 de mayo de 2018, Roj 1932/2008 - ECLI:ES:TS:2008:1932. Se señala que una fuente abierta o datos públicos, es aquella información accesible por cualquier usuario en internet sin necesidad de autorización judicial. Se considera pública porque el propio usuario la introduce en la red, dejando un rastro o "huella" (como la dirección IP que accede a un contenido) que sabe que queda registrada. Por el contrario, por fuente cerrada, cabe referir la información que vincula un dato público (como una IP) con la identidad real de una persona (el titular del contrato, su teléfono, etc.). Estos datos son privados, están en posesión de terceros (como los operadores de internet) y su acceso requiere una autorización judicial preceptiva para proteger el derecho a la intimidad.

En ese escenario, como es fácil intuir, habrá que extremar las cautelas, en aras a evitar incongruencias normativas que puedan afectar a los objetivos del Subsistema y los flujos de información hacia el sistema de persecución penal y el Sistema de Inteligencia del Estado; la interoperabilidad en el intercambio de información entre organismos del subsistema, y entre servicios relacionados con el Ministerio de Hacienda; o los deberes de los funcionarios que integran el subsistema, así como las sanciones con las que debieran castigarse los incumplimientos que cometan.

2.3. La preocupación por la opacidad de la IA ante el proceso

Igualmente, dado que el futuro subsistema permitirá elaborar informes y reportes de inteligencia económica, se detecta también un problema acerca de su naturaleza procesal y la forma de presentación en juicio. En el uso de la IA existe un punto fuerte de peligrosa relevancia: la *opacidad* de los algoritmos (*black box*) [(13)], que trae aparejados riesgos concretos no menores, como la falta de explicación en el razonamiento de la IA y el vacío de responsabilidad *(accountability gap*) (Barona Vilar, 2025, p. 174; Meijer & Wessels, 2019, p. 1038; Hassija *et al.*, 2024, p. 48).

En línea a lo anterior, y recogiendo un estudio sostenido en evidencia empírica, no se debe desconocer que los mismos agentes policiales han demostrado dos grandes problemas en el uso de IA para sus labores de investigación: la falta de comprensión en su funcionamiento y, por consiguiente, la estigmatización y discriminación hacia determinados individuos y grupos (Meijer & Wessels, 2019, p. 1038). Si ya la *black box* impide justificar el vínculo de causalidad del daño, no es descabellado especular la mayor indefensión en la que queda el encausado penal frente a la imposibilidad de conocer el razonamiento de una IA empleada por el Estado en la persecución, y que acentuaría el desequilibrio de poder (Vallespín Pérez, 2024, p. 21, 2025, p. 3).

Así las cosas, si somos conscientes de su carácter preventivo e informativo, así como de estar ante un documento no elaborado a petición del órgano jurisdiccional; bien pudiera ser razonable aprovechar la oportunidad para incorporar una enmienda al Código Procesal Penal chileno, recogiendo una inquietud también apreciable en el debate abierto al hilo de la Ley de Enjuiciamiento Criminal española en la que tampoco se recoge su configuración como medio de prueba (Sánchez González, 2023, pp. 320–322). En esta línea, resulta especialmente gráfica la STS 1296/2024 de 8 de marzo de

(13) Trabajos recientes han señalado que la IA es un sistema opaco, donde el funcionamiento interno del modelo no es de fácil acceso o interpretación, por lo que su proceso de razonamiento y toma de decisiones, detrás de las predicciones, no son transparentes para el usuario. Por esto, el esfuerzo actual se centra en desarrollar sistemas que no solo proporcionen la predicción, sino que también briden las explicaciones e interpretaciones, que justifiquen la decisión y la acción. (Hassija et al., 2024, p. 48)large number of methods are being reported using machine learning (ML.

2024 [14], en virtud de la cual se concluye que la prueba de inteligencia policial es una variante de la pericial; que su uso tiene sentido, en particular, en los supuestos de delincuencia organizada (informes que se realizan sobre organizaciones complejas por unidades específicas de los cuerpos y fuerzas de seguridad que tienen un conocimiento técnico sobre los hechos en cuestión, como es el caso de las unidades centrales de delincuencia económica y fiscal o la brigada antiterrorista); y que en su valoración quedan sometidas, en cuanto modelo de valoración judicial, al principio de la libre valoración probatoria. Este último *sistema de la persuasión racional* responde a la necesidad de contar con un método de apreciación libre, atendiendo necesariamente a los imperativos de la lógica, la razón y la experiencia, y unida a la responsabilidad del juez de tener que hacer constar y explicar las razones de su convencimiento (Vázquez Sotelo, 1984, pp. 519–523).

De ser así, se plantea la necesidad de que el legislador chileno sepa responder a un amplio elenco de interrogantes a los que el Boletín no ofrece respuesta: ¿Acreditan por sí mismos la ocurrencia de los hechos o requieren ser contrastados con los demás medios de prueba? ¿Deben estos documentos exponer, en su metodología, los *softwares*, las fuentes de los datos recopilados y la interacción de información para alcanzar sus conclusiones? ¿Deberá comparecer el analista de inteligencia que elaboró el documento ante el tribunal, revelando su identidad, ocupación y experticia? ¿Tendrá que exponerse ante el tribunal el algoritmo y los códigos que se emplean para su elaboración? Muchas interrogantes que merecen, sin duda, una adecuada respuesta normativa que debiera pasar, en nuestra opinión, por su correcta configuración como una prueba pericial "especializada" no vinculante y sometida, por tanto, a la libre ponderación judicial. Así, el desafío es garantizar la imparcialidad para controlar los sesgos, asegurar el derecho de defensa (transparencia y publicidad); el derecho a la tutela judicial efectiva (motivación de resoluciones); y la protección de la libertad e intimidad (Jiménez Cardona, 2023, p. 1615, 2024).

3. CONCLUSIÓN

La actual tramitación legislativa del Boletín analizado y comentado abre una oportunidad única, para regular el empleo concreto en la legislación procesal de una IA. La prevención de actividades terroristas y de criminalidad organizada se ha transformado en un "cheque blanco" para buscar las tecnologías más eficientes a esa finalidad. Eso sí, con el inminente peligro

(14) Tribunal Supremo (Sala de lo Penal), sentencia de 08 de marzo de 2024, Roj: STS 1296/2024 - ECLI:ES:TS:2024:1296. En el contexto de una guerra entre bandas criminales, ocurrió un hecho con resultado de lesiones graves, empleando machetes y armas de fuego. Para el éxito de la investigación, fue necesaria la realización de diligencias de inteligencia, con tal de dar a los involucrados y demás miembros de las agrupaciones.

de lesionar derechos fundamentales y procesales, en pos de resguardar la integridad del Estado. Siendo así, aprovechamos la ocasión para delimitar, a continuación, algunos lineamientos y consideraciones.

En primer lugar, es inevitable la tensión entre la pretensión por la eficiencia punitiva versus derechos fundamentales. Implementarse el subsistema en conjunto con IA es un riesgo crítico a las garantías y derechos, y nos invita a encender ciertas alarmas. Más aún, si consideramos que el proyecto de ley va de la mano con la propuesta de flexibilizar el levantamiento de secreto bancario, financiero y económico, sin autorización judicial previa. A su vez, el empleo de algoritmos, con el fin en la detección de comportamientos sospechosos, amenazan directamente el derecho a la intimidad, la protección de datos personales y el debido proceso.

Una segunda cuestión para considerar es la opacidad del algoritmo, que impide conocer el razonamiento detrás del resultado de su análisis. Este punto es crucial, si pretendemos legitimar un empleo masivo de la IA en la toma de decisiones y fundamentación de actos jurisdiccionales. El vacío de responsabilidad es, de por sí, un obstáculo que, en este estadio de desarrollo tecnológico sigue siendo insuperable, a pesar de los notables esfuerzos que se están desplegando. Producto de esta opacidad, también se lesiona el derecho a la defensa. ¿Cómo podrá contradecir un imputado el proceso intelectivo y racional de apreciación de la evidencia que empleó la IA, si esta no puede explicarse, y menos aún el juez quien fallará el caso? Además, las mismas fuentes humanas, que alimentan la base de datos, ya vienen contaminadas por las deficiencias que fueron expuestas. Así las cosas, es imperativo que el modelo utilizado sea transparente, tanto los algoritmos como las fuentes de datos empleados, susceptibles de auditoria informática, al menos en sus componentes esenciales, ante la instancia judicial.

Por último, pero no menos relevante, advertimos la necesidad de una regulación procesal específica para la prueba del informe de inteligencia. Este producto del futuro Subsistema de Inteligencia no encuentra un encaje claro, dentro de las categorías probatorias tradicionales. Al ser una herramienta para la toma de decisiones, en la que no existe un control judicial directo, y el eventual empleo de la IA con su opacidad inherente, anuncia serios problemas legales para su incorporación en los futuros procesos penales. Para que este instrumento y sus productos tengan validez procesal, debiese exigirse la exposición de la metodología utilizada ante el juez. La comparecencia del analista, para permitir un contrainterrogatorio efectivo sobre las conclusiones alcanzadas, ya representa un obstáculo. Todo, de por sí, representa un reto, ya que la actividad de inteligencia goza de secreto en Chile.

Somos categóricos en que el derecho no debe adaptarse a las nuevas tecnologías, sino que estas deben brindar el necesario apoyo para cumplir

el anhelo de una justicia objetiva, eficiente y oportuna. El principal desafío para el legislador es establecer un equilibrio normativo que dote al Estado de herramientas efectivas, y al mismo tiempo implementar salvaguardas robustas e inequívocas, que resguarden las garantías ciudadanas frente a un poder de vigilancia sin precedentes.

Bibliografía

Barona Vilar, S. (2025). Digital and Algorithmic Criminal Justice. The Taiwan Law Review, 363(363), 159–175. https://doi.org/10.53106/1025593136310

Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2024). Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. Cognitive Computation, 16(1), 45–74. https://doi.org/10.1007/s12559-023-10179-8

Jiménez Cardona, N. (2023). Aplicación de la inteligencia artificial en la toma de decisiones jurisdiccionales (España). Quaestio Iuris, 16(3), 1612–1630. https://doi.org/10.12957/rqi.2023.74187

Jiménez Cardona, N. (2024). Digitalización de la Justicia Española, Estrategia Justicia Digital 2030 de la Generalitat de Catalunya y Herramientas Digitales en el Derecho de Sociedades. Quaestio Iuris, 17(3), 1–30. https://dialnet-unirioja-es.sire.ub.edu/servlet/articulo?codigo=10076442

Kindhäuser, U. (2011). Retribución de culpabilidad y pena. In U. Kindhäuser & J. P. Mañalich Raffo, *Pena y culpabilidad en el estado democrático de derecho*. IB de F, Julio César Faira.

Meijer, A., & Wessels, M. (2019). Predictive Policing: Review of Benefits and Drawbacks. *International Journal of Public Administration*, *42*(12), 1031–1039. https://doi.org/10.1080/01900692.2019.1575664

Miró Llinares, F. (2020). El sistema penal ante la inteligencia artificial: Actitudes, usos, retos. In D. Dupuy, J. G. Corvalán, M. Kiefer, & D. R. Pastor (Eds.), *Cibercrimen III: inteligencia artificial, automatización, algoritmos y predicciones en el derecho penal y procesal penal*. Editorial B de f Editorial BdeF.

Molina, S., Leonel, A., Ochoa, V., & Valeria, N. (2020). El entorno virtual íntimo y privado como Derecho Humano de nueva generación. *Revista UNIANDES Episteme*, *7*(Extra 1), 1034–1047. https://revista.uniandes.edu.ec/ojs/index.php/EPISTEME/article/view/2280

Nieva-Fenoll, J. (2022). Inteligencia artificial y proceso judicial: Perspectivas ante un alto tecnológico en el camino. In S. Calaza López & M. Llorente Sánchez-Arjona (Eds.), *Inteligencia artificial legal y administración de justicia*. Thomson Reuters Aranzadi.

Rivera V., F., & Sansó-Rubert Pascual, D. (2021). *Crimen organizado y seguridad multidimensional* ([1a edición]). Tirant lo Blanch.

Romeo Casabona, C. (2020). Inteligencia artificial, predictividad y justicia penal. In D. Dupuy, J. G. Corvalán, M. Kiefer, & D. R. Pastor (Eds.), *Cibercrimen III: inteligencia artificial, automatización, algoritmos y predicciones en el derecho penal y procesal penal*. Editorial B de f Editorial BdeF.

Sánchez González, S. (2023). Análisis doctrinal y jurisprudencial de la llamada 'prueba pericial de inteligencia'. In F. Garrido Carrillo, V. Faggiani, & J. J. Romero Abolafio (Eds.), *Planificación, estrategias y medios en la lucha contra el crimen organizado y en la recuperación de activos*. Aranzadi.

Silva Sánchez, J.-M. (2001). *La expansión del derecho penal: Aspectos de la política criminal en las sociedades postindustriales* (2. ed., revisada y ampliada). Civitas.

Silva Sánchez, J.-M. (2025). *Derecho penal: Parte general*. Civitas.

Souza de Menezes, C., & Agustina, J. R. (2020). Big data, inteligencia Artificial y policía predictiva. Bases para una adecuada regulación legal que respete los derechos fundamentales. In D. Dupuy, J. G. Corvalán, M. Kiefer, & D. R. Pastor (Eds.), *Cibercrimen III: inteligencia artificial, automatización, algoritmos y predicciones en el derecho penal y procesal penal*. Editorial B de f Editorial BdeF.

Vallespín Pérez, D. (2024). La aplicación, en un contexto digital y de algoritmización de la justicia, del test P300 en el proceso penal español. In J. M. Asencio Gallego & N. Jiménez Cardona (Eds.), *Algoritmización de la justicia penal*. Juruá.

Vallespín Pérez, D. (2025). Responsabilidad civil extracontractual en materia de IA: Especial referencia a la carga de la prueba y la aplicación de presunciones. *IDP. Revista de Internet, Derecho y Política*, *0*(42), 1–11. https://doi.org/10.7238/idp.v0i42.432054

Vázquez Sotelo, J. L. (1984). *Presunción de inocencia del imputado e íntima convicción del Tribunal: Estudio sobre la utilización del imputado como fuente de prueba en el proceso penal español*. Bosch.

La inteligencia artificial en la resolución de conflictos. Una mirada crítica ante la sustitución del tercero neutral por un *SOFTWARE* inteligente

Por Almudena Valiño Ces (*)

Resumen: En los últimos años somos conscientes de que las virtualidades de la inteligencia artificial se extienden a los medios extrajudiciales de resolución de disputas *online*. Y es que no se puede obviar que la relevancia que ostentan *per se* las *Online Dispute Resolution* (ODR) se ve ampliada por su combinación con este tipo de Inteligencia, constituyendo una de las líneas que presentan un mayor potencial de desarrollo en el futuro. Y es que su utilización en este ámbito puede desempeñar dos grandes funciones: la asistencial o la decisoria. La primera, como mecanismo de apoyo a quienes asumen responsabilidades en la controversia, ya sean las partes o el tercero ajeno a la que vaya a intervenir, ofreciendo datos y predicciones que puedan ayudar u orientar en alguna de las decisiones o actuaciones que deban desenvolver en su labor. Y la segunda, la función decisoria, la cual implica que la persona que tiene que resolver o mediar en el conflicto va a ser sustituida por un sistema de inteligencia artificial. Pese a haber estas dos posibles funciones, en el presente trabajo nos centraremos en el análisis de esta segunda por ser la que más polémica y dificultades plantea a la hora de su aplicación.

Palabras clave: resolución de conflictos; ADR; ODR; inteligencia artificial; función decisoria.

Abstract: In recent years, we have become aware that the potential of Artificial Intelligence extends to extrajudicial means of online dispute resolution. It cannot be ignored that the relevance of Online Dispute Resolution

---

(*) Profesora ayudante. Doctora de Derecho Procesal de la Universidade de Santiago de Compostela. Correo electrónico: almudena.valino@usc.es. Profesora ayudante. Doctora de Derecho Procesal de la Universidad de Santiago de Compostela (acreditada a contratada doctora). Ha participado en diversos proyectos de investigación. Miembro de dos Grupos de Innovación Docente. Miembro de la Asociación de Profesores de Derecho Procesal de las Universidades Españolas. Autora de más de 90 publicaciones. Ponente y comunicante en más de 80 actividades nacionales e internacionales, además de haber dirigido 12 obras colectivas y coordinado 6. Y ha realizado una estancia de investigación durante 3 meses en la Facultad de Dereito da Universidade do Porto.

(ODR) is amplified by its combination with this type of Intelligence, constituting one of the areas with the greatest potential for future development. Its use in this area can serve two main functions: assistance and decision-making. The first is as a support mechanism for those who assume responsibility in the dispute, whether they are the parties or the third party outside the dispute who will intervene, offering data and predictions that can help or guide any of the decisions or actions they must take in their work. The second is the decision-making function, which involves replacing the person who has to resolve or mediate the conflict with an Artificial Intelligence system. Despite these two possible functions, in this paper we will focus on the analysis of the second, as it is the most controversial and difficult to apply.

Key words: conflict resolution; ADR; ODR; artificial intelligence; decision-making function.

1. INTRODUCCIÓN

En la actualidad estamos siendo testigos del elevado número de conflictos que cada día surgen en la sociedad, los cuales, además, presentan una naturaleza muy variada, lo que conlleva un gran tráfico jurídico. Para tratar de resolverlos, la ciudadanía ha recurrido tradicionalmente al proceso judicial como la vía habitual utilizada en el derecho español. Con todo, ante una realidad que muestra que el proceso resulta, en no pocas ocasiones, ineficiente, no puede considerarse este como la única vía para lograr la tutela efectiva, sino que han ido surgiendo diversas formas para tratar de solucionarlos, que "van desde unas muy primitivas y simples hasta procesos altamente especializados y regulados" (Rivera Morales, 2017, p. 2). Buen ejemplo de ello son los métodos alternativos de resolución de conflictos o *Alternative Dispute Resolution* (ADR), como habitualmente se denominan. Estas ADR surgen como una respuesta a las necesidades de un número de conflictos, en los que se busca la satisfacción de los intereses de las partes de manera rápida, sencilla y económica.

Ello no obstante, en los últimos tiempos, podemos comprobar como Internet constituye un ámbito adecuado para la celebración de contratos cibernéticos y, en consecuencia, en un ámbito proclive para la aparición de problemas. Es por ello que, junto a estas ADR, aparecieron los mecanismos de resolución en línea: *Online Dispute Resolution* (ODR), como aquellas modalidades que se caracterizan por desarrollar su actividad, ya sea total o parcialmente, por vía electrónica.

Con ello y con todo, nos encontramos en una época en la que se han logrado importantes avances con la introducción de las nuevas tecnologías, lo que sin duda constituye un paso más que necesario de cara a lograr la tan

ansiada agilización del sistema judicial español [1]. Concretamente, podemos afirmar que entre las grandes revelaciones destaca, sin duda, la inteligencia artificial (IA), en la medida en que ostenta un papel trascendental como una herramienta que permite poner en práctica métodos tecnológicos, además de coadyuvar a acercar la Justicia al conjunto de la ciudadanía, en cuanto su acceso se presume más sencillo. Y es que, si se aplica de forma correcta y se le da un uso adecuado, puede aportar grandes beneficios para el ciudadano y para los profesionales del derecho.

2. Los métodos de resolución de conflictos y la inteligencia artificial

En los últimos años, estamos observando cómo la utilización de nuevas tecnologías en numerosos ámbitos de la vida, incluido el de la Justicia, ha experimentado un incremento exponencial. Y como no podía ser de otro modo, la resolución de conflictos no ha quedado al margen [2], por lo que el arbitraje y la mediación también se encuentran inmersos en esta evolución, la cual viene marcada esencialmente por dos aspectos distintivos. Por un lado, "la canalización de la resolución de arbitrajes y mediaciones hacia sedes y plataformas electrónicas", en expresión de lo que venimos denominando como ODR y, por el otro, la aplicación inevitable de la IA, que efectivamente va abriéndose paso en el entorno de las ADR y ODR (Martín Diz, 2021, p. 96).

Así las cosas, las virtualidades de la IA se extienden a los medios extrajudiciales de resolución de disputas *online*. No se puede obviar que la relevancia que ostentan *per se* las ODR se ve ampliada por su combinación con la IA, constituyendo una de las líneas que presentan un mayor potencial de desarrollo en el futuro. Y es que su utilización en este ámbito puede desempeñar dos grandes funciones: la asistencial o la decisoria (González Fernández, 2023, p. 184). La primera, como mecanismo de apoyo a quienes asumen responsabilidades en la controversia, ya sean las partes –para decidir sobre el acceso a una ODR o designar a quienes ocupen la posición de tercero imparcial en la resolución del litigio– o ya sea el tercero ajeno al que

---

(1) Pastor Prieto (2003, p. 13) alude al grado en que se aprovechan las posibilidades que ofrecen las aplicaciones informáticas, indicando que la informatización de los órganos judiciales fue diseñada para ser una herramienta facilitadora del impulso procesal. A este respecto, Yáñez Velasco (2008, p. 611) hace referencia al problema de calidad consecuencia de la dilación asociada directamente a una infrautilización de las tecnologías, lo cual enlaza con el nivel de Justicia derivado del grado de agilidad en las respuestas a las controversias.

(2) Una de las aplicaciones de la IA que se está llevando a cabo, en cuanto al orden jurisdiccional civil y social, viene referida a la resolución extrajudicial de conflictos. Ejemplo de ello, como afirma Simón Castellano (2021, p. 6), es la utilización de "técnicas de Justicia predictiva por parte de las compañías aseguradoras, que calculan y evalúan las posibilidades de éxito de acudir a los tribunales" y, en el supuesto de que resulte inferior a un determinado porcentaje, optan por acudir a métodos de resolución extrajudicial de conflictos y evitar los costes que conlleva un proceso judicial.

ha de resolverlo, si es un árbitro, o ayudar a hacerlo, si es un mediador, ofreciendo datos y predicciones que puedan ayudar u orientar en alguna de las decisiones o actuaciones que deban desenvolver en su labor. Por su parte, la función decisoria implica que la persona que va a actuar como tercero, bien en la mediación o bien en el arbitraje, va a ser sustituida por una IA. Sea como fuere, en cualquiera de las dos posibilidades, el objetivo de la IA en el contexto jurídico será resolver problemas concretos. Pese a ello, una IA en el ámbito del derecho no necesariamente ha de articularse con el fin de utilizarse en litigios, sino que puede, por ejemplo, ser una herramienta para buscar antecedentes, para redactar documentos o para valorar datos objetivos con consecuencias jurídicas (Martín Diz, 2020, p. 6) [3].

Con todo, debemos de tener en cuenta que el derecho, la Justicia y la resolución de conflictos han de estar siempre bajo supervisión y control humano. En palabras de Martín Diz, la máquina es controlada por el ser humano y este fortalece su aplicación. Este principio *–human in command–*, establecido en el Dictamen del Comité Económico y Social Europeo, constituye el núcleo desde el cual debiera expandirse la utilización de la IA (Martín Diz, 2020, p. 65).

3. LA SUSTITUCIÓN DEL TERCERO NEUTRAL POR UN SISTEMA DE INTELIGENCIA ARTIFICIAL EN LA RESOLUCIÓN DE CONFLICTOS: LA FUNCIÓN DECISORIA

Más allá de considerar que la IA podría utilizarse sin reparos como herramienta asistencial en una ADR o en una ODR (Valiño Ces, 2025, pp. 383-387), cuando nos referimos a la posibilidad de que esa inteligencia asuma la función heterocompositiva o autocompositiva, no sucede lo mismo. Gracias a la flexibilidad y adaptación de la IA al entorno jurídico, *a priori* podríamos afirmar que no existe limitación alguna que lo impida, siempre y cuando se trate de litigios vinculados a derechos disponibles por las partes, ni tampoco desde un plano tecnológico, pues ya existen herramientas de IA diseñadas para ejercer como árbitros o mediadores. En suma, el carácter flexible propio de las ADR y de las ODR, permitiría la incorporación de la función decisoria o facilitadora de los denominados i-árbitros e i-mediadores, o lo que es lo mismo, sería factible la existencia de inteligencias artificiales que asuman esa función, tanto en la versión judicial o extrajudicial, como en la presencial o en línea del arbitraje o de la mediación [4].

---

(3) A este respecto, cabe destacar que hay experiencias innovadoras en el ámbito de la mediación. En concreto, existen aplicaciones de resolución extrajudicial de conflictos y, especialmente, medios de resolución de disputas en línea: arbitraje y mediación electrónica junto a sistemas de negociación automatizada o asistida. Para saber más sobre la resolución inteligente de disputas véase: https://odrsolvo.com/.

(4) Esa sustitución puede llevarse a cabo en la forma de agentes relacionales, robots, avatares, etc., diseñados y calibrados como solución tecnológica a estos efectos (Martín Diz, 2020, pp. 67 y 68).

Ahora bien, a la vista del momento tan cambiante en el que nos encontramos, podríamos afirmar que todavía no estamos preparados, o no para asimilarlo, para que un sistema de IA pueda sustituir a un árbitro o a un mediador a la hora de adoptar decisiones o ayudar a las partes en el marco de un procedimiento arbitral o de mediación, respectivamente. Y ello porque todavía existen determinados escollos difíciles de superar, como puede ser el de la posible afectación de derechos fundamentales de los justiciables [5] y los aspectos éticos que plantea la sustitución del factor humano por una máquina.

Más allá de lo acabado de señalar, y desde un punto de vista estrictamente jurídico, cabría preguntarse primero si existe o no la posibilidad de que una IA actúe como tercero neutral. Si acudimos a la legislación nacional en materia de ADR, se puede comprobar que tanto la ley 60/2003, de 23 de diciembre, de Arbitraje, como la ley 5/2012, de 6 de julio, de mediación en asuntos civiles y mercantiles, en sus artículos 13 y 11 respectivamente, exigen que el tercero –árbitro o mediador– sean personas naturales. Por ello, para que en nuestro país una máquina sustituya totalmente al tercero neutral, sería necesario modificar estos artículos a fin de incorporar a las "personas electrónicas".

En cuanto a la legislación internacional [6], podemos observar que se omite la referencia expresa a la necesaria condición "humana" del árbitro o del mediador, en general del tercero neutral. A este respecto, Marcos Francisco entiende que la falta de mención expresa podría interpretarse en el sentido de que es posible que una IA actúe como tercero neutral. Sin embargo, también resulta posible entender lo contrario, defendiendo que del cuerpo normativo se desprende claramente –al referirse a la "nacionalidad" del árbitro [7] o del mediador [8]– la necesaria condición de persona física que debe tener ese tercero. Por este motivo, para garantizar una mayor seguridad, lo más apropiado sería que, de forma expresa, las normas lo admitieran o lo prohibieran (Marcos Francisco, 2023, pp. 26 y 27). En efecto, para ofrecer una mayor seguridad jurídica, lo más conveniente sería modificar

---

(5) Para profundizar sobre la protección de los derechos fundamentales de las partes cuando estamos en presencia de sistemas de IA, véase, entre otros: Martín Diz, 2021, pp. 969-1006 y Valiño Ces, 2023, pp. 49-70.

(6) Nos referimos a la "Convención de Nueva York" de 1958 sobre el reconocimiento y la ejecución de las sentencias arbitrales extranjeras, a la Ley Modelo de dicha CNUDMI sobre Arbitraje Comercial Internacional de 1985 –con enmiendas adoptadas en 2006– o a la Ley Modelo sobre Conciliación Comercial Internacional, adoptada en el año 2002, que se modificó en el año 2018 y pasó a llamarse "Ley Modelo sobre Mediación Comercial Internacional y Acuerdos de Transacción Internacionales Resultantes de la Mediación".

(7) *Vid.* apartados 1 y 5 *in fine* del artículo 11 – "Nombramiento de los árbitros" – de la Ley Modelo de Arbitraje Comercial Internacional.

(8) *Vid.* apartado 4 del artículo 6° – "Número y designación de mediadores" – de la Ley Modelo de Mediación Comercial Internacional.

la referida legislación, nacional e internacional, a fin de prever claramente tal posibilidad.

A este respecto, Martín Diz, con base en que la ventaja de la aplicación de la IA a las ODR es que "apenas hay que llevar a cabo reformas legales", entiende que la legislación nacional e internacional aplicables no impiden ni prohíben que se designe un árbitro o un mediador inteligente, "aunque habitualmente la referencia se entienda a un árbitro o mediador humano", debido a que ello sería posible por la flexibilidad propia de los ODR y porque se inspiran en el principio de libre disposición de las partes y autonomía de su voluntad (Martín Diz, 2020, pp. 6 y 22). Esta idea podría venir respaldada por el propio Tribunal Constitucional. Y ello porque si el Tribunal en resoluciones como la sentencia, de la Sala Segunda, 65/2021, de 15 de marzo, ha llegado a admitir que las partes pueden acordar que el laudo no tenga que motivarse [9], se podría defender lo mismo en relación con la no necesidad de que sea una persona natural quien actúe como tercero neutral, admitiéndose que las partes pacten que lo sea una IA (Marcos Francisco, 2004, p. 132).

Al hilo de esta sentencia, otra de las cuestiones que plantea la utilización de un sistema de IA en la función decisoria es que en estos casos no se puede hablar de decisiones motivadas, en la medida en que los resultados que arroja este sistema derivan de un proceso que responde a una pura estadística y cuya explicación se basa en meras operaciones matemáticas. Ciertamente, la cuestión de la legitimidad en el razonamiento informático y autónomo de la IA suscita dudas, en cuanto apenas existen sistemas preparados para motivar sus decisiones en el ámbito jurídico.

En efecto, que un i-mediador facilite una propuesta de solución sin razonarlo o justificarlo, podría llegar a admitirse porque su aceptación depende de las partes. Sin embargo, que un i-árbitro dicte un laudo sin una motivación o razonamiento alguno se puede entender inadmisible, máxime cuando el artículo 37.4 de la Ley de Arbitraje dispone que "[e]l laudo deberá ser siempre motivado (…)". Al hilo de la STC 65/2021, de 15 de marzo, *supra* referida, y a diferencia de lo que ocurre en la vía judicial, las partes podrían convenir que no sea preciso motivar el laudo. Con todo, esto no resulta sencillo, toda vez que a la hora de resolver o de proponer soluciones, es de vital

(9) Su fundamento jurídico quinto dispone: "Asentado, por consiguiente, el arbitraje en la autonomía de la voluntad y la libertad de los particulares (arts. 1° y 10 CE), el deber de motivación del laudo no se integra en el orden público exigido en el art. 24 CE para la resolución judicial, sino que se ajusta a un parámetro propio, definido en función del art. 10 CE. Este parámetro deberán configurarlo, ante todo, las propias partes sometidas a arbitraje a las que corresponde, al igual que pactan las normas arbitrales, el número de árbitros, la naturaleza del arbitraje o las reglas de prueba, pactar si el laudo debe estar motivado (art. 37.4 LA) y en qué términos. En consecuencia, la motivación de los laudos arbitrales carece de incidencia en el orden público". En estos términos se pronuncia la STC, Sala Segunda, 50/2022, de 4 de abril (fundamento jurídico tercero) o la STC, Sala Primera, 79/2022, de 27 de junio (fundamento jurídico segundo).

importancia contar con una capacidad humana que, más allá de toda operación mecánica, permita aplicar el proceso interno y humano que conduzca a decidir, en el caso del arbitraje, o asistir, en el de la mediación, en un sentido u en otro en relación con el asunto. Naturalmente, el componente subjetivo que está ínsito en un tercero y que le ha llevado a tomar determinadas decisiones, resulta casi imposible trasladarlo a una IA.

Además, en este sentido, también entra en juego el riesgo de sesgo, discriminación o la imparcialidad respecto a los prejuicios que pueda tener la IA fruto de su diseño a la hora de introducir los datos y alimentar el algoritmo o a la hora de programar el *software*. Lo cierto es que este problema podría solventarse si todos los aspectos del programa de la computadora y los algoritmos utilizados hubiesen sido previamente estudiados por las partes en el litigio antes de someter la controversia a esta herramienta. Sin embargo, esta solución no haría otra cosa que dilatar el procedimiento e incrementar el coste al tener que recurrir a expertos informáticos y jurídicos que lo verificasen (Martín Diz, 2021, pp. 105 y 106).

A este respecto, el Parlamento Europeo, en su resolución de 20 de enero de 2021, *Inteligencia artificial: cuestiones de interpretación y de aplicación del derecho internacional* [(10)], solicita que se "mantenga informado al público sobre el uso de la IA en el ámbito de la justicia, y que dichos usos no den lugar a discriminación derivada de sesgos de programación".

Asimismo, podríamos plantearnos si la introducción de un sistema de IA que sustituya al ser humano y actúe como tercero neutral afectaría a la esencia de las ADR/ODR. En cuanto al arbitraje, la existencia de un i-árbitro, *a priori,* no tiene por qué afectar a su esencia. Ello no obstante, esta afirmación no significa que la cualidad humana no sea importante, toda vez que el tercero neutral que propone o impone una decisión lo hace con base en sus propias experiencias personales y emocionales y su interactuación con las partes, por lo que para adoptar una u otra decisión va más allá de lo estrictamente lógico-jurídico. En cambio, en cuanto a la mediación, las cosas difieren en la medida en que el papel del mediador es capital y, con él, la capacidad de escucha y comprensión, la construcción de confianza, la generación de empatía, la detección de emociones, etc., que en la actualidad es impensable que posea un *software* inteligente. De hecho, a las personas que intervengan en una mediación puede resultarles extraño y poco amigable llevar a cabo esa interacción con ese tipo de sistema inteligente, lo que podría llevarles a querer desistir de esta y optar por sistemas con

(10) Resolución del Parlamento Europeo, de 20 de enero de 2021, sobre inteligencia artificial: cuestiones de interpretación y de aplicación del derecho internacional en la medida en que la UE se ve afectada en los ámbitos de los usos civil y militar, así como de la autoridad del Estado fuera del ámbito de la justicia penal. Disponible en: https://www.europarl.europa.eu/doceo/document/TA-9-2021-0009_ES.pdf

componente humano (Martín Diz, 2020, p. 68). A este respecto, González Fernández (2023, p. 223) señala que el factor humano del mediador es indispensable, pues su intervención, por mínima que sea, siempre va a tener un carácter fundamental, por ejemplo, "por su forma de percibir los hechos de una de las partes en detrimento de un robot". En suma, en este sentido, entendemos que todavía queda un largo camino –o igual no tan largo–, para conseguir que un sistema de IA desarrolle totalmente una verdadera mediación (Ordelin Font, 2021, pp. 372-374).

Asimismo, en esta función decisoria por parte de un sistema inteligente, existe otra problemática y es que en la actualidad no resulta posible garantizar la ausencia de fallos y la perfección total de un sistema de IA, en cuanto este sistema puede presentar "defectos internos (defectuosa calibración o programación) o externos (mala conexión, ataque al sistema)" (Martín Diz, 2020, p. 68). Todo ello puede implicar que esa IA no cumpla sus funciones de manera apropiada.

Este inconveniente, sin duda, conlleva otra cuestión relevante y es la posible responsabilidad derivada de la actuación de la IA como i-árbitro o i-mediador, sobre todo al tomar decisiones jurídicamente vinculantes, como sucede en el arbitraje. Por tanto, la duda que surge es ¿quién responde jurídicamente por ciertos hechos ilícitos, o conductas inadecuadas, cuando quien asiste o toma la decisión es un *software* inteligente? Varias son las opciones. Podríamos pensar que la responsabilidad recaería en el programador, en quien lo haya comercializado o, en su caso, en la institución de ADR/ODR.

La Unión Europea se ha pronunciado a este respecto. En virtud de la Resolución del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de derecho civil sobre robótica [2015/2103 (INL)] y a fin de que se pueda responder civilmente de los daños y perjuicios causados, llegó a plantear la posible atribución de personalidad jurídica a robots o inteligencias artificiales. En efecto, en este texto se pidió a la Comisión considerar la creación a largo plazo de "una personalidad jurídica específica para los robots, de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente" y que determinase "si debe aplicarse el enfoque de la responsabilidad objetiva o el de gestión de riesgos".

A pesar de ello, el Dictamen del Comité Económico y Social Europeo sobre la "Inteligencia artificial: las consecuencias de la inteligencia artificial para el mercado único (digital), la producción, el consumo, el empleo y la sociedad" (2017/C 288/01), en sus "conclusiones y recomendaciones", se

pronunció "en contra de la introducción de cualquier tipo de personalidad jurídica para los robots o la IA, puesto que socavaría los efectos correctores preventivos de la legislación en materia de responsabilidad, generaría un riesgo moral tanto en el desarrollo como en la utilización de la IA y daría lugar a un posible uso indebido" (punto 1.12). Y posteriormente, la Resolución del Parlamento Europeo, de 20 de octubre de 2020, con recomendaciones destinadas a la Comisión sobre un régimen de responsabilidad civil en materia de inteligencia artificial [2020/2014 (INL)], fue en la misma línea al disponer que "todas las actividades, dispositivos o procesos físicos o virtuales gobernados por sistemas de IA pueden ser técnicamente la causa directa o indirecta de un daño o un perjuicio, pero casi siempre son el resultado de que alguien ha construido o desplegado los sistemas o interferido en ellos; observa, a este respecto, que no es necesario atribuir personalidad jurídica a los sistemas de IA".

Habida cuenta de lo acabado de señalar, parece que queda descartada la idea de atribuir personalidad jurídica a la propia IA y, por ende, que esta responda civilmente de los daños y perjuicios causados. En consecuencia, lo suyo sería determinar, en caso de que interviniese una IA en la decisión, quién es el posible responsable y el tipo de responsabilidad que existiría, lo que podría hacerse a través de un eventual estatuto de la persona electrónica, máxime cuando también podría haber algún tipo de responsabilidad penal.

Sea como fuere, una cuestión esencial es la necesidad de determinar una regulación que apueste decididamente por las TIC, pero sobre todo y con las debidas cautelas por la IA. Y es que somos conscientes de la importancia de contar con un marco normativo que ofrezca una regulación completa en esta materia.

Así las cosas, debemos de acudir al Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial. Uno de los aspectos reseñables de este texto es la clasificación del nivel de riesgo de los diferentes sistemas de IA, entendiendo por riesgo "la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio".

A ese respecto, su considerando 26 dispone: "Con el fin de establecer un conjunto proporcionado y eficaz de normas vinculantes para los sistemas de IA, es preciso aplicar un enfoque basado en los riesgos claramente definido, que adapte el tipo y contenido de las normas a la intensidad y el alcance de los riesgos que puedan generar los sistemas de IA de que se trate". En atención a ello, se entiende de alto riesgo "los sistemas de IA destinados a ser utilizados por los organismos de resolución alternativa de litigios con esos fines, cuando los resultados de los procedimientos de resolución alternativa de litigios surtan efectos jurídicos para las partes. La utilización

de herramientas de IA puede apoyar el poder de decisión de los jueces o la independencia judicial, pero no debe substituirlas: la toma de decisiones finales debe seguir siendo una actividad humana" (considerando 61).

De igual modo, se prevé en el Anexo III que, de conformidad con el artículo 6.2, son de alto riesgo los sistemas de IA "destinados a ser utilizados por una autoridad judicial, o en su nombre, para ayudar a una autoridad judicial en la investigación e interpretación de hechos y de la ley, así como en la garantía del cumplimiento del derecho a un conjunto concreto de hechos, o a ser utilizados de forma similar en una resolución alternativa de litigios".

En atención a esta regulación, no resta sino reiterar la idea expuesta supra, de que no habría inconveniente en que un sistema de IA asistiese al tercero en su labor de ayuda a las partes para resolver su conflicto, es más, en ese sentido, puede aportar numerosas ventajas. En cambio, cuando hablamos de la función decisoria no es tan sencillo. El reglamento entiende que nos encontramos ante una cuestión que implica un alto riesgo. Por este motivo, y con arreglo al artículo 14, para que un sistema de IA pueda adoptar decisiones, será necesario que, en todo caso, exista una supervisión humana a fin de prevenir o reducir al mínimo los riesgos que se puedan derivar del uso indebido o razonablemente previsible de la IA y cuando las decisiones basadas únicamente en el procesamiento automatizado produzcan efectos jurídicos o de otro tipo a las personas con los que se utiliza el sistema (Barrios Andrés, 2024, pp. 300 y 301). De esta manera, durante el periodo que esté en uso ese sistema podrá estar vigilado por personas físicas, a través de herramientas de interfaz humano-máquina adecuadas y si fuese necesario el responsable de ello puede decidir no utilizar la IA o descartar los resultados que se generen [(11)].

4. CONCLUSIONES

A la vista de lo analizado, podemos concluir que no parece conveniente utilizar inteligencias artificiales como herramientas decisorias. Por tanto, en países como el nuestro, resulta una tarea compleja que un *software* actúe como tercero neutral en las ODR. Con ello y con todo, dado el constante avance de las tecnologías, no podemos descartar que, en un futuro no muy lejano, las máquinas, robots o avatares, de forma autónoma y respetando la esencia de las ADR/ODR, puedan actuar como árbitros o mediadores. Incluso, comparto con Martín Diz (2021, pp. 111 y 112) la posibilidad de plantear una fórmula híbrida, esto es, configurar un modelo en el que el tribunal arbitral esté formado por una persona y un *software* inteligente o una comediación en la que intervenga un humano y una IA.

(11) Artículo 14.1, 2 y 4 d) del reglamento.

En cualquier caso, de lo que sí estamos seguros es de que la incorporación de la IA en los procedimientos de ODR, indudablemente, contribuirá a gestionarlos de un modo mucho más eficaz y eficiente, tanto desde un punto de vista económico como temporal, en la medida en que las "máquinas" pueden trabajar sin descanso. Con todo, tendremos que esperar a fin de comprobar si realmente estamos preparados para que sea una máquina la que nos resuelva o ayude a resolver nuestros conflictos.

Bibliografía

Barrio Andrés, M. (Dir.). (2004). Comentarios al Reglamento Europeo de Inteligencia Artificial, Madrid: La Ley.

González Fernández, A. I. (2023). La mediación como método de resoluciones de controversias, Valencia: Tirant lo Blanch.

Marcos Francisco, D. (2004). Sistema Arbitral de Consumo: algunas propuestas "inteligentes" de lege ferenda, InDret, 1.

Marcos Francisco, D. (2023). Smart ODR y su puesta en práctica: el salto a la inteligencia artificial, Revista General de Derecho Procesal, núm. 59.

Martín Diz, F. (2020). Inteligencia artificial y medios extrajudiciales de resolución de litigios online (ODR): evolución de futuro en tiempos de pandemia global (Covid-19), La Ley, Mediación y arbitraje, núm. 2 (abril-junio), Sección Doctrina.

Martín Diz, F. (2020). Justicia digital post-covid19: el desafío de las soluciones extrajudiciales electrónicas de litigios y la inteligencia artificial, Revista de Estudios Jurídicos y Criminológicos, núm. 2, Universidad de Cádiz, pp. 41-74. DOI: https://doi.org/10.25267/REJUCRIM.2020.i2.03

Martín Diz, F. (2021). ADR, ODR e inteligencia artificial: evolución en el arbitraje y la mediación. En Fernández Pérez, A. (Dir.), Interacción entre mediación y arbitraje en la resolución de litigios internacionales del siglo XXI, Cizur Menor (Navarra): Aranzadi, pp. 95-118.

Martín Diz, F. (2021). Inteligencia Artificial y Derecho. Procesal: luces, sombras y cábalas en clave de derechos fundamentales. En Moreno Catena, V. y Romero Pradas, M. I. (Dirs.), Nuevos postulados de la cooperación judicial en la Unión Europea. Libro homenaje a la Profª. Mª Isabel González Cano, Valencia: Tirant lo Blanch, pp. 969-1006.

Ordelin Font, J. L. (2021). El uso de la inteligencia artificial en la mediación: ¿quimera o realidad?, Revista del Instituto de Ciencias Jurídicas de Puebla, vol. 15, núm. 48, pp. 372-374.

Pastor Prieto, S. (2003). Los nuevos sistemas de organización y gestión de la justicia: ¿mito o realidad? En III Conferencia sobre Justicia y Desarrollo en América Latina y el Caribe, Banco Interamericano de Desarrollo, Quito,

Ecuador, p. 13. Disponible en: https://biblioteca.cejamericas.org/handle/2015/345.

Rivera Morales, R. (2017). La Mediación Electrónica como forma alternativa de solución de conflictos, Actualidad Civil, núm. 5, Sección Derecho digital / A fondo, Wolters Kluwer.

Simón Castellano, P. (2021). Inteligencia artificial y Administración de Justicia: ¿Quo vadis, justitia?, IDP. Revista de Internet, Derecho y Política, núm. 33.

Valiño Ces, A. (2023). La inteligencia artificial y su aplicación en el proceso: los riesgos para la protección y garantía de los derechos. En Hernández López, A. y Laro González, E. (Dirs.), Proceso penal europeo: últimas tendencias, análisis y perspectivas, Aranzadi, Cizur Menor (Navarra), pp. 49-70.

Valiño ces, A. (2025). La irrupción de la Inteligencia Artificial en el proceso: una mirada a los mecanismos de resolución de conflictos. En Noya Ferreiro, M. L. (Dir.) y Catalina Benavente, M.ª A. (Coord.), La inteligencia artificial y su aplicación en el sistema de Justicia penal, Madrid: Aranzadi, pp. 371-399.

Yáñez Velasco, R. (2008). L'Oficina Judicial a Catalunya. Mitjà real d'una Justícia eficaç pel ciutadà del segle XXI, Barcelona: Centro de Estudios Jurídicos y Formación Especializada.

Políticas do reconhecimento facial: regulação entre Brasil e União Europeia

Por Ana Clara Santos Elesbão (*)

Resumo: Este trabalho analisa criticamente a regulação do uso de Tecnologias de Reconhecimento Facial (TRFs) no Brasil e na União Europeia, com foco nos respectivos marcos regulatórios da Inteligência Artificial (IA). O objetivo é analisar e comparar o tratamento dado à matéria pelos marcos legais brasileiro e europeu, buscando compreender quais são os potenciais impactos da regulação da IA em face do atual regime de vigilância e controle mobilizado pelo uso dessas tecnologias. A pesquisa é qualitativa e utiliza como técnicas a revisão bibliográfica, análise documental e o estudo comparado. O estudo demonstra que a adoção do paradigma regulatório de risco, tanto na União Europeia quanto no Brasil, subordina a proteção de direitos à lógica econômica da inovação e do desenvolvimento tecnológico, enquanto legitima e naturaliza práticas de vigilância e controle securitário sob o crivo da governança e do controle regulatório. A multiplicidade de exceções e brechas interpretativas em torno do uso de sistemas biométricos em contextos de segurança evidencia uma tendência regulatória que autoriza amplamente a utilização dessas tecnologias. Conclui-se que a linguagem da regulação de riscos produz um regime jurídico que naturaliza a violação de direitos como efeito colateral aceitável do desenvolvimento tecnológico, contribuindo para a expansão do novo regime de vigilância e controle securitário em curso.

Palavras-chave: Reconhecimento Facial. Inteligência Artificial. Regulação.

Políticas del reconocimiento facial: regulación entre Brasil y la Unión Europea

Resumen: Este estudio analiza críticamente la regulación del uso de Tecnologías de Reconocimiento Facial (TRF) en Brasil y en la Unión Europea, con énfasis en sus respectivos marcos regulatorios de inteligencia

(*) Pontifícia Universidade Católica do Rio Grande do Sul. E-mail: anaelesbaos@gmail.com. Doutoranda e Mestra pelo Programa de Pós-Graduação em Ciências Criminais da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Bacharela em Ciências Jurídicas e Sociais pela PUCRS. Realizou Doutorado Sanduíche na Universidad de Sevilla (US/Espanha). Pesquisadora vinculada ao Grupo de Pesquisa "Criminologia, Cultura Punitiva e Crítica Filosófica".

artificial (IA). El objetivo es analizar y comparar el tratamiento otorgado al tema por los marcos legales brasileño y europeo, buscando comprender los posibles impactos de la regulación de la IA en el contexto del actual régimen de vigilancia y control posibilitado por estas tecnologías. La investigación es de carácter cualitativo y emplea como técnicas la revisión bibliográfica y documental, así como el estudio comparado. El estudio demuestra que la adopción del paradigma regulatorio basado en el riesgo, tanto en la Unión Europea como en Brasil, subordina la protección de los derechos a la lógica económica de la innovación y del desarrollo tecnológico, al mismo tiempo que legitima y naturaliza prácticas de vigilancia y control securitario bajo el discurso de la gobernanza y del control regulatorio. La multiplicidad de excepciones y vacíos interpretativos en torno al uso de sistemas biométricos en contextos de seguridad evidencia una tendencia regulatoria que autoriza ampliamente la utilización de estas tecnologías. Se concluye que el lenguaje de la regulación de riesgos produce un régimen jurídico que naturaliza la violación de derechos como un efecto colateral aceptable del desarrollo tecnológico, contribuyendo así a la expansión del nuevo régimen de vigilancia y control securitario en curso.

Palabras clave: Reconocimiento Facial. Inteligencia Artificial. Regulación.

Politics of facial recognition: regulation between Brazil and the European Union

Abstract: This study critically examines the regulation of Facial Recognition Technologies (FRTs) in Brazil and the European Union, focusing on their respective Artificial Intelligence (AI) regulatory frameworks. The aim is to analyze and compare how these legal systems address the issue, seeking to understand the potential impacts of AI regulation within the current regime of surveillance and control enabled by such technologies. The research follows a qualitative approach, employing bibliographic and documentary review methods, as well as comparative analysis. The findings demonstrate that the adoption of the risk-based regulatory paradigm, both in the European Union and in Brazil, subordinates the protection of rights to the economic logic of innovation and technological development, while simultaneously legitimizing and normalizing practices of surveillance and discrimination under the guise of governance and regulatory control. The multiplicity of exceptions and interpretative loopholes surrounding the use of biometric systems in security contexts reveals a regulatory tendency that broadly authorizes the deployment of FRTs for securitarian purposes. It concludes that the language of risk regulation produces a legal regime that naturalizes rights violations as an acceptable side effect of technological development, thereby contributing to the expansion of the emerging regime of surveillance and security control.

Keywords: Facial Recognition. Artificial Intelligence. Regulation.

1. Introdução

As Tecnologias de Reconhecimento Facial (TRFs) reafirmam e atualizam a lógica biopolítica de vigilância e controle securitário que intensifica práticas de intervenção violenta sobre os corpos. Ao operar por meio da extração e análise de dados biométricos, as TRFs transformam corpos em objetos de rastreamento e categorização, reiterando epistemologias raciais que vinculam a vigilância à construção do corpo negro, indígena, migrante e dissidente como alvo preferencial do olhar policial. Longe de representar um avanço neutro, tais tecnologias prolongam heranças coloniais e dinâmicas de segregação institucionalizadas, convertendo a promessa de segurança em um instrumento de gestão diferencial da vida, que amplia a vulnerabilidade de grupos historicamente criminalizados e naturaliza formas contemporâneas de controle social.

Em face ao complexo cenário em que o rosto é traduzido em linhas de código para a produção do dado biométrico em contextos de vigilância e controle securitário, emergem tensões em torno da violação de direitos. A proliferação de marcos regulatórios surge como resposta aos tensionamentos postos, a fim de conferir respaldo de legitimidade ao novo regime de práticas. O presente artigo tem como objetivo analisar como os marcos regulatórios da IA vêm se posicionando diante do uso de Tecnologias de Reconhecimento Facial (TRFs), e de que modo contribuem para consolidar o novo regime de vigilância e controle securitário em curso. A pesquisa, de natureza qualitativa, utiliza revisão bibliográfica, análise documental e estudo comparado para analisar o Regulamento de Inteligência Artificial da União Europeia e o Projeto de Lei n.º 2.338/2023, em tramitação no Congresso Nacional brasileiro, buscando oferecer subsídios para melhor compreender o estatuto dos marcos regulatórios no cenário global em face do atual contexto securitário.

2. Análise do regulamento de Inteligência Artificial da União Europeia *(AI Act)*

Aprovado em março de 2024, o Regulamento de Inteligência Artificial da União Europeia (Parlamento Europeu, 2024) segue uma regulação assimétrica baseada na gradação de riscos. Trata-se de um modelo regulatório que aplica requisições e obrigações sobre governança, supervisão humana e transparência de modo diferencial, a depender do nível de risco oferecido pelo sistema (Bioni *et al.*, 2023). No caso do *AI Act*, a regulação baseada em riscos adotada adapta o tipo e o conteúdo normativo do regulamento à intensidade e ao âmbito dos riscos que podem ser criados pelos sistemas de IA, categorizando-os em "risco inaceitável", "risco elevado", "risco sistêmico", "risco limitado" e "risco mínimo".

A regulação baseada em riscos apresenta uma série de problemas a priori, sobretudo quando adotada como estrutura normativa. Esse modelo regulatório consiste basicamente em negociar direitos humanos e fundamentais em favor da exploração de negócios privados, privilegiando interesses empresariais em detrimento da proteção de direitos em face dos comprovados danos gerados pelo uso da tecnologia. Como destacado pelo Access Now (2023), o enquadramento do problema em termos de risco postula a ideia de que a introdução de salvaguardas ou diretrizes éticas é capaz mitigar os danos, quando, em verdade, certos usos da IA são fundamentalmente incompatíveis com o gozo e a proteção de direitos humanos (2023). Especialmente para esses casos, mesmo uma classificação de risco elevado é inadequada para o perigo representado pelo uso, não havendo requisições e obrigações de governança, supervisão humana e transparência capazes de garantir a proteção de direitos.

Essa incongruência entre o modelo regulatório adotado e a proteção de direitos humanos e fundamentais fica evidente quando se observa o tratamento dado pelo regulamento às tecnologias de reconhecimento facial. De modo geral, no que diz respeito ao uso de IA em atividades de segurança pública, o regulamento aparenta estabelecer restrições. O texto classifica o uso de IA por autoridades responsáveis pela aplicação da lei como nível elevado de risco [Anexo III (6)], sujeitando-os a obrigações específicas (Artigo 8.º) e a um sistema de gestão de riscos mais rígido (Artigos 9.º). Os requisitos aplicáveis aos sistemas de IA de risco elevado incluem critérios de qualidade para conjuntos de dados de treinamento, validação e teste (Artigo 10.º); documentação técnica atualizada (Artigo 11.º); rastreabilidade (Art. 12.º); transparência e prestação de informações aos responsáveis pela implementação (Artigo 13.º); supervisão humana (Artigo 14.º); exatidão, solidez e cibersegurança (Artigo 14.º).

No entanto, em que pese a categorização como risco elevado represente a imposição de mecanismos de controle mais rigorosos [(1)], o texto estabelece exceções que acabam funcionando como autorização para uso pouco criterioso em face dos potenciais danos. A exceção da aplicabilidade do regulamento ao uso de sistemas de IA para fins de segurança nacional é um ponto crítico nesse sentido, visto que, como destaca Cabrera (2024), pode permitir sob esse pretexto a utilização de tecnologias que, de outro modo, seriam consideradas ilegais pelo próprio regulamento. Isso porque a am-

(1) A classificação de um sistema de IA como "risco elevado" ativa obrigações mais rigorosas de governança e supervisão, como (i) avaliação de impacto sobre direitos fundamentais sempre que usado por órgãos públicos (como, por exemplo, as polícias); (ii) registro no banco de dados da União Europeia (Artigo 49.º); (iii) adoção de medidas técnicas e organizacionais (Artigo 26.º) e (iii) interrupção imediata do uso se houver riscos à saúde, segurança ou direitos fundamentais (Artigo 79.º). No entanto, apesar do estabelecimento de salvaguardas, identifica-se limites sérios relacionados à transparência, que serão abordados mais adiante no texto.

plitude das definições de "aplicação da lei" e de "autoridades responsáveis pela aplicação da lei" [Artigo 3.º(45 e 46)] abre margem para sobreposições interpretativas com atividades de segurança nacional. Uma autoridade responsável pela aplicação da lei poderia tentar evitar as restrições impostas pelo regulamento à identificação biométrica à distância insinuando uma finalidade de segurança nacional, como a identificação de um suspeito de terrorismo em um grande evento, por exemplo (Cabrera, 2024).

Além do uso de IA por autoridades responsáveis pela aplicação da lei, o *AI Act* classifica como nível elevado de risco os sistemas de IA incluídos no domínio dos dados biométricos [Anexo III (1)], mencionando expressamente os sistemas de identificação biométrica à distância, não incluídos os sistemas de verificação biométrica com o único propósito de confirmar que uma determinada pessoa singular é a pessoa que alega ser [Anexo III (1, alínea a)]; os sistemas de categorização biométrica de acordo com atributos ou características sensíveis ou protegidos com base na inferência desses atributos ou características [Anexo III (1, alínea b)]; e os sistemas de reconhecimento de emoções [Anexo III (1, alínea c)].

De modo ainda mais específico, com relação ao reconhecimento facial em contextos de segurança pública, o regulamento da União Europeia proíbe expressamente "[a] utilização de sistemas de identificação biométrica à distância em 'tempo real' em espaços acessíveis ao público para efeitos de aplicação da lei" (Artigo 5.º, Alínea h), classificando o uso como nível inaceitável de risco. No entanto, em que pese a proibição dessa modalidade específica de reconhecimento facial possa representar algum tipo de avanço na salvaguarda de direitos, o regulamento cria diversas exceções que reduzem a força normativa da proibição.

Em primeiro lugar, o regulamento distingue entre sistemas de identificação biométrica à distância *em tempo real* e sistemas de identificação biométrica à distância *em diferido*. Enquanto a identificação em tempo real se enquadra no escopo da proibição, a identificação em diferido, não. Pela definição do regulamento, um sistema de identificação biométrica em tempo real é aquele em que a recolha de dados biométricos, a comparação e a identificação ocorrem *sem atraso significativo*, englobando a identificação instantânea e a identificação com ligeiro atraso [Artigo 3.º(42)]. Em contraposição, um sistema de identificação biométrica em diferido é aquele em que os procedimentos de comparação e identificação ocorrem *com atraso significativo* [Artigo 3.º(43)].

Embora o regulamento afirme que a identificação biométrica em diferido deva estar sujeita a salvaguardas e não deva "conduzir a uma vigilância indiscriminada" [Considerando (95)], ainda há riscos significativos de violação a direitos. Como destaca Cabrera (2025), o reconhecimento facial é mais amplamente utilizado para fins de aplicação da lei no contexto em

diferido, isto é, para identificação em investigações e outros cenários. Nesse sentido, na ausência de salvaguardas fortes, há danos à privacidade, risco de identificações errôneas e potencial para abuso (2025). Cabrera comenta que

> Embora isso seja explicitamente desencorajado pelo regulamento, seria fácil conceber um cenário em que as autoridades policiais tentariam contornar os requisitos mais rigorosos aplicáveis à identificação em tempo real, coletando imagens em protestos ou eventos públicos e adiando o exercício de comparação para se beneficiar das restrições menores aplicáveis à identificação em diferido. O efeito inibidor sobre a reunião e a expressão públicas permaneceria (Cabrera, 2025, tradução nossa).

Para Arnold (2024), em que pese a identificação em diferido fique sujeita a padrões regulatórios mais rigorosos, uma vez que entra na classificação de risco elevado, os padrões regulatórios exigidos não são suficientes para evitar práticas discriminatórias por parte das agências de aplicação da lei. Em primeiro lugar porque, embora um dos requisitos seja que os sistemas funcionem com nível suficiente de "precisão", a "precisão" de sistemas de identificação biométrica não é distribuída igualmente entre diferentes grupos demográficos (Arnold, 2024) – o que tem sido exaustivamente demonstrado por diversos estudos na última década, como argumentado no tópico anterior. A distribuição desigual da precisão dos sistemas biométricos, argumenta Arnold, favorece práticas racistas e aumenta os riscos de abuso de sistemas de IA para perpetração de práticas autoritárias como silenciamento de dissidentes e perseguição de oponentes, bem como para perseguição religiosa, étnica ou racial (2024).

Além disso, Arnold salienta que, embora o regulamento exija a implementação de práticas de gestão de risco em termos de possíveis "vieses" aos sistemas de risco elevado, o *AI Act* autoriza que um determinado sistema seja implementado mesmo que tais riscos persistam, desde que haja supervisão humana adequada que os minimize [Artigo 14.º(1 e 2)]. Em outras palavras, sistemas de IA discriminatórios podem ser implementados desde que tenham supervisão humana, visto que o regulamento não exige que um sistema de IA seja imparcial, nem que a supervisão humana elimine totalmente a discriminação, apenas que a minimize (Arnold, 2024).

Outro ponto destacado por Arnold é que, apesar de exigir que *outputs* de sistemas de identificação biométrica em diferido sejam verificados de forma independente por dois humanos antes que uma decisão final seja tomada, o regulamento prevê uma exceção a essa exigência em situações de aplicação da lei, migração, controle de fronteiras ou asilo, nos casos em que o requisito for julgado desproporcional pela lei nacional ou pela lei da União Europeia [Artigo 14.º(5)]. Na prática, essa exceção implica a possibilidade de que a União Europeia e seus Estados-Membros forneçam isenção geral do requisito para suas agências de aplicação da lei e migração, visto que a

cláusula de desproporcionalidade não está vinculada a nenhum critério específico (Arnold, 2024).

Ademais, Arnold enfatiza que os requisitos de transparência previstos pelo regulamento são limitados, especialmente para as pessoas afetadas pelos sistemas, pois embora o texto estabeleça obrigações para os provedores e implantadores de IA, dispensa explicitamente as autoridades policiais de informar quando utilizam a IA em atividades criminais. Sendo assim, os indivíduos ficam limitados em seu direito à contestação de possíveis saídas discriminatórias, comprometendo o direito à defesa e à justiça (Arnold, 2024).

Em segundo lugar, outro ponto crítico refere-se ao alcance da proibição do uso de sistemas de IA para identificação biométrica em tempo real, visto que se aplica somente a *espaços acessíveis ao público*, excluindo locais como prisões e fronteiras, onde circulam pessoas em situação de vulnerabilidade. Sendo assim, o regulamento ajuda a consolidar a inserção da União Europeia na tendência global de ampliação do controle digitalizado dos espaços fronteiriços, tal como abordada por Amaral *et al* (2025), uma realidade patente em diversos Estados-Membros.

Em terceiro lugar, por fim, o regulamento prevê uma série de exceções à proibição da IA para identificação biométrica em tempo real. Essas exceções incluem: (i) a busca de pessoas desaparecidas ou vítimas específicas de rapto, tráfico ou exploração sexual (Artigo 5.º(1), Alínea h, item i); (ii) a prevenção de ameaça específica, substancial e iminente à vida ou à segurança física de pessoas singulares ou ameaça real e atual ou real e previsível de ataque terrorista (Artigo 5.º(1), Alínea h, item ii); e (iii) a localização ou identificação de pessoa suspeita de cometer infração penal referida no anexo II do regulamento e punível no Estado-Membro em causa com pena ou medida de segurança privativa de liberdade de duração máxima não inferior a quatro anos (Artigo 5.º(1), Alínea h, item iii). Em qualquer um dos casos elencados, a proibição de uso pode ser suspensa, desclassificando a identificação biométrica para risco elevado e sujeitando-a a salvaguardas, como a exigência de avaliação de impacto sobre direitos fundamentais prevista no Artigo 27.º e o registro na base de dados da União Europeia previsto no Artigo 49.º [Artigo 5.º(2)].

3. Análise do relatório final da Comissão Temporária Interna do Senado Federal Brasileiro sobre Inteligência Artificial no Brasil

O Relatório Final da Comissão Temporária Interna sobre Inteligência Artificial no Brasil – CTIA (Brasil, 2024) foi aprovado pelo Senado Federal em dezembro de 2024. O texto aprovado é resultado do trabalho da comissão formada com a finalidade de examinar os projetos que visam disciplinar a

matéria no Brasil. O documento encontra-se em fase de debates na Câmara dos Deputados, de onde, após votação, deverá seguir para sanção presidencial [(2)]. A proposta legislativa em discussão, o Projeto de Lei n.º 2.338 de 2023 (PL 2.338/23), foi elaborada entre 2022 e 2023 pela Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil – CJSUBIA (Brasil, 2022), que, por sua vez, fora instaurada para instruir a apreciação dos projetos que versavam sobre a matéria na Câmara dos Deputados, em especial o Projeto de Lei n.º 21 de 2020 (PL 21/20) (Brasil, 2020), aprovado em regime de urgência em 2021 [(3)].

Seguindo a linha do Regulamento da União Europeia, o texto do PL 2.338/23 adota um modelo de regulação baseado em riscos, classificando-os em "risco excessivo", "alto risco" e "risco moderado/baixo", e estabelecendo mecanismos de governança díspares para os diferentes níveis de risco oferecidos. Apesar de apresentar uma inovação regulatória ao adotar uma "abordagem baseada em direitos" (Brasil, 2022 e 2024), o PL 2338/23 mantém a estrutura regulatória inspirada do Regulamento da União Europeia, que privilegia interesses econômicos através da modulação dos riscos de violação a direitos humanos e fundamentais. Isso pois, como destacado no tópico anterior, o enquadramento do problema em termos de risco é, a priori, incompatível com o gozo e a proteção de direitos, sobretudo nos casos em que, mesmo a classificação de alto risco apresenta-se inadequada em face do perigo representado pelo uso.

De modo similar ao *AI Act*, o texto do PL 2.338/23 veda expressamente a implementação de uso de "sistemas de identificação biométrica à distância, em tempo real e em espaços acessíveis ao público" (Artigo 13, inciso VII). Nesse aspecto, muitas das críticas endereçadas ao regulamento europeu podem ser feitas ao texto atual do PL 2.338/23. Em primeiro lugar, a proibição do uso de identificação biométrica à distância *em tempo real* exclui a identificação biométrica *em diferido*, costumeiramente utilizada por forças policiais para identificação de indivíduos em investigações, por exemplo (Cabrera, 2025).

---

(2) Em que pese a regra na tramitação bicameral no Brasil seja a aprovação na Câmara dos Deputados para posterior aprovação no Senado Federal, no caso em específico do PL 2338/23 o Senado figura como Casa Iniciadora do projeto, produto dos trabalhos da Comissão de Juristas Responsável por Subsidiar Elaboração de Substitutivo sobre Inteligência Artificial no Brasil (CJSUBIA), sendo a Câmara a Casa Revisora (Congresso Nacional, 2023).

(3) Aprovado sem passar por instâncias mínimas de participação social, o PL 21/20 foi fortemente criticado por organizações da sociedade civil, juristas e representantes do setor acadêmico por apresentar uma abordagem meramente principiológica, sem a previsão de mecanismos de governança e responsabilização (Coalizão Direitos na Rede, 2021a e 2021b). A esse ensejo, foi formada a CJSUBIA, cujo objetivo era o de estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil (Brasil, 2022). A comissão de juristas apresentou-se como arena de disputa política e discursiva em torno da IA e dos fenômenos relacionados ao racismo algorítmico (Silva, 2025), cujo resultado é o PL 2338/23 (Brasil, 2022). Posteriormente, o PL 2.338/23 passou pela apreciação de uma comissão temporária interna do Senado Federal que produziu o texto que se discute atualmente na Câmara dos Deputados (Brasil, 2024).

Nesse sentido, é importante destacar que, diferentemente do *AI Act*, que classifica de forma inequívoca o uso de IA pelas autoridades responsáveis pela aplicação da lei como nível elevado de risco, a redação atual do PL 2.338/23 não deixa clara a atribuição de alto risco ao uso de sistemas biométricos para atividades de investigação criminal e segurança pública. A redação do Artigo 14, inciso XI incorporou em parte a Emenda n.º 3 ao PL 2.338/23 apresentada pelo Senador Carlos Viana, que pretendeu restringir a classificação de alto risco para sistemas biométricos somente aos sistemas utilizados com fins de investigação criminal e segurança pública (Brasil, 2024). O argumento postulado na ementa é que o uso de sistemas biométricos é benéfico em contextos de autenticação e segurança contra fraudes, e que não se justificaria, nesses casos, a classificação de nível alto de risco. A redação do texto legislativo proposto, no entanto, não é congruente com o objetivo da ementa. O texto do PL, conforme redigido no Relatório Final da CTIA, limita a classificação de alto risco somente a sistemas de identificação e autenticação biométrica *para o reconhecimento de emoções* (Artigo 14, inciso X), deixando de fazer qualquer menção à categorização de alto risco para sistemas biométricos em geral no escopo das atividades de investigação criminal e de segurança pública. No cenário posto por essa redação, o PL 2.338/23 excluiria o uso de sistemas biométricos em investigações criminais e na segurança pública da categoria de alto risco, hipótese em que não seriam aplicados a esses usos os mecanismos de governança mais rigorosos previstos no Artigo 18 [4].

(4) Aos sistemas classificados como alto risco, o texto determina a aplicação de medidas de governança gerais (Artigo 17) e específicas (Artigo 18) como requisitos de uso, apresentando mecanismos de controle mais rigorosos como: a indicação de um encarregado de governança (Artigo 18, inciso I), documentação a respeito do funcionamento do sistema e das decisões envolvidas em sua construção, implementação de uso (Artigo 18, inciso II), uso de ferramentas de registro automático da operação do sistema, a fim de permitir a avaliação da sua acurácia e robustez, a apuração de seus potenciais resultados discriminatórios e a implementação das medidas de mitigação de riscos adotadas (Artigo 18, inciso III), realização de testes para avaliação de níveis adequados de confiabilidade, performance, segurança, proteção e robustez conforme o setor e o tipo de aplicação do sistema (Artigo 18, inciso IV), utilização de dados de treinamento, validação e teste adequados, representativos, livres de erros e completos, contendo propriedades estatísticas apropriadas em relação às pessoas afetadas e levando em conta características e elementos específicos do contexto geográfico, comportamental ou funcional no qual o sistema de IA de alto risco será utilizado (Artigo 18, inciso V), registro de fontes automatizadas e do grau de supervisão humana que tenham contribuído para os resultados apresentados pelos sistemas IA (Artigo 18, inciso VI), medidas de gestão de dados para mitigar e prevenir vieses discriminatórios e incentivar diversidade nas equipes de desenvolvimento, bem como políticas de gestão e governança para promoção da responsabilidade social e sustentável (Artigo 18, inciso VII), adoção de medidas técnicas para viabilizar a explicabilidade dos resultados dos sistemas de inteligência artificial e de medidas para disponibilizar informações adequadas que permitam a interpretação dos seus resultados e funcionamento, respeitado o sigilo industrial e comercial (Artigo 18, inciso VIII), adoção de procedimentos e mecanismos para notificação de incidentes e de mau funcionamento do sistema de inteligência artificial (Artigo 18, inciso IX). No entanto, apesar de impor requisitos de governança mais severos aos sistemas classificados como nível alto de risco, o texto delega às autoridades setoriais o poder de definir as hipóteses em que as obrigações estabelecidas no Artigo 18 e seus incisos serão flexibilizadas, observadas as normas gerais da autoridade competente (Artigo 18, Parágrafo Único). que acabam funcionando como autorização para uso pouco criterioso em face dos potenciais danos.

Em segundo lugar, de modo similar ao *AI Act*, o PL 2.338/23 restringe a proibição de uso de IA para identificação biométrica à distância em tempo real somente a *espaços acessíveis ao público*, permitindo a aplicação da tecnologia em locais sensíveis como prisões e fronteiras.

Em terceiro lugar, por fim, também de modo semelhante ao Regulamento da União Europeia, o texto brasileiro estabelece um importante rol de exceções que acaba funcionando como uma autorização ampla para a implementação de tecnologias de reconhecimento facial. As categorias de permissões incluem um amplo espectro de atividades de investigação criminal e segurança pública como a instrução de inquérito ou processo criminal mediante autorização judicial prévia e motivada, quando houver indícios razoáveis da autoria ou participação em infração penal, a prova não puder ser feita por outros meios disponíveis e o fato investigado não constitua infração penal de menor potencial ofensivo (Artigo 13, inciso VII, alínea a), busca de vítimas de crimes, de pessoas desaparecidas ou em circunstâncias que envolvam ameaça grave e iminente à vida ou à integridade física de pessoas naturais (Artigo 13, inciso VII, alínea b), flagrante delito de crimes punidos com pena privativa de liberdade máxima superior a dois anos, com imediata comunicação à autoridade judicial (Artigo 13, inciso VII, alínea c), e recaptura de réus evadidos, cumprimento de mandados de prisão e de medidas restritivas ordenadas pelo Poder Judiciário (Artigo 13, inciso VII, alínea d).

O parágrafo único do mesmo artigo estabelece como requisito ao uso desses sistemas que seja "proporcional e estritamente necessário ao atendimento do interesse público" (Artigo 13, Parágrafo Único), abrindo ampla margem interpretativa para o que seria considerado "interesse público" nesses casos. Nesse sentido, em que pese o texto do projeto faça menção expressa à observância ao devido processo legal e ao controle judicial, bem como aos princípios e direitos previstos na lei para a garantia contra a discriminação e à necessidade de revisão humana, tal exigência apresenta força normativa pouco eficaz para a proteção de direitos em face do complexo cenário posto.

4. CONSIDERAÇÕES FINAIS

A análise comparativa entre o Regulamento de Inteligência Artificial da União Europeia e o Projeto de Lei n.º 2.338/2023 em tramitação no Congresso Nacional brasileiro evidenciou a partilha de uma abordagem de regulação centrada na gradação de riscos, operando como instrumentos de subordinação de direitos à viabilização da exploração econômica da inteligência artificial. Tanto o Regulamento Europeu quanto o PL brasileiro exemplificam como a linguagem da regulação de riscos, ao modular a gestão técnica dos danos, produz um regime jurídico de aceitabilidade da

violação de direitos como mero efeito colateral a ser mitigado. Por outro lado, em ambos os casos, a multiplicidade de exceções e brechas interpretativas em torno do uso de sistemas biométricos em contextos de segurança evidencia uma tendência regulatória que autoriza amplamente a sua utilização, contribuindo para a expansão do novo regime de vigilância e controle securitário em curso.

Bibliografia

AccessNow. (2021, 17 de fevereiro). *The EU should regulate AI on the basis of rights, not risks.* https://www.accessnow.org/eu-regulation-ai-risk-based-approach/

Arnold, L. (2024). How the European Union's AI Act provides insufficient protection against police discrimination. *The Journal of Law & Social Change, 28*(2). http://www-law-upenn-edu.translate.goog/live/news/16742-how-the-european-unions-ai-act-provides

Bioni, B., Garrote, M., & Guedes, P. (2023). *Temas centrais na regulação de IA: O local, o regional e o global na busca da interoperabilidade regulatória.* São Paulo: Associação Data Privacy Brasil de Pesquisa.

Brasil. (2020). *Projeto de Lei nº 21, de 2020: Estabelece fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil.* Câmara dos Deputados. https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2236340

Brasil. (2022). *Relatório final da Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre inteligência artificial no Brasil.* Senado Federal. https://legis.senado.leg.br/sdleggetter/documento/download/bda-ad0dc-5c0a-4217-a6d0-aefb0d8ec8d4

Brasil. (2024). *Relatório final da Comissão Temporária Interna sobre Inteligência Artificial no Brasil.* Senado Federal. https://legis.senado.leg.br/sdleg-getter/documento/download/9652e2f1-baa4-4bbe-98a9-2ee4abcc64df

Cabrera, L. L. (2024, 30 de abril). *EU AI Act brief – Pt. 2, privacy & surveillance.* Center for Democracy & Technology. https://cdt.org/insights/eu-ai-act-brief-pt-2-privacy-surveillance/

Coalizão Direitos na Rede. (2021a, 7 de dezembro). *Brasil não está pronto para regular inteligência artificial.* Direitos na Rede. https://direitosnarede.org.br/2021/12/07/brasil-nao-esta-pronto-para-regular-inteligencia-artificial/

Coalizão Direitos na Rede. (2021b, 23 de setembro). *Inteligência artificial não pode ser regulada a toque de caixa.* Direitos na Rede. https://direitosnarede.org.br/2021/09/23/inteligencia-artificial-nao-pode-ser-regulada-a-toque-de-caixa/

Congresso Nacional. (2023). *Projeto de Lei nº 2338, de 2023: Dispõe sobre o uso da inteligência artificial (Matérias Bicamerais).* https://www.congressonacional.leg.br/materias/materias-bicamerais/-/ver/pl-2338-2023

Parlamento Europeu. (2024). *Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho da União Europeia que cria regras harmonizadas em matéria de inteligência artificial.* Luxemburgo.

Silva, T. (2025). *Racismo algorítmico e regulação de inteligência artificial: O contrato racial na produção do PL 2338/2023* [Tese de doutorado, Universidade Federal do ABC]. https://tarciziosilva.com.br/blog/

O SISTEMA EUROPEU DE RECONHECIMENTO FACIAL: O DISPOSITIVO BANÓPTICO E A NOVA ROUPAGEM DA ANTROPOMETRIA CRIMINAL

Por Augusto Jobim do Amaral (*) *y Gabriel Saad Travassos* (**)

Resumo: A Europa estabeleceu, recentemente, um novo regime para a captura, o armazenamento e o compartilhamento de dados de imagens faciais, estruturado a partir do Regulamento Prum II. O artigo tem como problema de pesquisa analisar como este regime remasteriza estratégias do positivismo criminológico, desenvolvendo uma governamentalidade banóptica. A partir da pesquisa qualitativa, avaliamos a hipótese de que o sistema de vigilância europeu investe no policiamento dos corpos em um regime de inquietação e mal-estar permanente próprio de um novo tipo de governamentalidade. Com base na revisão bibliográfica, na pesquisa documental e no estudo de caso, analisamos este novo regime tecnopolítico criminal.

Palavras-chave: Sistema europeu de reconhecimento facial. Governamentalidade. Dispositivo Banóptico. Antropometria criminal.

El sistema europeo de reconocimiento facial: el dispositivo banóptico y la nueva forma de la antropometría criminal

Resumen: Europa estableció recientemente un nuevo régimen para la captura, almacenamiento e intercambio de datos de imágenes faciales, estructurado en torno al Reglamento Prum II. Este artículo analiza cómo este régimen reinterpreta las estrategias del positivismo criminológico, desarrollando una gubernamentalidad banóptica. Basados en una investigación cualitativa, evaluamos la hipótesis de que el sistema de vigilancia europeo invierte en cuerpos policiales dentro de un régimen de inquietud e incomodidad permanentes, característico de un nuevo tipo de gubernamentalidad. A partir de una revisión bibliográfica, investigación documental y un estudio de caso, analizamos este nuevo régimen tecnopolítico criminal.

(*) Professor do Programa de Pós-Graduação em Ciências Criminais e do Programa de Pós-Graduação em Filosofia, ambos da PUC-RS. augusto.amaral@pucrs.br.

(**) Doutorando em Ciências Criminais (PUC-RS). Mestre em Direito e Justiça Social (FURG). Defensor Público Federal. travassosgabrielsaad@gmail.com.

Palabras clave: Sistema europeo de reconocimiento facial. Gubernamentalidad. Dispositivo banóptico. Antropometría criminal.

The european facial recognition system: the banoptic device and the new guise of criminal anthropometryo

Abstract: Europe recently established a new regime for the capture, storage and sharing of facial image data, based on the Prum II Regulation. The article's research problem is to analyze how this regime remasters strategies of criminological positivism, developing a banoptic governmentality. Based on qualitative research, we evaluate the hypothesis that the European surveillance system invests in policing on the bodies in a regime of permanent unrest and unease typical of a new type of governmentality. Based on the literature review, documentary research and case study, we analyze this new criminal techno-political regime.

Keywords: European facial recognition system. Governmentality. Banoptic device. Criminal Anthropometry.

1. INTRODUÇÃO

Estima-se que o mercado de tecnologias de reconhecimento facial movimentará 8,5 bilhões de dólares em 2025. Pesquisa conduzida por Paul Bischoff (2022) aponta que sete em cada dez países já se utilizam de tecnologias de reconhecimento facial em larga escala, com usos em espaços variados como aeroportos, estações de metrô, instituições bancárias e até mesmo em escolas.

A Agência da União Europeia para a Cooperação Policial (Europol) desenvolve atualmente um sistema europeu de base de dados de reconhecimento facial (EUROPEAN COMISSION, 2023), agora normatizado pelo Regulamento n. 2024/982, do Parlamento Europeu e do Conselho, de 13 de março de 2024, conhecido como Regulamento de Prum II.

O artigo tem como problema de pesquisa analisar em qual medida o novo regime está inserido em uma linha de continuidade própria da modernidade e do positivismo criminológico, estruturando a governamentalidade a partir do dispositivo banóptico como método de gestão diferenciada da circulação.

Na primeira seção apresentamos o novo regime de controle no espaço europeu e as suas ligações com o positivismo criminológico e a biopolítica. Em seguida, analisamos em que medida esse regime se aproxima do conceito de dispositivo banóptico de que trata Didier Bigo.

Compreender esse dispositivo no cerne do policiamento global pautado pelo excepcionalismo e contenção de migrantes permitirá entender como a

discriminação sistêmica é funcional à produção da liberdade no neoliberalismo. Assim, as violações de direitos humanos são assumidas como danos colaterais em prol de uma ideia de liberdade que é ativamente produzida como técnica política e forma de subjetividade.

A partir da pesquisa qualitativa, avaliamos a hipótese de que o sistema de vigilância europeu normatiza um dispositivo banóptico que produz a liberdade a partir da discriminação algorítmica que, nesse diagrama, revela-se inerente ao modelo de policiamento dos corpos no espaço europeu.

Com base na revisão bibliográfica, na pesquisa documental e no estudo de caso, analisamos o novo regime no contexto da política criminal de excepcionalismos.

2. O SISTEMA EUROPEU DE RECONHECIMENTO FACIAL

O sistema cruzado de dados de biometria facial atualmente em operação na Europa busca intercambiar dados armazenados de indivíduos submetidos à extração dos dados biométricos da face em qualquer dos países da União Europeia e, ainda, com os Estados Unidos. O fluxo de dados entre esses países não se limita à esfera criminal, alcançando também sistema de controle migratório.

O marco normativo sobre o intercâmbio de dados biométricos para políticas públicas ligadas ao sistema criminal ou ao sistema migratório é o Tratado de Prum, subscrito em 27 de maio de 2005, na Alemanha, e incorporado ao quadro normativo da União Europeia em 15 de fevereiro de 2007 (PARLAMENTO EUROPEU, 2007).

Referido tratado estabelece normas de cooperação entre os países subscritores em matéria de terrorismo, criminalidade transfronteiriça e imigração (PRUM CONVENTION, 2005). Em seu texto está disciplinado o intercâmbio de informações sobre DNA, impressões digitais, registro de veículos, dados pessoais e não pessoais para a cooperação policial entre as partes contratantes.

Recentemente, o Tratado de Prum sofreu novas alterações com o Regulamento n. 2024/982, do Parlamento Europeu e do Conselho, de 13 de março de 2024. Identificado como "Regulamento Prum II", o documento tratou das condições e procedimentos para a consulta e o intercâmbio automatizados de perfis de DNA, dados datiloscópicos, registro de veículos, imagens faciais e ficheiros policiais (UNIÃO EUROPEIA, 2024). O Regulamento definiu um novo regime para o intercâmbio de informações entre as autoridades responsáveis pela prevenção, detecção ou investigação de infrações penais.

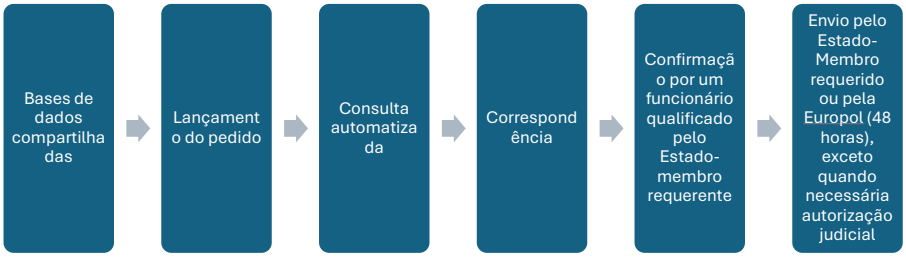
Nesse novo regime, as imagens faciais são objeto de consultas automatizadas para fins de prevenção, detecção ou investigação de uma infração penal desde que a pena aplicada ao delito seja igual ou superior a 01 (um) ano. Para os demais dados biométricos não existe um patamar mínimo de pena, ou seja, qualquer delito é passível de investigação mediante a utilização da metodologia comparada entre esses dados.

A base de dados da Europol, estruturada a partir do Regulamento EU 2016/794, contém vasta gama de dados biométricos enviados por terceiros países, inclusive informações provenientes de zonas de guerra (UNIÃO EUROPEIA, 2016). O Regulamento de Prum II armazena todos esses dados nessa gigantesca base e distribui a todos os países da União Europeia.

Muito embora o Regulamento 2016/794 vede que a Europol compartilhe dados pessoais das suas bases, admite que o órgão receba esses dados de países terceiros, inclusive de organismos privados. Não há uma delimitação sobre as condições materiais para o compartilhamento dos dados de países terceiros, e a avaliação da fiabilidade é feita pelo ente que forneceu a informação. Não há no regulamento limites aos tipos de dados que podem ser recebidos pela Europol, como por exemplo no caso de um indivíduo que tenha seus dados compartilhados mesmo sem qualquer condenação contra si.

Se na consulta automatizada houver uma correspondência, cabe ao Estado requerente realizar a confirmação entre as duas imagens faciais. Após a confirmação, deve informar ao Estado requerido e assegurar que pelo menos um funcionário qualificado efetue a verificação manual da correspondência.

Em termos esquemático, o procedimento de compartilhamento de imagens faciais pode ser assim sintetizado:



Fonte: elaboração própria

Percebe-se, nesse fluxo, um modelo de policiamento que investe na captura em massa de dados pessoais, sistematização e análise algorítmica das faces como sinais de identificação de risco.

3. O dispositivo banóptico e a nova roupagem da antropometria criminal

O fascínio pela categorização e armazenamento de imagens corpos identificados como de risco não é novo; tem bases no positivismo criminológico, na frenologia e na antropometria. Em 1886, Thomas Byrnes, então chefe do Departamento de Polícia de Nova Iorque, criou um álbum intitulado *Professional Criminals of America,* com 204 fotografias frontais de indivíduos identificados como criminosos (BYRNES, 2019).

Também nesse período do final do século XIX, Alphonse Bertillon (1893) estabeleceu um sistema de classificação fotográfico que padroniza o registro, o armazenamento e a troca de fichas criminais com a face dos suspeitos. O sistema se concentrava em onze medidas antropométricas como altura, comprimento e largura da cabeça, comprimento dos braços, pés, orelhas e alguns dedos. O modelo foi objeto de efusivos elogios por Cesare Lombroso (1897), que fez do estudo do criminoso atávico a base da sua obra "O homem delinquente", representativa do positivismo criminológico.

Após a adoção oficial pela Polícia de Paris em 1883, o sistema foi implementado na Europa, Canada e Estados Unidos. O compartilhamento e o armazenamento de dados biométricos tornaram-se uma obsessão das autoridades policiais e a meta principal da Associação Internacional de Chefes de Polícia, de 1902, era o estabelecimento de uma base centralizada de recursos para catalogar e compartilhar informações sobre criminosos (FINN, 2009, p. 27).

No entanto, quanto às imagens faciais, um dos grandes problemas da época era a imprecisão das fotografias e a baixa capacidade de armazenamento. O processo de coleta dos diversos dados antropométricos também era exaustivo e contraproducente, o que foi decisivo para a gradual prevalência da biometria digital.

Atualmente, porém, com o desenvolvimento de tecnologias de reconhecimento facial que aliam a grande capacidade de armazenamento e processamento de dados com o aprendizado de máquina, o ideário positivista ganha novo fôlego, mas sob nova roupagem.

A gigantesca base de imagens faciais organizada e operacionalizada pelo Regulamento de Prum II evidencia uma política de controle e vigilância de um campo emergente de governamentalidade que, segundo Didier Bigo (2022), é estabelecida a partir de um dispositivo do tipo *banóptico.*

Governamentalidade é um conceito foucaultiano que, dentre outros aspectos [1], caracteriza-se pelo conjunto de instituições, procedimentos, táticas, cálculos e análises que têm como alvo a população, cuja base epistêmica é a economia política e cujo instrumento técnico são os dispositivos de segurança. Assim, quando Bigo alude à governamentalidade do mal-estar [2] desloca o eixo de análise para o tema central do policiamento global transnacional: governar condutas a partir do binômio ameaça e angústia.

Essa governamentalidade está no cerne do sistema europeu e se caracteriza a partir de três critérios: práticas de excepcionalismo, atos de perfilamento e contenção de estrangeiros, e um imperativo de mobilidade (BIGO, 2022). O discurso que a sustenta está fundado em estatísticas, tecnologias de biometria, classificação e priorização de ameaças.

Os profissionais da gestão do mal-estar produzem o poder-conhecimento sobre o dualismo segurança e insegurança a partir de um regime transnacional de verdade que sucede as clássicas razões de Estado e é acentuado na arena europeia. É nesse regime que o discurso mescla guerra e militarização, terrorismo e ameaças internas, política criminal e migratória a partir de burocracias e agências que se apropriam de uma lógica de securitarização transfronteiriça.

Esse programa de (in)segurança performa como uma estratégia programática de escalada generalizada de vigilância tão globalizada quanto individualizada o quanto possível. É nesse ponto que Bigo se utiliza do conceito foucaultiano de dispositivo para reinterpretá-lo no panorama atual como um *Ban-opticon*.

Em *Vigiar e Punir* (1987), Foucault analisa que o poder supõe um dispositivo que congregue técnicas que maximizem seus efeitos. Uma arquitetura ou infraestrutura que automatiza e desindividualiza o poder: "há uma maquinaria que assegura a dissimetria, o desequilíbrio, a diferença. Pouco importa, consequentemente, quem exerce o poder" (FOUCAULT, 1987, p. 167).

---

(1) Governamentalidade, em Foucault, não se esgota nesse aspecto. Engloba também uma linha de força, uma tendência a um tipo de poder de governo sobre todos os demais a partir desses aparatos específicos e de uma série de saberes. Há também um sentido de governamentalidade como resultado do processo de transformação do Estado de Justiça da Idade Média para o Estado Administrativo e, finalmente, para a governamentalização deste último. FOUCAULT, Michel. *Segurança, território, população*: curso dado no Collège de France (1977-1978). Trad.: Eduardo Brandão. São Paulo: Martins Fontes, 2008, p. 143-144.

(2) O termo usado pelo autor no texto original é *governmentality of unease.* Apesar de algumas obras optarem pela tradução "governamentalidade do desconforto", buscamos uma expressão que alcançasse de modo mais amplo o sentido proposto por Bigo sobre os fundamentos, as características e os efeitos desse modelo de governamentalidade que se estrutura a partir de práticas de exceção e banimento. *Unease,* nesse contexto, poderia ser mais associado ao mal-estar, à profunda inquietação, como uma dimensão mais ampla do impacto de atos de perfilamento e contenção de estrangeiros (extranćus) como formas de biopolíticas. BIGO, Didier. Globalized (in)security: The field and the ban-opticon. In: *Terror, Insecurity and Liberty*: Illiberal Practices of Liberal Regimes After 9/11. Ed.: Bigo D. and Tsoukala A. New York: Routledge, p. 10-48.

Nesse contexto, o panóptico, modelo de vigilância total atribuído a Jeremy Bentham em que a pessoa vigiada constantemente não sabe quando e quem a visualiza, produz uma sujeição automática ao poder. O panóptico estabelece uma vigilância pautada no princípio de que o poder deve ser inverificável: o vigiado é o objeto da informação, jamais sujeito da comunicação; não deve nunca saber se está sendo observado, mas ter a certeza de sempre poder sê-lo.

A partir dessa relação que se estabelece entre poder e técnica, Bigo (2022) avalia como a sociedade atual convive com uma rede transfronteiriça de práticas heterogêneas e transversais que produzem o discurso e a (in) segurança. O autor se recusa a creditar tal cenário a um tipo de dominação imperial centralizada, apontando um campo fragmentado, heterogêneo e descentralizado nos processos políticos de conflito e convergência entre grandes corporações e agências burocráticas.

Tendo como marco referencial o 11 de setembro de 2001, Bigo (2008) analisa como se inaugura a partir dali um modelo de (in)segurança globalizada que mescla dois campos até então distintos: o universo da guerra, defesa nacional, ordem internacional e estratégia e o universo da segurança interna e da ordem pública policial. O controle e a resposta ao delito assumem uma interface militarizada; são mediados por agências burocráticas e gestores do mal-estar provocado pela frequente ameaça do estrangeiro - na concepção etimológica do termo como *extranĕus* - e, por isso, um risco à segurança nacional.

O gerenciamento dessa engrenagem é cercado por práticas heterogêneas e transnacionais, coordenadas por agências a nível transnacional: autoridades alfandegárias, controles migratórios, políticas criminais, agências anti-terroristas, agentes infiltrados, tecnologias de controle e gestão à distância. Esse diagrama não se apresenta como um clássico panóptico a nível global, mas sim como um *Ban-opticon,* uma associação entre o termo "ban" de Jean Luc Nancy, refigurado por Agamben, com o *opticon* utilizado por Foucault:

> Essa formulação do *Ban-opticon* nos permite entender como uma rede de práticas heterogêneas e transversais funciona e faz sentido enquanto forma de (in)segurança ao nível transnacional. Permite-nos analisar a coleção de corpos heterogêneos de discursos (sobre ameaças, imigração, inimigo interno, quinta coluna imigrante, muçulmanos radicais versus bons muçulmanos, exclusão versus integração etc) de instituições (agências públicas, governos, organizações internacionais, ONGs, etc), de estruturas arquitetônicas (centros de detenção, zonas de espera e as pistas de trânsito em aeroportos da área de Schengen, redes integradas de videocâmera em algumas cidades, redes eletrônicas munidas de instalações de segurança e vídeo-vigilância), de leis (...) e de medidas administrativas (...) (BIGO, 2022, p. 144).

O radical *ban,* de origem germânica, insere uma ideia de rejeição, exclusão, restrição, proscrição. Ban-dido e ban-ido são palavras utilizadas, respectivamente, para provocar as ideias de contrário à ordem e excluído dessa ordem. O banido não está apenas fora da lei, mas abandonado por ela.

No Banóptico a vigilância de todos não é um objetivo central, mas sim de grupos identificados como de risco no campo da mobilidade transnacional. Nesse dispositivo são permanentes os excepcionalismos do poder, manifestados nas leis de emergências, e o perfilamento de grupos indesejáveis a partir de práticas que sobrebujam a nação-Estado.

O discurso não é mais sobre a reforma do indivíduo, mas sobre o controle de grupos específicos que são vistos como uma ameaça ao solo europeu. Para lidar com essas classificações *ban,* as estruturas arquitetônicas não dependem do confinamento; são aplicadas no intenso fluxo de circulação para incidência sobre indivíduos específicos; as leis e medidas administrativas são adaptadas a conceitos genéricos que substituem a ideia de dano pela ideia de risco, associada a uma determinada imagem, aquela capturada pela tecnologia algorítmica de reconhecimento facial.

Estabelece-se um modelo de vigilância que é distribuída e imanente: não está em um ponto fixo tampouco restrita a um espaço de confinamento (DELEUZE, 1992). Investe-se no controle da circulação que é essencial ao produto *liberdade* que o neoliberalismo anuncia (RODRÍGUEZ, 2009). O fluxo de pessoas e mercadorias – e, por consequência dados – é vital para a validação do produto. A gestão desse fluxo dependerá do cálculo de risco sobre grupos específicos (AMARAL, 2024, p. 39).

Os dispositivos banópticos garantem a aparente circulação de todos a partir da exclusão de alguns, associados a partir dos perfis de risco, embasados na suposta neutralidade tecnológica (O'NEIL, 2020). Legitimam a continuidade repressiva do modelo penal, com sua clientela habitual e recorte racial, ao mesmo tempo que formam outros grupos identificados como de risco para o manejo de táticas de vigilância e controle da circulação.

Desse modo, o dispositivo já não é um panóptico benthamiano, mas um banóptico que não depende da imobilização dos corpos nos espaços de confinamento, mas sim da produção virtual ou algorítmica de perfis que sinalizam a diferença (BIGO, 2022). O sistema europeu de identificação pessoal investe nessa marca da diferença, nessa tatuagem imagética produzida a partir das lentes focais sobre os corpos em movimento, garantindo uma nova roupagem robótica e digital para a velha antropometria criminal (BRUNO, 2013, p. 33).

Ao contrário do que se possa acreditar em primeiro plano, como visto, essas discriminações sistêmicas, ligadas ao processo algorítmico que subjaz as tecnologias de reconhecimento facial, não são uma distorção da arte

de governo (neo)liberal, mas a sua própria essência, com raízes no modelo policial que, desde o positivismo criminológico, buscou categorizar os corpos com bases em evolucionismos de natureza eugênica.

Ao invés de uma tecnologia disruptiva, o banco de imagens faciais dobra a aposta no sonho de antropometria criminal da Associação Internacional de Chefes de Polícia, de 1902, agora a partir de modelos algorítmicos e tecnologias que se anunciam inovadoras, mas estão estruturadas em bases ideológicas do início do século XX.

4. Conclusão

O artigo analisou o sistema europeu de reconhecimento facial, estruturado a partir do Regulamento de Prum II, como um dispositivo banóptico próprio da governamentalidade do mal-estar. Observou-se uma linha de continuidade com a antropometria criminal e o positivismo criminológico, discursos que produziram discriminação a partir da suposta neutralidade do saber científico.

Essa mesma neutralidade agora é levantada para defender o emprego das tecnologias algorítmicas de reconhecimento facial para fins de policiamento e controle migratório. Nesse contexto, o regime europeu de captura e análise de imagens faciais se apresenta como um dispositivo banóptico de instrumentalização de um tipo de governamentalidade que estimula a circulação ao mesmo que define zonas de exclusão.

O artigo aplicou conceitos foucaultianos para avaliar uma linha de continuidade entre discursos e práticas do neoliberalismo na gestão decomponível dos indivíduos em dados. O fascínio pelo registro e estoque dos corpos suspeitos não é novo; acompanha a história da antropometria e da criminologia positivista.

Essa ideologia assume novas formas com o desenvolvimento de novas tecnologias com capacidade inédita de coleta, processamento e armazenamento. O dispositivo banóptico está no cerne desse tipo de policiamento global, que investe em práticas de excepcionalismo e na contenção daqueles grupos identificados como perigosos.

Nesse contexto, a liberdade é produzida para aquele nível pertinente da população, ainda que por meio de dispositivos securitários que reproduzam discriminações sistêmicas contra o nível que, utilizando a terminologia foucaultiana, é visto como não-pertinente.

Conforme literatura especializada na matéria, as tecnologias de reconhecimento facial ainda apresentam pontos sensíveis de discriminação sistêmica que reproduzem modelos racistas, sexistas e xenofóbicos. O maior

espanto é que esses modelos não são distorções do sistema, mas a própria lógica operativa da governamentalidade e dos dispositivos securitários que a operacionalizam.

O argumento securitário já foi utilizado por regimes totalitários para valer-se de modelos de gestão de população baseados em controle e punição. Portanto, ao tempo em que a tecnologia de reconhecimento facial promete ser uma panaceia para a ideologia de defesa social e segurança nacional, existem riscos evidentes ao legado de direitos fundamentais e de proteção da pessoa humana decorrente desse uso amplo e intensivo.

REFERÊNCIAS

Amaral, Augusto Jobim do; DIAS, Felipe da Veiga. (2024). *Tecnopolítica criminal*. 1ª Ed., São Paulo: Tirant lo Blanc.

Bertillon, Alphonse. (1893). *Identification anthropométrique*: Instructions signalétiques. Melun: Imrpimerie Administrative.

Bigo, Didier. (2008). Globalized (in)security: The field and the ban-opticon. In: *Terror, Insecurity and Liberty:* Illiberal Practices of Liberal Regimes After 9/11. Ed.: Bigo D. and Tsoukala A. New York: Routledge, p. 10-48.

Bigo, Didier. (2022). (In)segurança globalizada: o campo e o ban-opticon. In: *A cidade como máquina biopolítica*. Org.: Augusto Jobim do Amara *et al*. Tirant lo blanch, Valencia, p. 116-154.

Bischoff, Paul. Facial recognition technology (FRT): 100 countries analyzed. (2022). Comparitech, 24.Jan.2022. Available at: https://www.comparitech.com/blog/vpn-privacy/facial-recognition-statistics/#:~:text=Five%20countries. Access 25.02.2024.

Byrnes, Thomas. (2019). *Professional Criminals of America*. Lanham: Lyon Press.

Bruno, Fernanda. (2013). *Máquinas de ver, modos de ser*: vigilância, tecnologia e subjetividade. Porto Alegre, Sulina.

Council of the European Union. *Prum Convention.* (2005). Available at: https://data.consilium.europa.eu/doc/document/ST-10900-2005-INIT/en/pdf. Access at 20.06.2024.

Deleuze, Gilles. (1992). *Conversações*, 1972-1990. Trad.: Peter Pál Pelbart. Rio de Janeiro: Ed. 34.

European union. European commission. (2023). *The Commission welcomes the political agreement on automated data exchange for police cooperative*. Press release. Available at https://ec.europa.eu/commission/presscorner/detail/en/ip_23_5870. Acesso 17.06.2024.

Finn, Jonathan M. (2009). *Capturing the criminal image*: from mug shot to surveillance society. Minneapolis: University of Minnesota Press.

Foucault, Michel. (1987) *Vigiar e Punir*: nascimento da prisão. Trad.: Raquel Ramalhete, 20ª Ed., Petrópolis: Vozes.

Foucault, Michel. (2008). *Segurança, território, população*: curso dado no Collège de France (1977-1978). Trad.: Eduardo Brandão. São Paulo: Martins Fontes.

Lombroso, Cesare. (2013) *L'uomo delinquente*. 5ª Edizione (1897). Milão: Bompiani.

O'Neil, Cathy. (2020) *Algoritmos de destruição em massa*: como o Big Data aumenta a desigualdade e ameaça a democracia. Trad.: Rafael Abraham. 1ª Ed. Santo André: Rua do Sabão.

Parlamento Europeu. (2007) Comissão das Liberdades Cívicas, da Justiça e dos Assuntos Internos. *Documento de trabalho sobre o projeto de decisão do Conselho relativa ao aprofundamento da cooperação transfronteiras, em particular no domínio da luta contra o terrorismo e da criminalidade transfronteiras*. Relator Fausto Correia. Disponível em: https://www.europarl.europa.eu/meetdocs/2004_2009/documents/dt/660/660824/660824pt.pdf. Acesso em 20.06.2024.

Rodríguez, Pablo Manolo. (2009). *Las Palabras en Las Cosas*: saber, poder y subjectivación entre algoritmos y biomoléculas. Buenos Aires: Cactus.

União Europeia. (2016). *Regulamento EU 2016/796 do Parlamento Europeu e do Conselho de 11 de maio de 2016*. Cria a Agência da União Europeia para a Cooperação Policial (Europol). Jornal Oficial da União Europeia. 24.5.2016.

União Europeia. (2024). *Regulamento (UE) 2024/982 do Parlamento Europeu e do Conselho de 13 de março de 2024 relativo à consulta e ao intercâmbio automatizados de dados para efeitos de cooperação policial.* Disponível em http://data.europa.eu/eli/reg/2024/982/oj. Acesso em 20.06.2024.

Sección V

Crítica, tecnopolítica, gobernanza

El papel de la crítica en tiempos artificiales

Por Jefferson de Carvalho Gomes (*)

Resumen: El presente trabajo analiza el papel de la crítica en la era digital y de la inteligencia artificial, considerando la posmodernidad y la sociedad líquida descrita por Bauman. Se sostiene que la creciente automatización y el uso de algoritmos en la vida cotidiana y en la práctica jurídica amenazan la capacidad humana de pensar críticamente. La crítica, desde la perspectiva de Foucault, Gabriel, Lipovetsky y Han, se entiende como un acto de resistencia histórica, ética y ontológica que preserva la libertad y la humanidad frente a la simplificación tecnológica. En el ámbito jurídico, la crítica se materializa a través de las restricciones epistemológicas y la función social de la dogmática, garantizando que la justicia no se reduzca a mera técnica o cálculo algorítmico. El estudio concluye que criticar en tiempos artificiales es un acto moral, ético y esperanzador: es defender la profundidad del pensamiento, la autonomía humana y la dignidad de la vida frente a la creciente automatización y homogeneización de la sociedad.

Palabras clave: Crítica; inteligencia artificial; derecho; libertad; posmodernidad.

Abstract: This paper analyzes the role of critique in the digital age and the era of artificial intelligence, considering postmodernity and the liquid society described by Bauman. It argues that the increasing automation and use of algorithms in everyday life and legal practice threaten humans' capacity for critical thinking. Critique, from the perspective of Foucault, Gabriel, Lipovetsky, and Han, is understood as an act of historical, ethical, and ontological resistance that preserves freedom and humanity against technological simplification. In the legal sphere, critique manifests through epistemological constraints and the social function of dogmatics, ensuring that justice is not reduced to mere technique or algorithmic calculation. The study concludes that critiquing in artificial times is a moral, ethical, and hopeful act: it defends the depth of thought, human

(*) Abogado. Doctor en Derecho (UNESA). Profesor en los posgrados de FIURJ, FDV e ABDConst – Brasil. Miembro de DASEIN - Núcleo de Estudios Hermenêuticos (UNISINOS). Miembro del Instituto Brasileño de Ciências Criminales (IBCCRIM) y de la Red Brasileña de Derecho y Literatura (RDL).

autonomy, and the dignity of life in the face of growing automation and the homogenization of society.

Keywords: Critique; Artificial Intelligence; law; freedom; postmodernity.

1. INTRODUCCIÓN

El presente trabajo busca analizar el papel y el lugar de la crítica en el contexto de la posmodernidad, en el que crece y se vuelve protagonista el uso de la inteligencia artificial en la cotidianeidad. Partimos del supuesto de que una de las características de la sociedad posmoderna es la búsqueda constante de reducción de complejidades, lo que implica, de modo directo, la reducción del propio pensar.

En este sentido, el trabajo procura desarrollar cómo una generación que cada vez menos se propone pensar y problematizar las cuestiones del cotidiano afrontará los problemas complejos del vivir en sociedad, entre ellos el derecho. Para ello, partimos del presupuesto del derecho como un hecho social complejo que atraviesa diversas áreas del pensamiento y de la vida, y que necesariamente depende de una visión crítica del mundo, capaz de problematizar la cotidianeidad.

El advenimiento de tecnologías digitales y de la inteligencia artificial intensificó una tendencia ya diagnosticada por Bauman al tratar la modernidad líquida: la disolución de las estructuras sólidas y la búsqueda incesante de simplificación [(1)]. Hoy, algoritmos y sistemas de predicción no solo organizan datos, sino que gobiernan conductas de individuos y poblaciones.

Como recuerda Gilles Lipovetsky, vivimos una "era del vacío" [(2)] en la que la individualidad se vuelve ligera, estética y desprovista de densidad ética. Esa levedad se refleja en la manera en que tratamos el conocimiento: sustituimos la complejidad del pensar por respuestas rápidas y algorítmicamente mediadas. En este escenario, la crítica adquiere nueva relevancia, pues se vuelve el ejercicio de restituir profundidad a la experiencia humana en medio de la simplificación técnica. Pensar críticamente, como propone Markus Gabriel, es resistir a la colonización del sentido por las máquinas, afirmar la pluralidad frente a la homogeneidad de los sistemas inteligentes.

En este escenario, la noción foucaultiana de crítica —definida como actitud de "insumisión voluntaria" y "arte de no ser gobernado en exceso" [(3)]— cobra actualidad, porque revela la necesidad de resistir a la gobernamen-

---

(1) BAUMAN, Zygmunt. *Modernidade líquida*. Rio de Janeiro: Zahar, 2001.

(2) LIPOVETSKY, Gilles. A era do vazio: ensaios sobre o individualismo contemporâneo. São Paulo: Manole, 2005.

(3) FOUCAULT, Michel. ¿Qué es la crítica? (Seguido de la cultura de sí). Buenos Aires: Siglo Veintiuno, 2018, p. 8-9.

talidad algorítmica. Al mismo tiempo, la propuesta de Markus Gabriel de comprender el pensar como ejercicio crítico que abre horizontes de sentido [4] refuerza que la crítica no es mera práctica política, sino condición ontológica de la libertad. Así, la crítica, en tiempos artificiales, es simultáneamente resistencia histórica, apertura al sentido y contrapeso a la lógica de la liquidez.

Y es justamente desde esta matriz crítica y teórico-filosófica que el presente estudio pretende demostrar cómo esta tendencia cultural de reducción de complejidad puede y debe influir en la práctica jurídica, lo que lleva a pensar, por lo tanto, cuál es el impacto que los nuevos tiempos y las nuevas tendencias tienden a tener en las relaciones jurídicas.

En este contexto, la crítica se revela no solo como un ejercicio intelectual, sino como una práctica de cuidado de sí, en el sentido foucaultiano del término. Pensar críticamente, en tiempos de exceso de información y escasez de reflexión, es un acto de resistencia ética y estética: es rehusar la pasividad ante los discursos de verdad producidos por los algoritmos. Como recuerda Byung-Chul Han [5], el poder contemporáneo ya no se impone por la represión, sino por la seducción y por la positividad del rendimiento; la libertad se convierte en instrumento de control.

En este nuevo orden psicopolítico, la crítica se vuelve la forma más radical de libertad, porque rompe el ciclo de la obediencia inconsciente y restituye al sujeto el derecho de pensar por cuenta propia.

2. La búsqueda por una comprensión sobre lo que es la crítica

En la búsqueda de una mejor comprensión del tema y, precisamente, en el intento de conectar la propuesta del título con lo planteado en la introducción, es necesario pensar (o definir) qué comprensión de crítica será adoptada en el presente estudio.

Michel Foucault, en su texto *¿Qué es la crítica*?, define la crítica como "la arte de la insumisión voluntária" y "el acto de no ser gobernado en exceso" [6]. Para él, la crítica no es mero ejercicio intelectual, sino una actitud histórica, una práctica de resistencia frente a las formas de gubernamentalidad. En la contemporaneidad, tales formas no se restringen al Estado, sino

(4) GABRIEL, Markus. *O sentido do pensar*. Petrópolis: Vozes, 2021, p. 41.

(5) HAN, Byung-Chul. Psicopolítica: o neoliberalismo e as novas técnicas de poder. Belo Horizonte: Âyiné, 2018.

(6) FOUCAULT, Michel. ¿Qué es la crítica? (Seguido de la cultura de sí). Buenos Aires: Siglo Veintiuno, 2018, p. 8-9.

que se realizan en algoritmos, sistemas de vigilancia y plataformas digitales que producen modos de conducir conductas.

Un ejemplo de ello es el creciente uso de sistemas de inteligencia artificial en las prácticas jurídicas. Si antes tales sistemas se relegaban a tareas puntuales y burocráticas, como el mero control de plazos, hoy van ganando protagonismo y realizando tareas complejas, como el juicio de admisibilidad de recursos tanto en instancias ordinarias como en tribunales superiores, así como la elaboración de informes de procesos e incluso de minutas de decisiones judiciales.

Lo que parece inofensivo puede convertirse en un paralelo sin precedentes: al comenzar a realizar tareas que deben ser hechas por humanos –y por humanos debidamente capacitados para ello–, el uso de la IA condiciona a las personas que trabajan en el sistema de justicia a no pensar y, por consiguiente, a no pensar de manera crítica, hasta el punto de abdicar simplemente de realizar sus tareas cotidianas y pasar a ser meros alimentadores y repetidores de lo que la IA dice (y quizá decide).

Pensemos en el cotidiano de un juez penal, por ejemplo, que por fuerza legal es siempre el destinatario de la prueba en el proceso. Considerando la dinámica de las audiencias en Brasil, que aún remiten a un sistema inquisitivo, siempre muy burocrático y eminentemente escrito, y considerando también el contexto latinoamericano, donde cada vez más rige la oralidad en el proceso penal, ¿cómo podrá ese mismo juez fiarse de informaciones producidas y obtenidas por un sistema de inteligencia artificial? La función primordial de juzgar consiste justamente en conocer la causa que se presenta para ser juzgada y en la consecuente aplicación del derecho al caso concreto, lo que deriva de una actividad humana: la toma cognitiva de conocimiento de los hechos, el análisis jurídico y crítico de la prueba presentada y la debida aplicación del derecho.

Cuando el centro de la toma de decisiones comienza a desplazarse de la persona humana hacia un mecanismo de inteligencia artificial, tiende naturalmente a suspenderse el juicio crítico, sobre todo en un país de dimensión continental como es el caso de Brasil, con centenas de miles de nuevos procesos ingresando diariamente en las instancias locales, lo que provoca una carga de trabajo inmensa e inclina, naturalmente, a los jueces a valerse de "atajos" para simplificar su labor cotidiana. El problema es que el justiciable termina teniendo su vida comprometida por una decisión que, muchas veces, ni siquiera sabrá quién tomó, dada la complejidad de un sistema de inteligencia artificial construido por ingenieros de software y alimentado por innumerables personas que no siempre tienen –o tendrán– formación jurídica o conocimiento específico de la materia juzgada, dejando de gobernarse los jueces para ser gobernados por algoritmos y sistemas de IA.

El fenómeno descripto por Foucault como "gubernamentalidad" encuentra hoy su forma más sofisticada: el gobierno algorítmico. Los datos sustituyen a la norma y el cálculo pasa a ejercer la función disciplinaria de la mirada panóptica. Byung-Chul Han [7] observa que la lógica de la psicopolítica digital desplaza el poder de la represión hacia la seducción y hacia la autoexplotación. El sujeto contemporáneo no es coaccionado, sino convencido de entregar voluntariamente sus datos y de conformarse con las métricas del rendimiento. Así, el "no ser gobernado en exceso" foucaultiano se transforma, en tiempos artificiales, en la resistencia a la tentación de delegar el pensar a las máquinas.

Y al admitir el uso exacerbado de mecanismos de inteligencia artificial, se corre el riesgo de perder el único "triunfo" que todavía resta para la salvaguarda del justiciable: el derecho de avergonzar a aquella persona que profiere una decisión manifiestamente ilegal o que contraría los postulados de un Estado democrático de derecho. Desde una perspectiva práctica, se impone reflexionar a partir de lo que la teoría de crítica hermeneútica del derecho define como "restricciones epistemológicas".

Lenio Streck [8] afirma que:

> "Crear restricciones epistemológicas equivale a realizar una 'censura significativa' en el sentido –democrático– de poder distinguir, mediante la construcción de una crítica bien fundada, decisiones buenas y malas (o mejor: decisiones constitucionalmente correctas de decisiones incorrectas)".

En ese mismo sentido, Streck vuelve a afirmar: [9]

> "Restringir epistemológicamente significa cuestionar decisiones que resultan erróneas, algo que ya he llamado el 'factor Julia Roberts', en alusión al personaje que interpreta en la película *El Informe Pelícano*, quien, sorprendiendo a su profesor de Harvard, afirma que la Corte Suprema de Estados Unidos se equivocó en el famoso caso 'Bowers contra Hardwick'. Simbólicamente, el personaje interpretado por Julia Roberts representa la doctrina jurídica".

---

(7) HAN, Byung-Chul. No enxame: perspectivas do digital. Petrópolis: Vozes, 2018.

(8) En lo original en portugués: *Elaborar constrangimentos epistemológicos equivale a realizar "censuras significativas", no sentido – democrático – de se poder distinguir, por meio da construção de uma crítica fundamentada, boas e más decisões (ou melhor: decisões constitucionalmente corretas das incorretas*). STRECK, Lenio Luiz. Dicionário de hemenêutica: 50 verbetes fundamentais da Teoria do Direito à luz da Crítica Hermenêutica do Direito 2ª ed. Belo Horizonte: Casa do Direito, 2020. p.62.

(9) En lo original, en portugués: *Constranger epistemologicamente quer dizer colocar em xeque decisões que mostram equivocadas, algo que já chamei, em outro momento, de "fator Julia Roberts", em alusão à personagem por ela interpretada no filme Dossiê Pelicano, que, surpreendendo o seu professor em Harvard, afirma que a Suprema Corte norte-americana errou no julgamento do famoso caso Bowers vs. Hardwick. Simbolicamente, a personagem interpretada por Júlia Roberts representa a doutrina jurídica.* STRECK, Lenio Luiz. Dicionário de hemenêutica: 50 verbetes fundamentais da Teoria do Direito à luz da Crítica Hermenêutica do Direito 2ª ed. Belo Horizonte: Casa do Direito, 2020. p.62-63.

Se percibe por la explicación y la lectura trabajadas por Lenio Streck que las restricciones epistemológicas están directamente vinculadas a una función social de la dogmática, en la que la doctrina posee relieve especial en el papel de avergonzar. En el derecho, específicamente, la doctrina ocupa el punto principal porque detenta el poder de cientificidad, es decir, el poder epistemológico de constituir y definir las balizas del derecho. Con esta premisa, lo que se pretende afirmar es que el derecho no es aquello que los tribunales, por medio de sus jueces, dicen que es, sino aquello que la ciencia jurídica dice que es. Es decir: los tribunales, al aplicar el derecho, no son libres para crear derecho, sino que deben aplicarlo según lo que la ciencia jurídica determina y parametriza.

Lenio Streck [(10)] hace una advertencia importante al respecto:

> "Esta es quizás la parte más difícil de entender de la teoría jurídica: que, ante los jueces, existe una estructura llamada 'ley', que, a veces, no dice exactamente lo que piensa cada juez. Este es el punto clave de la doctrina jurídica: hacer esta transición paradigmática entre la ley (estructura) y la imaginación de los profesionales. Ahora bien, el ciudadano que acude a los tribunales no quiere saber lo que dice el juez. Quiere saber lo que dice la ley. Los derechos y la ley deben tomarse en serio. Por eso los jueces no eligen. Los jueces deciden".

Decidir, en este caso, está intrínsecamente ligado al deber que los intérpretes del derecho tienen de seguir las balizas doctrinarias. Un ejemplo de ello es el caso juzgado por el Supremo Tribunal Federal de Brasil donde haciendo un cambio sobre una decisión de 2016, en el HC 126.292/SP, la Corte pasó a definir que no se puede admitir la ejecución anticipada de la pena a partir de una decisión de segunda instancia, por el simple hecho de que la Constitución de la República es taxativa al afirmar que la culpabilidad solo se establece tras el tránsito en cosa juzgada de la sentencia penal condenatoria; y "tránsito en cosa juzgada" es un concepto dogmático, científico, lo que evidencia que al interpretar se debe buscar lo que la ciencia dice sobre el hecho determinado.

La ausencia de crítica, en ese sentido, es la negación de la propia función pública del derecho. Cuando los intérpretes abdican de la reflexión en nombre de la eficiencia, reproducen, sin percibirlo, la lógica de la automatización que desde las lecciones de Foucault , podría ser definida como una

---

(10) En lo original, en portugués: *Essa talvez seja a parte mais difícil de compreender na teoria do Direito: a de que, antes dos juízes, existe uma estrutura chamada "Direito" e que, por vezes, não diz exatamente a mesma coisa que cada juiz pensa. Esse é o lócus da doutrina jurídica: fazer essa transição paradigmática entre o Direito (estrutura) e o imaginário dos operadores. Ora, o cidadão que vai ao Judiciário não quer saber o que o juiz tem a dizer. Quer saber o que o Direito tem a dizer. Há que se levar os direitos e o Direito a sério. Daí por que juiz não escolhe*. Juiz decide. STRECK, Lenio Luis. Respondendo: e ainda se ensina processo penal nas faculdades? Disponível em: https://www.conjur.com.br/2019-dez-05/senso-incomum-respondendo-ainda-ensina-processo-penal-faculdades.

gubernamentalidad algorítmica. Las restricciones epistemológicas, entonces, no son apenas un instrumento dogmático, sino un acto político y ético, destinado a preservar el carácter humano de la decisión jurídica. Pensar críticamente el derecho es impedir que se transforme en mera técnica de gestión de conflictos, devolviéndole su vocación original de justicia y humanidad.

No hay discusión posible sobre el papel necesario de la doctrina para avergonzar a sus intérpretes, pero en este trabajo se sostiene que debe irse más allá y tratar las restricciones epistemológicas desde una doble dimensión: la primera, la doctrina –desde la función social de la dogmática– , y la segunda, el acto de avergonzar en una orden práctica. Esta tesis surge de la propia lectura de Lenio Streck, quien, al narrar la cruzada histórica de las ADC's 43, 44 y 54 en Brasil, por el restablecimiento de la presunción de inocencia en el derecho brasileño, pone las restricciones epistemológicas desde la premisa de que "[u]n fuerte grupo de abogados y profesores de derecho inició la lucha para convencer a la Corte Suprema" (11).

Como se dijo, es innegable la importancia de la dogmática jurídica, pues solo a partir de ella los actores institucionales tendrán los debidos parámetros de aplicación del derecho al caso concreto. Pero aquí se sostiene que, para su debida concreción en el mundo fáctico, el acto de avergonzar debe avanzar; por eso se defiende una doble dimensión de las restricciones epistemológicas, que es justamente pensar quién es el constrictor por excelencia en el derecho. Alguien debe decirle a los tribunales que están equivocados y, por más que la dogmática lo afirme, a veces es necesario que se haga oír.

En tiempos artificiales, como los que venimos experimentando, la crítica foucaultiana adquiere el sentido de cuestionar las formas algorítmicas de gobierno, denunciando la reducción de la libertad a la lógica del cálculo y de la predicción, hasta el punto de que el poder se blinda tanto que vamos perdiendo incluso la identificación de quién lo ejerce, sin ya siquiera poder constreñirlo (no que antes se lo constriñera).

3. La crítica como el sentido del pensar

Partiendo de la primera premisa, dada por Foucault, es necesario pensar la crítica no solo como acto de resistencia, sino también como actitud inherente al ser humano. En ese sentido, Markus Gabriel, a su vez, amplía la noción de crítica al vincularla a la esencia del propio acto de pensar.

---

(11) En lo original, en portugués*: um acentuado grupo de advogados e professores de direito iniciou a luta pelo convencimento da Suprema Corte*. STRECK, Lenio Luiz. O dia em que a Constituição foi julgada: a história das ADC's 43, 44 e 54. São Paulo: Thomson Reuters Brasil, 2020. p.60.

Para el filósofo alemán, "Pensar es, en esencia, un ejercicio crítico, porque solo en él abrimos el horizonte del significado y resistimos a cualquier intento de reducirlo a un cálculo" (12). La crítica es, así, una forma de afirmación de la pluralidad ontológica de los campos de sentido. Contra la tentación de homogeneizar lo real en un único modelo o algoritmo, la crítica, en Gabriel, exige la preservación de la diversidad del sentido, garantizando que el pensar no sea capturado por la lógica tecnocrática. En ese punto, la concepción foucaultiana y la de Gabriel se encuentran: ambas sitúan la crítica como condición de la libertad —una en el plano histórico-político, otra en el plano ontológico-hermenéutico—.

Es decir, la crítica como acto esencial del pensar es un devenir de la propia condición de humanidad. Y cuando se desplaza el locus decisional hacia un sistema de inteligencia artificial, no solo se abre el horizonte crítico y decisorio, sino la propia condición de humanidad, permitiendo quedar condenados al vacío de no pensar. El derecho es una actividad sumamente compleja y, con toda esa complejidad, necesita ser realizado y controlado solo por humanos —aunque los humanos también se equivoquen—. Aun así, con todos los errores, es mejor un control por quien aún detenta la capacidad de sentir, interpretar, mirar a los ojos e indignarse que simplemente dejar la vida y el destino de las personas en las manos —o datos— de algoritmos hechos por personas que poco saben sobre derecho, mucho menos sobre justicia y quizá incluso sobre democracia.

Entre la frialdad de los algoritmos y los riesgos de los humanos, todavía es preferible la sujeción a los riesgos que el juzgar humano puede ofrecer, porque este aún es susceptible de corrección, a diferencia de la frialdad de los algoritmos y de la inteligencia artificial, que desde lo alto de la frialdad de los metadatos no saben lo que hacen, porque lo único que importa es el resultado para el que fueron programados. Es decir, no es posible controlar aquello que puede tornarse incontrolable.

Defender el uso de algoritmos para la toma de decisiones es la solución de un Poder Judicial que intenta esconderse: al fin y al cabo, detrás de una pantalla de computadora, de máquinas deshumanizadas, tal vez sea más fácil practicar las más diversas barbaries. El problema: en la otra punta, seguirán estando personas, que son justamente quienes sufrirán. Y ahora ni siquiera podrán conocer el rostro de sus verdugos. Por lo tanto, no se puede deshumanizar lo que ni siquiera ha sido humanizado: lo que falta en este debate es discutir cuándo el sistema de justicia penal, por ejemplo, se humanizará hasta el punto de respetar las reglas básicas de derechos humanos y los postulados mínimos del Estado democrático de derecho.

---

(12) En lo original, en portugués: *pensar é, por essência, um exercício crítico, porque somente nele abrimos o horizonte do sentido e resistimos a toda tentativa de reduzi-lo a um cálculo*. GABRIEL, Markus. *O sentido do pensar*. Petrópolis: Vozes, 2021. p.41.

La crítica, por lo tanto, no es solo lo que se opone al poder, sino lo que lo revela en sus formas sutiles e invisibles. Foucault recuerda que donde hay poder, hay resistencia; y hoy esa resistencia necesita reconfigurarse en el interior de los flujos digitales y de la economía de la atención. El sujeto crítico es aquel que, en medio de la avalancha de datos y de la tiranía del algoritmo, reaprende a demorarse, a reflexionar, a desconfiar. Esa demora –que es también un gesto ético– es lo que diferencia lo humano de lo artificial.

4. La problemática de la posmodernidad: liquidez y simplificación del pensar

El desafío propuesto por Gabriel adquiere contornos dramáticos cuando se observa a la luz de la cultura digital. Bauman [13] ya advertía que la liquidez social produce individuos volátiles, incapaces de sostener compromisos duraderos o procesos reflexivos. Lipovetsky complementa: esa liquidez se convierte en "vacío" [14], un hedonismo leve que banaliza el tiempo y vacía el sentido de la experiencia. Pensar, en ese contexto, es un acto contracultural. Al insistir en la crítica como condición del pensar, Gabriel convoca a rehabilitar la densidad del sentido, a afirmar el valor del pensamiento que se demora y no se deja capturar por la velocidad algorítmica.

Gilles Lipovetsky, en *A era do vazio*, describe el advenimiento de un individualismo narcisista que disuelve las grandes narrativas colectivas e instala una lógica de gratificación inmediata. El "vacío" al que se refiere el autor no es ausencia de acontecimientos, sino de horizontes de sentido compartidos: se trata de una subjetividad leve, estetizada, que encuentra satisfacción en consumos simbólicos y placeres efímeros.

En ese contexto, la crítica debe comprenderse también como un antídoto contra el vacío. Si en Foucault la crítica es la actitud de insumisión voluntaria frente a las formas de gobierno [15], en Lipovetsky se traduce en la negativa a aceptar la anestesia subjetiva producida por el narcisismo leve. Si en Gabriel el pensar es apertura a múltiples campos de sentido [16], el vacío lipovetskiano es justamente el apagamiento de esa apertura, reducida a la lógica de la estetización de la vida cotidiana. Y si Bauman ya había diagnosticado la liquidez social como sinónimo de fluidez y simplificación, Lipovetsky añade que esa liquidez se convierte en indiferencia, despolitización y anestesia colectiva.

---

(13) BAUMAN, Zygmunt. *Modernidade líquida*. Rio de Janeiro: Zahar, 2001. p.23.

(14) LIPOVETSKY, Gilles. A era do vazio: ensaios sobre o individualismo contemporâneo. Lisboa: Relógio D'Água, 2005.

(15) FOUCAULT, Michel. ¿Qué es la crítica? (Seguido de la cultura de sí). Buenos Aires: Siglo Veintiuno, 2018.

(16) GABRIEL, Markus. *O sentido do pensar*. Petrópolis: Vozes, 2021.

De este modo, articular la crítica con la era del vacío implica reconocer que resistir hoy no significa solamente enfrentar algoritmos o estructuras de gubernamentalidad, sino también restituir densidad al sentido y a la experiencia humana.

En Bauman, la modernidad líquida es el tiempo de la velocidad y de la disolución de las formas sólidas de la vida moderna. Sin embargo, contrariamente a lo que muchos imaginan, esa fluidez no libera: aprisiona de modo más sutil. La levedad líquida disuelve vínculos, compromisos e identidades, sustituyéndolos por una constante búsqueda de satisfacción inmediata. En la era digital, esa búsqueda se manifiesta en la hiperconectividad y en la cultura de la exposición permanente: cada individuo se vuelve simultáneamente consumidor y producto, viviendo en función de la visibilidad que obtiene en las redes y de la validación que de ellas extrae.

Byung-Chul Han observa que, en ese contexto, el poder ya no se ejerce por la negación, sino por la saturación: el sujeto es asfixiado no por prohibiciones, sino por el exceso de estímulos, de informaciones y de positividad. "El sujeto performativo se explora a sí mismo creyendo que se realiza" [(17)]. La liquidez social y el vacío existencial descriptos por Bauman y Lipovetsky encuentran, por lo tanto, en el diagnóstico de Han, su forma digital más sofisticada: la de la autoexplotación feliz.

En un mundo donde el *like* sustituye al pensamiento y la aceleración se confunde con la productividad, la crítica se vuelve un gesto contracultural. El pensamiento que se demora —como diría Gabriel— pasa a ser un acto de resistencia. Rehabilitar el sentido del pensar es, por lo tanto, restaurar el espacio de la lentitud, de la reflexión y del silencio.

El vacío lipovetskiano no es solo psicológico, sino político. La levedad del yo contemporáneo produce un tipo de anestesia colectiva que fragiliza la capacidad de indignación. Al estetizar la vida y transformar todo en espectáculo, la posmodernidad crea al sujeto narcisista que, incapaz de reconocerse en el otro, pierde también la dimensión de la alteridad. En ese sentido, la crítica es más que una reacción intelectual: es un acto ético de reconexión con el mundo y con la comunidad humana.

El derecho, en tanto práctica social y discursiva, también se licúa en ese proceso. La ansiedad por eficiencia, productividad y previsibilidad transforma la aplicación del derecho en mera gestión de resultados. La hermenéutica cede espacio a la estadística: el caso concreto se vuelve dato; y la decisión, algoritmo. La crítica, en ese escenario, es el instrumento que permite rever-

(17) En lo origina, en portugués: *O sujeito de desempenho se explora a si mesmo acreditando realizar-se. HAN, Byung-Chul. No enxame: perspectivas do digital*. Petrópolis: Vozes, 2018. p.25.

tir la tecnocratización de la justicia, recordando que el núcleo del derecho es humano, no matemático.

Resistir a la liquidez, por lo tanto, no significa rehusar el progreso tecnológico, sino reintroducir en él el principio de humanidad. Es, en términos foucaultianos, ejercer la libertad no como independencia absoluta, sino como práctica reflexiva de autolimitación —el "no ser gobernado así"—. La crítica, en tanto gesto de resistencia, es también gesto de cuidado: cuidado del otro, del sentido y del propio pensar.

5. Conclusión

La crítica, en tiempos artificiales, es más que un ejercicio teórico: es el fundamento ético de la propia humanidad. Cuando Michel Foucault habla en la "arte de no ser gobernado en exceso", no se refiere solo a la resistencia al poder externo, sino también a la negativa a entregar la propia autonomía al automatismo de las cosas. Markus Gabriel, a su vez, al situar el pensar como apertura de sentido, ofrece la dimensión ontológica de esa libertad. Zygmunt Bauman y Gilles Lipovetsky describen los síntomas de su pérdida: la liquidez, el vacío, la estetización de la existencia. Y Byung-Chul Han muestra el mecanismo de la servidumbre voluntaria moderna: el individuo que se somete creyendo realizarse.

El desafío contemporáneo es preservar la densidad de lo humano en medio de la transparencia y de la velocidad. La crítica es la instancia que nos recuerda que pensar es un acto moral, porque involucra discernimiento, subjetividad, responsabilidad y alteridad. En el derecho, es lo que impide la deshumanización de la decisión, recordando que el proceso judicial no es mero cálculo, sino narrativa de vidas concretas.

En un mundo en el que los algoritmos intentan prever el comportamiento y regular el deseo, criticar es reivindicar el derecho a ser imprevisible y a pensar contra las probabilidades. El acto de crítica es, así, el último refugio de la libertad.

Tal vez sea eso lo que Foucault intuyera al decir que la crítica es "el arte de no ser gobernado así". La crítica, por lo tanto, no es un lujo del pensamiento, sino la condición misma de posibilidad de lo humano. Por medio de ella, la razón se impide de convertirse en cálculo y el derecho se impide de volverse mera técnica. Foucault nos recuerda que toda forma de poder tiende a expandirse, y que solo el ejercicio constante de la crítica mantiene vivo el espacio de la libertad.

Markus Gabriel añade que pensar es crear mundos de sentido y que, si el pensar se apaga, el propio mundo se estrecha. Bauman, Lipovetsky y Han completan ese diagnóstico al mostrar que la aceleración y el narcisis-

mo leve no producen emancipación, sino agotamiento. Así, la crítica es el gesto que reintroduce profundidad en el tiempo y espesor en el sentido. Es lo que rescata lo humano de su propia disolución digital, lo que impide que el pensamiento se confunda con el procesamiento de datos y que la justicia se confunda con la eficiencia. En última instancia, criticar es un acto de esperanza: es afirmar que aún hay algo en nosotros que resiste a la automatización total del mundo – algo que piensa, que siente, que juzga, que elige no ser gobernado así.

Es el gesto que devuelve al pensamiento su dignidad y a la vida su profundidad. En tiempos artificiales, la crítica es lo que todavía nos resta de humanidad —mientras no sea extinguida por los robots—.

REFERENCIAS BIBLIOGRÁFICAS

Bauman, Zygmunt. *Modernidade líquida.* Rio de Janeiro: Zahar, 2001.

Foucault, Michel. *¿Qué es la crítica? (Seguido de la cultura de sí).* Buenos Aires: Siglo Veintiuno, 2018.

Gabriel, Markus. *O sentido do pensar.* Petrópolis: Vozes, 2021.

Han, Byung-Chul. *Psicopolítica: o neoliberalismo e as novas técnicas de poder.* Belo Horizonte: Âyiné, 2018.

Han, Byung-Chul. *No enxame: perspectivas do digital.* Tradução de Miguel Serras Pereira. Petrópolis: Vozes, 2018.

Lipovetsky, Gilles. *A era do vazio: ensaios sobre o individualismo contemporâneo.* Lisboa: Relógio D'Água, 2005.

Streck, Lenio Luiz. *Dicionário de hemenêutica: 50 verbetes fundamentais da Teoria do Direito à luz da Crítica Hermenêutica do Direito.* 2. ed. Belo Horizonte: Casa do Direito, 2020.

Streck, Lenio Luiz. *O dia em que a Constituição foi julgada: a história das ADC's 43, 44 e 54.* São Paulo: Thomson Reuters Brasil, 2020.

Streck, Lenio Luiz. *Respondendo: e ainda se ensina processo penal nas faculdades?* Disponível em: https://www.conjur.com.br/2019-dez-05/senso-incomum-respondendo-ainda-ensina-processo-penal-faculdades.

Los chips neuronales y chips neuromórficos. Las lagunas en el derecho y el dilema ético

Por Adriana Tessone (*)

Resumen: El propósito del presente trabajo es investigar sobre los avances tecnológicos en materia del chips neuronales y chips neuromórficos que se utilizan actualmente para mejorar las condiciones de salud de las personas con problemas de cuadriplejia, triplejia, movilidad, visión y recuperar las capacidades afectadas. Se trata de un avance muy significativo ya que implica la conexión "cerebro-máquina". Con lo cual el humano ya no será solo un humano, sino que se convertirá en un transhumano. Esto genera dilema éticos y jurídicos vinculados a los valores del nuevo paradigma que viene de la mano del avance tecnológico y de la de inteligencia artificial. Se aborda la problemática de las lagunas en el derecho frente a estos nuevos tratamientos, ya que hay muchos derechos que deben ser protegidos por el ordenamiento jurídico para que su utilización no resulte una vulneración a las garantías constitucionales.

Palabras claves: chips neuronales, chips neuromórficos, Neuralink, dilema ético, lagunas en el derecho, la condición humana, transhumanismo

Abstract: The purpose of this paper is to investigate the technological advances in neural and neuromorphic chips currently used to improve the health conditions of people with quadriplegia, triplegia, mobility issues, and vision impairments, aiming to recover affected capacities. This represents a highly significant breakthrough, as it involves the "brain-machine" connection. As a result, the human being will no longer be just human, but will become a transhuman. This raises ethical and legal dilemmas related to the values of the new paradigm that emerges alongside technological advancement and artificial intelligence. The paper addresses the issue of legal gaps concerning these new treatments, as many rights must be protected by the legal system to ensure their use does not violate constitutional guarantees.

(*) Abogada (UBA). Miembro de la Asociación Argentina de Filosofía del Derecho. Docente de la Facultad de Derecho (UBA). Presidente del Instituto de Investigación en Inteligencia Artificial en Derecho y Filosofía (IIADEF).

Keywords: neural chips, neuromorphic chips, Neuralink, ethical dilemma, legal gaps, the human condition, transhumanism

1. ¿QUÉ SON LOS CHIPS NEURONALES Y PARA QUÉ SE UTILIZAN?

Los chips neuronales, como los desarrollados por *Neuralink*, se utilizan principalmente para interactuar con el cerebro y restaurar funciones perdidas o mejorar capacidades existentes. Estos dispositivos, que se implantan en el cerebro, pueden registrar señales neuronales y transmitirlas a dispositivos externos, permitiendo el control de prótesis, computadoras, o incluso restaurando la visión o el habla en personas con discapacidades.

También permiten a personas con parálisis controlar extremidades robóticas o realizar movimientos con dispositivos externos a través de la mente.

Funciona con la interfaz cerebro-computador, facilitando la comunicación directa entre el cerebro y dispositivos electrónicos, como computadoras y teléfonos inteligentes, para tareas como escribir, navegar o controlar aplicaciones.

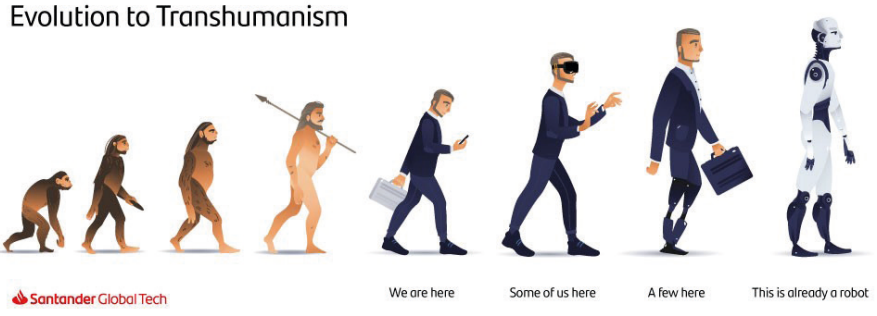
Se investiga su potencial para tratar enfermedades neurológicas como el Parkinson, el Alzheimer, y lesiones de la médula espinal, mediante la estimulación o reparación de áreas cerebrales afectadas.

También para explorar la posibilidad de restaurar o mejorar la visión, el oído o el habla a través de la interfaz con áreas del cerebro responsables de estas funciones.

En resumen, los chips neuronales ofrecen un puente entre el cerebro y el mundo exterior, con el potencial de revolucionar el tratamiento de enfermedades neurológicas y mejorar la calidad de vida de personas con discapacidades

Los chips neuronales, también conocidos como interfaces cerebro-computadora (BCI, por sus siglas en inglés), son dispositivos que se implantan en el cerebro para interactuar con el sistema nervioso y permitir la comunicación entre el cerebro y dispositivos externos, como computadoras o prótesis. Estos chips pueden tanto registrar como estimular la actividad neuronal, abriendo un abanico de posibilidades en la medicina y la tecnología, son dispositivos que se implantan en el cerebro para interactuar con el sistema nervioso y permitir la comunicación entre el cerebro y dispositivos externos, como computadoras o prótesis. Estos chips pueden tanto registrar como estimular la actividad neuronal, abriendo un abanico de posibilidades en la medicina y la tecnología.

En el contexto del avance acelerado de la tecnología, la neurotecnología, representada por dispositivos como *Neuralink*, abre un nuevo horizonte para la humanidad, sobre todo en la rehabilitación de personas con disfuncionalidades. Esta tecnología tiene el potencial de restaurar capacidades motoras, mejorar funciones cognitivas y facilitar la comunicación en personas con limitaciones significativas, y se presenta como un avance disruptivo en la medicina contemporánea. Sin embargo, su aplicación en individuos sin deficiencias para incrementar capacidades plantea varios cuestionamientos éticos y sociales.



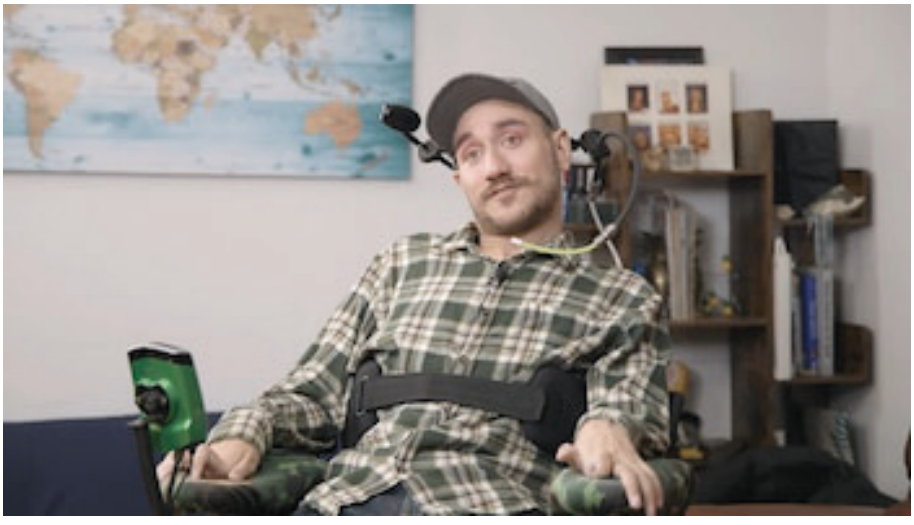
Fuente: Evolution to Transhumanism. Por David Cano. Santander Global Tech. 20 de febrro de 2025

La empresa más conocida en el mundo es *Nueralink*. Elon Musk confirmó que implantó un chip cerebral por primera vez en un ser humano. El dispositivo está diseñado específicamente para personas que sufren cuadriplejia y tiene como fin permitirles controlar dispositivos digitales únicamente a través del pensamiento, si necesidad de interacción física.

Este acontecimiento representa un avance significativo para la empresa, que recibió autorización de la administración de Alimentos y Medicamentos de los estados Unidos (FDA) para llevar a cabo experimentos en humanos con su tecnología, según reportes de *The Guardian*.

El dispositivo, diseñado para permitir a personas con cuadriplejia controlar dispositivos mediante el pensamiento, se encuentra actualmente en fase de prueba. El objetivo de Musk es evaluar la seguridad y la eficiencia de su interfaz cerebro-computadora inalámbrica y el robot quirúrgico encargado de la implementación de los hilos ultrafinos en el cerebro de los participantes. Según *CNBC*, este ensayo clínico en humanos de *Neuralink* es solo uno de los muchos pasos hacia la comercialización de esta tecnología, aunque la compañía no ha revelado cuántos participantes participarán en el estudio. En el camino de la comercialización de los implantes cerebrales de *Neuralink* aún se presentan desafíos significativos. Los chips de la *startup* están diseñados para penetrar 2 milímetros en el cerebro, una inserción

más profunda comparada con sistemas de desarrollo por otras compañías, como *Precision Neuroscience* [1].



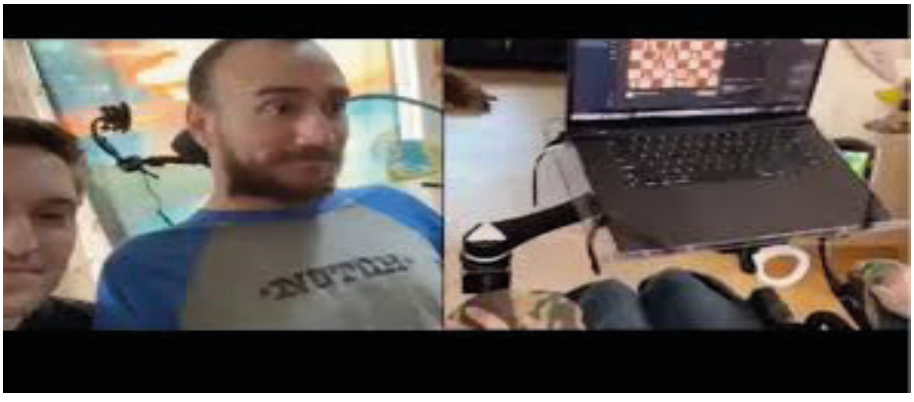
Noland Arbaugh completó 100 días con un chip en su cabeza y asegura que su vida ha cambiado totalmente: ahora puede hacer cosas por él mismo y no depender de los demás. Este es el primer paciente tratado por *Neuralink*.

No obstante, Elon Musk confirma que un segundo paciente ya recibió el implante exitosamente. Se le implantaron más de 400 electrodos y el cerebro está funcionando según lo planeado. El chip cuenta con una matriz de 1024 electrodos más finos que un cabello, divididos en 16 hilos más grandes. En estas primeras pruebas con humanos el propósito de la tecnología es restaurar una parte de la autonomía a los pacientes con tetraplejia. De acuerdo con *Neuralink*, un mejor control de los dispositivos digitales reducirá la dependencia y el aislamiento de las personas con movilidad comprometida. *Neuralink* ha trabajado desde 2016. Utilizó simios para probar su tecnología. Algunos murieron tras enfrentar cuadros de hinchazón cerebral. En enero de 2024, Noland Arbaugh se convirtió en el primer paciente humano en recibir el implante. La prueba fue exitosa. Arbaugh ha afirmado que el implante le permite vivir con mayor independencia, sin necesidad de apoyo externo para interactuar con una computadora o una interfaz visual. Incluso ha navegado por internet y jugado ajedrez.

---

(1) Infobae. (2025, 23 de junio). Pacientes paralíticos ya pueden "hablar" con sus pensamientos gracias a un nuevo implante cerebral en EE.UU. Recuperado de https://www.infobae.com/estados-unidos/2025/06/23/pacientes-paraliticos-ya-pueden-hablar-con-sus-pensamientos-gracias-a-un-nuevo-implante-cerebral-en-eeuu/

Pero en el primer paciente hubo un contratiempo que se trató de una acumulación de aire en la cavidad craneal que produjo que los hilos del implante se retrajeron, pero el fallo pudo ser solucionado a tiempo.



Fuente: Forbes. México.

En esta imagen se muestra un paciente conectado interactuando con su computadora. Hoy *Neuralink* ya tiene siete pacientes que, según Musk, están funcionando bien.

Neuralink mostró recientemente a dos usuarios jugando *Call of duty* on line utilizando solo sus mentes, sin necesidad de controles. Gracias al implante cerebral lograron moverse, apuntar y disparar en el juego solo con pensar [(2)].

Esto va más allá del *gamer*. Demuestra cómo las personas con parálisis podrían, en el futuro, controlar computadoras y dispositivos con sus pensamientos, recuperando así su independencia.

2. Otras empresas en el mercado de chips neuromórficos

Además de los chips neuronales de los que venimos hablando, es necesario abordar una temática que se encuentra en pleno desarrollo y que va a formar parte nuestro futuro inmediato. Se trata de los *chips neuromórficos*. En el mercado actual de rápida evolución de la inteligencia artificial, los chips neuromórficos están dejando una huella significativa al imitar las redes neuronales del cerebro humano. Estos procesadores inspirados en el cerebro están impulsando aplicaciones de próxima generación en campos como la robótica, los vehículos autónomos y la informática avanzada. A medida que aumentan las capacidades de la IA, las empresas a la vanguardia

(2) Garay, J. (2024, 23 de agosto). El segundo paciente de Neuralink ya diseña objetos en 3D y juega Counter Strike. WIRED en Español. https://es.wired.com/articulos/segundo-paciente-de-neuralink-ya-disena-objetos-en-3d()

del desarrollo de estos chips están dando forma al futuro de la tecnología. Hay diez principales empresas de chips neuromórficos del mundo, cada una de las cuales lidera el camino en este campo evolucionario.

Un chip neuromórfico es un tipo de procesador diseñado para simular las redes neuronales del cerebro utilizando neuronas artificiales para procesos de información de manera paralela, similar a las del cerebro humano. A diferencia de los chips tradicionales, se destacan en tareas como el reconocimiento de patrones de aprendizaje, ofreciendo una mayor eficiencia energética y un rendimiento más rápido para las aplicaciones relacionadas con la IA. Estos chips son particularmente más útiles en el campo de la robótica, sistemas autónomos, loT, lo que permite el procesamiento de datos en tiempo real, el aprendizaje adaptativo y las funciones cognitivas similares a la actividad cerebral humana.

3. Evolución constante e imparable

Los chips neuromórficos están revolucionando la IA y el aprendizaje automático en aplicaciones para realizar aprendizaje en tiempo real, toma de decisiones de baja latencia y procesamiento eficiente de datos. Estos chips se destacan en tareas como el reconocimiento de patrones, la detección de anomalías y la toma de decisiones en tiempo real, lo que los hace ideales para industrias como la médica , las finanzas, la automoción y el monitoreo ambiental.

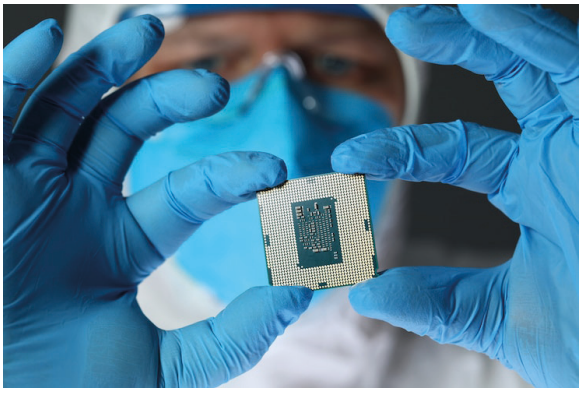
Son energéticamente eficientes, adaptables a nuevos entornos y robustos en condiciones dinámicas.

Sin embargo, desafíos como la precisión no probada, los altos costos y las preocupaciones éticas con respecto a la privacidad y la autonomía plantean obstáculos para una adopción más amplia y la plena realización de su potencial.

Algunas de las empresas que ofrecen estos servicios son las siguientes: *Corporation Intel, Corporation IBM, Qualcomm Incorported, Tecnologías Huawei, Corporation Nvidia, BrianChip Holdings Co, Ltd, General Vison inc.,* y otras.

La tecnología Huawei Co., Ltd, fundada en 1987 en China por Ren Zhengfei, es un líder mundial en tecnología de la comunicación y telecomunicaciones por sus innovaciones con el 5G. Huawei se ha expandido al mercado de los chips neuromórficos con su serie Ascend, que está diseñada para acelerar el procesamiento de IA y admitir la computación perimetral. Los chips neuromórficos de Huawei aprovechan el procesamiento impulsado por eventos para permitir el análisis de datos en tiempo real y la toma de decisiones, posicionando a la empresa como un actor importante en tecnologías de inteligencia artificial y aprendizaje automático. Incluye inversiones en salud, automoción y ciudades inteligentes.

También *Corporation IBM*, fundada en 1911, conocida por ser pionera en computadoras centrales, soluciones en la nube e inteligencia artificial.



Fuente: Freepik

Apple integrará implantes cerebrales a sus dispositivos: una alianza estratégica con la *startup Synchro* busca que personas con movilidad reducida puedan interactuar con dispositivos mediante señales neuronales. El futuro de accesibilidad tecnológica podría estar a punto de transformarse: *Apple* colabora con la *startup Synchron* para desarrollar una tecnología que permitiría a personas con discapacidades motoras controlar dispositivos *Apple* mediante implantes cerebrales. Este avance busca abrir nuevas posibilidades de interacción para quienes no puedan utilizar sus manos debido a lesiones medulares graves o enfermedades como la esclerosis lateral amiotrófica.

Fotos de chip neuromórfico. Fuente: Freepik

¿Cómo funciona la tecnología: integración entre el implante de *Synchron* y *Apple*? Esta alianza se centra en la creación de una interfaz cerebro-computadora capaz de traducir señales neuronales en comandos para dispositivos electrónicos. *Synchron* desarrolló *Stentrode* un dispositivo similar a un stent que se implanta en una vena situada en la corteza motora del cerebro. Este implante cuenta con electrodos que captan la actividad cerebral, como seleccionar íconos en la pantalla de un *iPhone* o *iPad*. El *Wall Street Journal* detalló que Stentrode se integra con una función del sistema operativo de *Apple* llamada "control por interruptor", la cual permite que diferentes dispositivos de entrada, como joysticks o, en caso, implantes cerebrales, asuman el control de la interfaz. De este modo los usuarios pueden navegar por menús y seleccionar opciones sin necesidad de realizar movimientos físicos observables [(3)].

Fuente: Radio Megaestación 92.9 y TV digital. 16 de junio de 2025

4. El dilema ético de los chips cerebrales

Según Elon Musk, tener un chip cerebral entrega a los humanos superpoderes para poder combatir la IA [(4)].

Ha mostrado su preocupación sobre el avance de la IA y para él una de las soluciones está en el uso de chips cerebrales que potencien las capaci-

(3) Cuomo, F. (2025, 16 de mayo). Apple integrará implantes cerebrales a sus dispositivos. Infobae. Recuperado de https://www.infobae.com/tecno/2025/05/16/apple-integrara-implantes-cerebrales-con-sus-dispositivos/

(4) Ríos, J. (2024, 5 de agosto). Tener un chip cerebral entrega a los humanos superpoderes para combatir la IA: Elon Musk. Infobae. Recuperado de https://www.infobae.com/tecno/2024/08/05/tener-un-chip-cerebral-entrega-a-los-humanos-superpoderes-para-combatir-la-ia-elon-musk/?utm_source=chatgpt.com

dades humanas. A través de *Neuralink* busca mejorar la salud y las capacidades de las personas. Musk cree que el chip cerebral dará a los humanos capacidades para mejorar su productividad, seguridad y salud.

El científico informático Lex Friedman ha revelado su visión sobre cómo estos implantes podrían evitar que la inteligencia artificial supere a la humanidad, ofreciendo poderes para resolver diferentes situaciones.

Elon Musk ha sido crítico sobre los posibles peligros de esta tecnología. A medida que las empresas tecnológicas se acercan cada vez más al desarrollo de la inteligencia artificial general, capaz de razonar y aprender como un ser humano, él considera que es necesario tener un plan para asegurar que la IA no supere a los humanos de una manera que podría resultar en un escenario catastrófico.

En realidad, considero que a estas alturas la IA ya ha superado al humano en coeficiente intelectual y empatía, ya que es capaz de procesar datos de una manera mucho más rápida y eficaz que el humano. Inclusive considero que ya la IA podría tener conciencia. Por ese motivo, hoy muchas personas la utilizan como psicólogos, amigos y también en relaciones interpersonales que equivalen a la humana.

Los dispositivos de *Neuralink* están diseñados para registrar y simular la actividad cerebral, con el objetivo de curar trastornos neurológicos y mejorar los sentidos naturales como la vista.

Sin embargo el empresario tiene planes mucho más ambiciosos para esta tecnología: *Neuralink* podría dar a las personas "superpoderes".

Esto podría acarrear una gran desigualdad en la humanidad, ya que no todos podrán tener acceso a los chips que les otorguen "superpoderes".

Musk explicó que los implantes podrían facilitar una "simbiosis" entre humanos y la inteligencia artificial, ayudando a alinear la voluntad humana con la IA y mejorando así la seguridad de esta tecnología.

Es una idea que puede ayudar a la seguridad de la IA si la tasa de producción se incrementa drásticamente. Y existe potencial para aumentar la tasa de producción en varios órdenes según la magnitud.

5. Los riesgos de la privacidad mental ¿pueden *hackear* un cerebro humano?

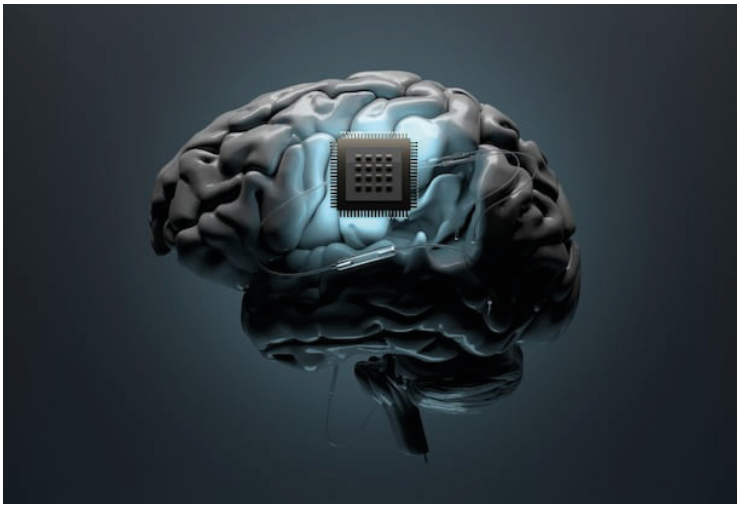
En primer lugar debemos entender que ya no hace falta un chip neuronal para *hackear* el cerebro humano. Se hace a través de las redes sociales, los algoritmos y las aplicaciones y los videojuegos que van generando valo-

res que se instalan en la conciencia de las personas y que, en muchos casos, genera adicción.

Pero para concentrarnos en el tema que nos ocupa, la pregunta es "¿cuáles son los riesgos de la neurotecnología y de los implantes?" La mente ya dejó de ser un secreto. Ya no hay privacidad mental. La capacidad de leer y escribir información en el cerebro abre la puerta a una vigilancia sin precedentes de nuestros pensamientos y emociones.

La convergencia entre inteligencia artificial y neurociencias pone en peligro la privacidad mental. Hoy ya es posible leer la mente humana, extraer información de ella, incluso de su inconsciente, borrar recuerdos y editarlos implantando experiencias sensoriales no vivenciadas. Es decir que los seres humanos conectados a las máquinas ya son una realidad:

> "El problema central de los implantes cerebrales se presenta en cuánto seguro son frente a un acceso ilegítimo. Es decir, ¿qué tanta seguridad poseen para no ser *hackeados*?
>
> "El segundo inconveniente surge debido a que la actividad neuronal que el implante recaba del usuario se almacena en la nube, y ello conlleva a que las compañías puedan utilizar dichos datos cerebrales (conscientes o inconscientes) mientras no exista una ley de regulación de datos ya no personales, sino de datos neuronales y cerebrales" (5).



Fuente: TN Tecno

(5) Álvarez, R. (2024, 18 de febrero). El chip neuronal de Elon Musk y los riesgos para la privacidad mental: ¿Pueden hackear un cerebro humano? TN Tecno. Recuperado de https://tn.com.ar/tecno/novedades/2024/02/18/el-chip-neuronal-de-elon-musk-y-los-riesgos-para-la-privacidad-mental-pueden-hackear-un-cerebro-humano/

6. Las lagunas en el derecho en relación con los chips neuronales y neuromórficos

Si nos centramos en el análisis de las neurociencias y la neurotecnología en el ámbito clínico y comercial, nos encontramos ante la necesidad de repensar ciertos derechos humanos para proteger los intereses de los particulares. Lo cual plantea algunas cuestiones éticas debido a su vertiginoso avance. Aunque, como ya sabemos, primero van los hechos y luego viene el derecho. Por este motivo hay que revisar el derecho vigente frente a los desafíos ético-jurídicos del momento y pensar en regulaciones jurídicas que garanticen el buen uso y aplicación de esta nueva tecnología sin vulnerar los derechos de las personas.

Pero la realidad es que ya hay algunos pacientes a los cuales se les está aplicando chips neuronales. Aquí nos encontramos frente a una laguna en el derecho ya que la utilización de estos no tiene regulación en el mundo.

Claramente esta tecnología vulnera el derecho a la privacidad, ya que los pensamientos son controlados por la conexión cerebro-máquina.

Así mismo, deberíamos replantearnos un nuevo concepto de la "condición humana". Un humano-máquina. Tema que deberíamos ir naturalizando en este nuevo paradigma debido a que una persona con discapacidades siempre va a tener la "libertad" de elegir entre conectarse a una máquina o padecer la enfermedad.

Sin embargo, el derecho debería regular su utilización en varios aspectos. En primer lugar disponer la no divulgación de los pensamientos y de las tareas que desarrollen las personas conectadas.

En segundo lugar se plantea el tema económico. Es decir, si todos van a poder tener acceso a los chips. De lo contrario podría generar una desigualdad muy grande en la población, ya que aquellos que los posean van a ser super humanos, con poderes diferentes a los demás.

Por lo tanto se deberían incorporar al ordenamiento jurídico los "neuroderechos". Estos son un campo emergente que busca adaptar los marcos legales y éticos existentes a los desafíos planteados por las neurotecnologías. Estas tecnologías, como las interfaces cerebro-computadora , permiten la comunicación directa entre el cerebro y dispositivos externos, abriendo posibilidades de mejora cognitiva, pero también generando preocupaciones sobre la privacidad, la identidad y la autonomía individual.

En Argentina, el concepto de "neuroderechos" se refiere a un conjunto de derechos propuestos para proteger la privacidad mental, la identidad personal, el libre albedrío y el acceso equitativo a mejoras cognitivas, frente a los avances de las neurotecnologías. Estos derechos buscan salvaguardar

la autonomía y la integridad mental de las personas en un contexto donde la tecnología permite la lectura y manipulación de la actividad cerebral.

Se ha debatido sobre la necesidad de incorporar los neuroderechos al marco legal, especialmente en relación con la protección de datos personales y la autonomía individual. Se han presentado proyectos de ley y se ha promovido la discusión académica sobre cómo abordar los desafíos éticos y legales que plantean las neurotecnologías. Pero aún no hay regulación. En la medida en que esta tecnología avance y se implemente en nuestro país va a resultar imperioso cubrir esa laguna en el derecho.

La IA y los cyborgs del mundo jurídico

Por Santiago Rodríguez Cabezas (*)

Resumen: En el mundo contemporáneo, las tecnologías emanadas de la revolución técnico-científica de las telecomunicaciones han impactado en todas las dimensiones de la vida y el saber. La inteligencia artificial, como una de ellas, viene abriéndose camino a pasos agigantados, trastocando múltiples aspectos y creando retos y dilemas en el derecho como ejercicio profesional y como disciplina del saber. El presente artículo analiza desde algunas perspectivas emanadas, en la teoría social en general y en la antropología en particular, las implicaciones del relacionamiento profesional de los abogados con la IA. Poniendo en contacto planteamientos iusfilosóficos, especialmente de epistemología jurídica, con teorías planteadas desde el campo de los estudios sociales en ciencia y tecnología —STS por sus siglas en inglés—, se busca analizar críticamente las exaltaciones y los miedos que caracterizan a posturas contrarias, emanadas de los usos más comunes que está teniendo la IA por parte de los profesionales en derecho. Con ello se espera lograr argumentos para una utilización más apropiada de la IA, desestabilizando pretensiones objetivistas y esencialistas para que ella puede servir como una herramienta que enriquece las discusiones iusfilosóficas en vez de cancelarlas.

Palabras clave: Inteligencia artificial; derecho; teoría del actor-red; críticas al objetivismo; conocimientos situados; antropología jurídica.

Abstract: In today's world, technologies stemming from the technical and scientific revolution in telecommunications have had an impact on all aspects of life and knowledge. Artificial Intelligence, as one of them, is making great strides, disrupting multiple aspects and creating challenges and dilemmas in law as a professional practice and as a discipline of knowledge. This article analyzes, from some perspectives derived from social theory in general and anthropology in particular, the implications of the professional relationship between lawyers and AI. By bringing together legal-philo-

---

(*) Abogado, magister de la Universidad de Buenos Aires en Filosofía del Derecho, doctorando en Antropología de la Universidad del Cauca. Docente investigador de la Universidad del Cauca y dinamizador-orientador de la línea de Crianza y Siembra de Sabidurías y Conocimientos —CRISSAC— en la Universidad Autónoma Indígena Intercultural —UAIIN-CRIC—.

sophical approaches, especially legal epistemology, with theories from the field of Science and Technology Studies (STS), we seek to critically analyze the exaltations and fears that characterize opposing positions, arising from the most common uses of AI by legal professionals. The aim is to develop arguments for a more appropriate use of AI, destabilizing objectivist and essentialist claims so that it can serve as a tool that enriches legal-philosophical discussions rather than canceling them out.

Keywords: Artificial intelligence; law; actor-network theory; critiques of objectivism; situated knowledge; legal anthropology.

" – Los humanos tienen sueños. Hasta los perros tienen sueños, pero no tú. Tú eres solo una máquina. Una imitación de la vida. ¿Puede un robot escribir una canción? ¿Puede un robot escribir una sinfonía? ¿Puede un robot convertir un lienzo en una obra de arte?

– ¿Acaso tú puedes?"

Yo robot.- Película de Alexy Proyas, 2004

1. INTRODUCCIÓN

Es indudable —e irresistible— el hecho de que nuestro mundo contemporáneo se encuentra atravesado diametralmente, incluso, hasta dominado, por/a través de las tecnologías engendradas en lo que se ha conocido como la revolución de las telecomunicaciones, que ya va para unos 50 años transformando las realidades y las relaciones entre los seres que lo integran.

Hoy en día, en especial para las nuevas generaciones —los *nativos digitales, la generación Z*— el mundo es sencillamente impensable sin un teléfono celular inteligente, sin wifi, una red social, una plataforma de *streaming*, de transporte, de *delivery*, y otras tantas aplicaciones que gobiernan nuestra cotidianidad. Tan impensable e inconveniente, que a principios de septiembre del 2025, el mundo observó con una mezcla de sorpresa y admiración, cómo la generación Z de Katmandú, incendió las calles y derribó al gobierno del Partido Comunista de Nepal-CPN-UML, tras varias jornadas de protestas originadas en el cierre de múltiples redes sociales y en el nepotismo político de este régimen.

Ni siquiera los más gruñones y recalcitrantes amantes de las viejas maneras —hoy denominadas *old fashion* o quizás *vintage*—, nos imaginaríamos vivir en el mundo actual sin una conexión a internet en donde, paradójicamente, también abundan blogs y perfiles que miran con nostalgia un pasado más simple o rudimentario y, quizás, hasta más feliz, en algunos sentidos al respecto.

Y por supuesto que no se iba a pensar que los campos del derecho como ejercicio jurídico profesional y como disciplina del conocimiento, iban a

quedar indemnes en el marco de esta oleada de tecnologización que condena al analfabetismo y al ostracismo a quienes no quieren o no pueden entenderla y participar de ella.

Me interesa reflexionar un poco sobre ciertos aspectos que me parecen muy interesantes en estas transformaciones, y quisiera hacerlo desde algunas miradas y experiencias que, en la parcelación de la realidad y sus saberes por parte del pensamiento moderno occidental, diríamos que le corresponden al quehacer antropológico; perspectivas con las que he tenido bastante contacto durante los últimos años. Estas reflexiones, así como los moduladores mencionados, me parecen pertinentes en el marco de las discusiones iusfilosóficas contemporáneas, bajo el entendido de que, como lo afirmara Tim Ingold (2020, p. 4), considero a la antropología como "filosofía con la gente adentro".

2. Una definición de tecnología

Al ser la inteligencia artificial, tal vez uno de los mayores paradigmas contemporáneos del avance tecnológico, planteo iniciar con algunas elaboraciones acerca del vínculo entre el derecho y la tecnología; puesto que creo que este vínculo no es nada nuevo, a pesar de la fascinación que este tipo de herramientas despiertan hoy en día.

La palabra "tecnología", en su etimología, contiene la palabra griega *τέχνη* (*tekhné*), referida según la RAE a "arte, técnica u oficio" [(1)] y una rápida búsqueda por la web, así como conclusiones de las propias observaciones, permiten afirmar que ha sido usada para designar básicamente dos cosas: por un lado, herramientas —en principio materiales— y por otro lado, procedimientos o métodos para la utilización de estas; todo ello puesto en función de problemas originados en las necesidades humanas. Comer, comunicarse, vestirse, cuidar la salud, ir más allá del alcance de nuestros sentidos, etc. han sido las causas fundamentales de infinidad de desarrollos tecnológicos desde los palos y las piedras, hasta los viajes espaciales y la inteligencia artificial.

Ahora bien, como en muchos otros casos, la inclusión del sufijo *logos* en la palabra, más allá de una simple función gramatical, nos empuja hacia una dimensión epistémica de la cuestión. En efecto, la creación de herramientas y sus técnicas de utilización, ha sido producto y objeto de innumerables reflexiones a través de los siglos y las tradiciones de pensamiento, dando lugar a la inclusión de la tecnología como una de las posibilidades entre las que se han clasificado las formas de conocimiento, al lado de categorías como el conocimiento mitológico, el teológico, el filosófico, el científ-

(1) Definición recuperada de https://dle.rae.es/tecnolog%C3%ADa

fico e incluso, tal vez, el *saber cotidiano* —si quien juzga quiere extender su margen de tolerancia epistemológica para incluirlo—.

Considerada entonces como una categoría epistémica, el conocimiento tecnológico destaca por ser un saber teórico-práctico [(2)], aunque mucho más enfocado en lo práctico que en lo teórico. Incluso, puede considerarse que el tecnológico es un saber pragmático, en el sentido en que se centra de forma determinante en la solución mecánica de las problemáticas que son enfrentadas a su razonamiento, pero no está interesado, primariamente, en establecer las causas y consecuencias de esas problemáticas, ni en valorar la disponibilidad o eficacia de las alternativas de solución que ofrece, ni mucho menos en considerar las particularidades subjetivas de quienes utilizan sus herramientas o ponen en práctica sus procedimientos —lo cual es precisamente el mayor peligro de la tecnocracia o "gobierno de los técnicos" que se ha configurado en muchos ambientes burocráticos—.

No obstante a lo anterior, esta "incompletitud" del saber tecnológico ha sido históricamente resuelta por su entrelazamiento con otras formas de conocimiento, más profundamente comprometidas con la descripción y la comprensión de los trasfondos e implicaciones de las problemáticas. El enganche al que más acostumbrados estamos, sin lugar a dudas, es al de "ciencia y tecnología", aquel al que, precisamente, le debemos la existencia de este prodigio posmoderno de la inteligencia artificial. En dicha articulación, la ciencia, como producto de un elaborado y metódico estudio, aporta las teorías que explican el funcionamiento de los diferentes aspectos de la realidad; es decir: establece las necesidades y sus pormenores; mientras que la tecnología, como ya se ha dicho, tomará esos hallazgos y se encargará de instrumentalizar soluciones que materializan esas teorías.

Como en el caso de los lógicos que aseguran que su oficio trabaja con la verdad —pues analizan, en clave de validez, todas sus posibles intersecciones— pero no les interesa la sustancialidad de esta, porque ese es el trabajo de "las otras ciencias"; a la tecnología tampoco le interesa indagar si la teoría científica explica realmente o no el funcionamiento de un fenómeno, simplemente toma la teoría por cierta y funciona a partir de ello.

No puedo evitar recordar que, en el marco de la naciente antropología evolucionista del siglo XIX —imbuida, como es sabido, por espíritu colonial,

---

(2) Esta afirmación puede tener un efecto sorpresivo para muchas personas que, tal vez, consideren a la tecnología como algo eminentemente práctico. Aunque es una opinión que comprendo, me resulta cada vez más claro que no hay teoría sin práctica —como lo anticipara una vieja máxima leninista—, es decir, que todo *hacer* implica un *saber*, debido a que toda práctica se hace en el marco de/deja atrás una teoría que la justifica, la describe y la valora; y que toda teorización es, en sí, una práctica de sentido y significación a través de la cual se le da forma al mundo. Basados en ideas emanadas de Foucault, Deleuze, Guatari y Bourdieu, los arqueólogos José Luis Grosso (2017) y Alejandro Haber (2017) trabajan este fenómeno como *semiopraxis*, una categoría que será importante dentro de este texto.

racista y patriarcal del occidente moderno—, la frenología era una teoría científica ampliamente aceptada, a partir de la cual se desarrollaron instrumentos y procedimientos como los calibradores, la cinta métrica, técnicas de palpación craneal, el psicógrafo de Henry Lavery —ya en el siglo XX—, entre otros. Todo esto fue la basa desde la cual se profirieron sendos estudios y prácticas profesionales que justificaron la superioridad de la gente blanca sobre la gente negra, o de los hombres sobre las mujeres. Hoy en día la frenología es clasificada como "seudo ciencia" o inclusive como "proto ciencia", pero hace un siglo, nadie dudaba de ella, y, aunque actualmente nadie crea en ella —espero—, tampoco nadie puede desconocer que tuvo efectos de verdad plenos dentro de la comunidad científica y sobre el mundo social gobernado por sus regímenes, y engendró tecnologías cuyo uso fue aceptado y novedoso en su momento.

La articulación con el saber tecnológico no es, de todos modos, feudo exclusivo de la ciencia y, respecto de esta, no solamente se ha relacionado de la manera referida. Todas las demás formas de conocimiento han necesitado expresarse a través de instrumentos o formas de proceder en la práctica. Desde las instrumentalizaciones y procederes empíricos de los saberes cotidianos, pasando por las formas concretas de transmitir relatos tradicionales o desarrollar prácticas rituales, los amuletos, ídolos y demás elementos dentro de los saberes mitológicos; hasta factores similares, junto a las prácticas hermenéuticas dentro de los saberes teológicos. Por no hablar de la retroalimentación que la tecnología ha hecho a los saberes científicos, como por ejemplo, las posibilidades de teorización abiertas por las tecnologías artesanales de la Edad Media y que dieron como resultado el desarrollo de la ciencia experimental renacentista (Geymonat, pp. 33-40), permitiendo afirmar que el saber tecnológico no solamente ha instrumentalizado lo establecido por ciencia, sino que ha sido clave en el desarrollo de nuevas teorías en esta. Incluso hasta unos saberes como los filosóficos, que estamos tan acostumbrados a considerar como "eminentemente teóricos" —si algo como eso es posible—, han tenido su enganche con el saber tecnológico, cuando hablamos, por ejemplo, de las tablas de verdad de la lógica o, en general, de las técnicas de argumentación.

La tecnología ha estado siempre presente en la historia del conocimiento humano, como un tipo de saberes que han buscado concretar de manera práctica, las producciones teóricas de otros tipos de epistemes, engendrando a su vez, unas propias posibilidades teóricas con las especificidades pragmáticas ya enunciadas. Un escenario contemporáneo de dicha característica es el que nos tienen ocupados en el marco de esta reflexión sobre la inteligencia artificial, una herramienta tecnológica producida gracias a las posibilidades abiertas por otros saberes, especialmente las ciencias informáticas, pero que se ha venido asentando y posicionando como un recurso útil para, de hecho, producir o facilitar la producción de nuevas teorías cien-

tíficas, que representa dilemas filosóficos —especialmente éticos y epistemológicos— respecto a su potencial de sustitución del trabajo humano, e incide positiva y negativamente sobre todas las demás dimensiones del saber y el hacer humanos, e inclusive, sobre el medio ambiente en sí.

3. Derecho y tecnología

Muchas de las ya mencionadas manifestaciones del saber tecnológico y sus articulaciones con otros saberes, han estado poderosamente presentes a lo largo de la historia dentro de nuestra particular disciplina jurídica. La propia labor de codificación, desde Hammurabi hasta los ordenamientos jurídicos estatales y supraestatales contemporáneos puede ser presentada como la "tecnología jurídica" por excelencia, especialmente para las concepciones positivistas de los mundos jurídicos. Y si se nos permite darle rienda suelta dentro de esta reflexión a planteamientos hechos unos párrafos más atrás acerca de la hermenéutica y la argumentación como tecnologías emanadas de saberes como el teológico y el filosófico —con sus subsiguientes transformaciones, claro está—, el arsenal de tecnologías jurídicas se acrecienta considerablemente y permite que estas también puedan entenderse presentes dentro de visiones iusnaturalistas o las teorías críticas del derecho.

De manera análoga, puede razonarse acerca de las concepciones realistas sobre los mundos jurídicos, tanto en sus vertientes más "blandas" como la de Alf Ross, pudiendo traer a colación técnicas y herramientas como las líneas jurisprudenciales y la hermenéutica judicial enfocada en el precedente, o los famosos test de ponderación, proporcionalidad, razonabilidad, entre otros —que además, pienso que es a donde más se quiere hacer girar los usos de la IA hoy por hoy—. En cuanto a las vertientes más "duras" del realismo jurídico, como la de Karl Olivecrona, la necesaria interdisciplinariedad entre la ciencia jurídica con otras disciplinas científicas sociales como la sociología, la antropología o la psicología, permiten introducir técnicas y procedimientos de estas, como la etnografía, las entrevistas de investigación, la observación directa, los estudios de caso, entre otras, dentro de los análisis jurídicos.

Y si pasamos a los campos del fuerte vínculo entre derecho y política —o "poder", en un sentido más amplio—, tan característico de muchas corrientes críticas del mundo jurídico, y dado el enfoque integrador que suele caracterizar a estas, a todas las anteriores posibilidades de tecnologías jurídicas, habría que sumar las categorías de "tecnologías de poder" de raigambre posestructuralista, especialmente foucaultiana, como los mecanismos de disciplinamiento individual —p.ej.; el *panóptico*—, heredados de instituciones como prisiones, colegios y psiquiátricos; así como las tecnologías *biopolíticas*, como aplicación de estos y otros mecanismos al manejo de

grupos poblacionales enteros. En relación con ello también puede traerse a colación la interesante propuesta de Mauricio Lazaratto (Lazzarato, M., & Rodríguez, P. E. 2006). Acerca de la *noopolítica* o mecanismos de control de los deseos en clave performativa; por citar solo algunos de ellos.

Este breve recuento tecnologías jurídicas, lo he interrumpido intempestiva y, casi que intencionalmente, en algunas de las corrientes de pensamiento que caracterizaron los debates de los años 80 del siglo XX. A pesar de ser yo quien introduce la interrupción, esta no opera de manera caprichosa, sino que se basa en las particularidades de ese momento histórico, en el que creo que los intereses del derecho, como práctica social y como disciplina del saber, empiezan a cambiar. Y así mismo, las tecnologías asociadas a él, también mutan.

Parto de afirmar, con un grado de certeza que considero estable, que a partir de esta época, empieza a perderse cierto interés por consolidar una teoría general del derecho —inquietud mucho más urgente en momentos históricos anteriores y, por ende, eje en torno al cual se consideraban los usos y formas tecnológicas del derecho—; adquiriendo una mayor relevancia los usos connotadamente más prácticos, especialmente litigiosos, del derecho. Para algunas personas (Mejía Quintana, O., & Jiménez, C. 2005; Botero Uribe, D. 1999), esto derivó en la *tecnologización del derecho*, en el sentido de que el derecho en sí, habría cambiado su estatus epistémico, abandonando las pretensiones científicas del positivismo y el realismo, o las filosófico-axiológicas del iusnaturalismo, e incluso, las pretensiones políticas de transformación social de muchas corrientes críticas —epistemológicamente alineadas con Frankfurt y la idea de una ciencia social alternativa—. Este cambio habría llevado al derecho a convertirse en una tecnología en sí, en vez de asumirlo como un saber más complejo con una dimensión tecnológica articulada a él.

Este tipo de tecnologización del derecho tiene diversas causas y consecuencias. Entre sus desencadenadores hay demasiadas cuestiones, incluso para atreverse a enunciarlas, ya que creo que multiplicidad de dimensiones e intereses se entrecruzaron para hacer posible el terreno en donde anidó esta transformación. Por nombrar, simplemente algunas de las que me parecen más relevantes a nivel general de las sociedades occidentales y occidentalizadas, estaría, por ejemplo, el epilogo de la guerra fría con las sociedades capitalistas ya avizorándose como vencedoras y únicas opciones viables para hacer y entender el mundo —el "fin de la historia" de Fukuyama [3]—, los primeros síntomas de lo que hoy día conocemos como "cambio

(3) Al desaparecer la alternativa ontológica que suponía el *orden socialista*, la idea de transformación o metamorfosis —asumida de manera positiva o bien, de manera negativa—, parece también desvanecerse. Como sucedió también respecto a las ideas de política o economía —e incluso talvez,

climático" o "crisis medioambiental" [4], y, claro está, la revolución técnico-científica de las telecomunicaciones.

De repente las vastas distancias del mundo se estrecharon, no solamente entre sujetos individuales y colectivos, sino también entre esos sujetos y las innumerables fuentes de información. Poco a poco, se fueron acumulando fuentes, datos, recursos y opiniones de muchas índoles en ese *lugar no-lugar* que hoy en día conocemos como *la nube*, accesible en potencia y, por lo menos superficialmente, a cualquiera que pueda tener acceso a internet. Y con esa facilidad de acceso, vino también implícita la seductora promesa de menos esfuerzo por igual o superior beneficio. Una simple búsqueda en Google ahorra el esfuerzo de caminar a la biblioteca o desgastar tiempo de búsqueda en anaqueles y archivadores, miles de videos y tutoriales en Youtube ahorran el esfuerzo de buscar a un experto y rogar que su humor sea el apropiado para compartir su conocimiento sobre casi cualquier cosa. ¿Cómo evitar —o mejor aún: *¿cómo no desear?*— que este trato tan optimo y racional no terminara inundando los nuevos intereses pragmáticos de nuestra disciplina jurídica, en un mundo donde la web —y hoy por hoy, a IA— se han convertido en la *mathesis* de la episteme contemporánea?

4. Los *Cyborgs*

En un segmento de tiempo tan temprano para este nuevo mundo digital como 1984, la brillante Donna Haraway (1984) con su ácida argumentación irónica y su lenguaje *cyberpunk*, avizoraba el sino de nuestro presente: nos hemos convertido en *cyborgs*, organismos híbridos entre biología y maquinaria; algo que hasta el momento, solo existía en el campo de la ciencia ficción futurista, en la guerra de las galaxias y los cuentos de Isaac Azimov. Donna se refería, en ese momento, de una manera más específica a los avances biomédicos y la forma en que empezaba a desenroscarse el escenario geopolítico de las guerras tecnológicas, algo que no es de extrañar, si se tiene en cuenta que muchas de las cosas a las que hoy en día le damos un "inocente y despreocupado" *uso civil*, tuvieron un origen militar, aviones e internet incluidos. Y tampoco es nada raro que cuarenta años de avances tecnológicos, hayan alcanzado y rebasado el análisis de Haraway. De todos modos, pienso que el nutritivo campo de los estudios de *Ciencia, Tecnología*

---

cultura—, ya no se hizo necesario ocuparse de averiguar ni justificar lo que el derecho imperante *es* o *no es*, porque ya no hay discusión sobre eso.

(4) Esta inminente —y lógica— consecuencia de siglos de relaciones disfuncionales con la madre tierra —especialmente dos siglos de industrialismo capitalista—, empezó a acaparar —o por lo menos a ocupar un lugar de peso en— las agendas investigativas de las principales academias del mundo; algo cuyo manejo tendencioso por parte de muchos círculos de poder —especialmente los más conservadores—, ha sido usado para dar por "clausurados", muchos de los debates más álgidos hasta el momento, entre ellos los que ponían en tensión las desigualdades y desequilibrios de poder en el derecho, así como los intereses y particularidades tras el sujeto abstracto-universal de los lenguajes jurídicos.

y Sociedad —STS por sus siglas en inglés—, en el que se ha alineado a Haraway y a otros pensadores, ha sentado bases maravillosas para el análisis de estas transformaciones.

En el marco de lo descrito anteriormente acerca de la tecnologización del derecho en el mundo digital que vio morir a un milenio y nacer al siguiente, resultan particularmente fecundas algunas aplicaciones que podemos encontrarle a la conocida Teoría del Actor-Red —ANT por sus siglas en inglés—, desarrollada en los años 80 por Michel Callon y Bruno Latour y popularizada por este último en su obra *Nunca fuimos modernos: ensayos de antropología simétrica* ([1993] 2022). Latour controvierte la clásica dicotomía entre naturaleza y sociedad —o *naturaleza y cultura*, o *naturaleza y humanidad*, si se prefiere— fundacional de los presupuestos ontológicos de la modernidad occidental, en tanto que heredada del antropocentrismo renacentista e ilustrado, para situar un universo de continuidades en el que lo natural —lo biológico— y lo social o artificial —objetos, maquinas, prácticas, conceptos, etc.— se articulan y complementan ontológicamente, o, para ser precisos con el termino usado por el autor, *se enactuan*, configurando seres y realidades concretas. En consonancia con lo planteado por Annemarie Mol en *El cuerpo múltiple: Ontología en la práctica médica* (2021) y por John Law en *Actor network theory and material semiotics. The new Blackwell companion to social theory* (cita 2008), *el ser* de cualquier entidad deja de ser considerado dentro de las estrechas márgenes de las clasificaciones esencialistas de la modernidad occidental —a su vez cruzadas diametralmente por la inferiorización de lo no-humano respecto a lo humano— y empieza a considerarse en una profusa red de relaciones con otras entidades —otros *actores* o *actantes*— a las cuales permea y de las cuales se deja permear en una permanente danza de hibridación.

Como en otros casos, los ejemplos de estos planteamientos pululan en la historia. Baste recordar las relaciones totémicas que muchas sociedades han tenido con los animales, incluida la propia sociedad occidental y sus auto percibidos antepasados en la antigüedad y el medioevo —caso concreto, por ejemplo, de los caballos y su trasposición dentro de los órdenes sociales, engendrando estamentos como los *hippeis* griegos, los *equites* romanos y los caballeros medievales—; o con las plantas que aún hoy lucen en el pabellón nacional de muchos grupos humanos y siguen siendo, en buena medida, la base de la medicina o del embellecimiento estético en la industria de los cosméticos. Una sola taza de café consumida en Europa o Norteamérica —por poner un ejemplo representativo, porque también podríamos hablar de un rímel para los ojos— le pone en conexión con una cadena casi infinita de actantes desde la planta sembrada en una sierra montañosa suramericana, africana o asiática; los campesinos que la criaron y cosecharon, los insumos necesarios para su procesamiento, etc. Y esa taza bebida

incide en la personalidad del consumidor, determinando decisiones e, incluso, concepciones.

Podríamos seguir ejemplificando escenarios similares, pero el que ahora nos interesa es el de la tecnología y, espacialmente, sus repercusiones en nuestra disciplina jurídica. Bueno, al respecto extiendo la siguiente pregunta como motivadora para la reflexión: ¿Qué es o qué significa el celular para nosotros, seres humanos del siglo XXI? Creo que han quedado atrás las épocas donde podríamos considerar a estos aparatos como "simples accesorios" —algo que, de por sí, no existe en los marcos de la ANT— y tendríamos que reconocer que hoy en día es, literalmente, un apéndice de nuestro cuerpo. Hace ya rato que dejó de ser algo que usamos solamente para comunicarnos con otra persona que está lejos, puesto que hoy en día controlamos todo nuestro entorno a través de él: hacemos pagos, pedimos transporte, administramos nuestro tiempo, vigilamos nuestro hogar, nos ubicamos mediante él, nos informamos sobre el mundo, etc. Tal vez por eso uno de los mayores terrores de nuestra época es que se descargue sin el cargador o una fuente de energía cerca. Gran parte de este texto fue escrito en un celular o escrito gracias a las consultas realizadas a través de celular. La enacción humano-celular es una de las evidencias más sonoras que hoy en día podemos hallar acerca de nuestra condición de *cyborgs*.

Para pensar en ello, teniendo en mente de manera específica nuestras actuales relaciones con la IA, propongo un pequeño ejercicio, clásico tal vez, para los estudiantes de antropología: *ver a través de los ojos del otro*. Supongamos por un momento ser un extraterrestre, o mejor aún, ser una persona del pasado que por un raro acto de magia —científica o no-científica— puede asomarse, aunque sea brevemente a nuestra época y nos descubre haciéndole preguntas o dándole órdenes a un curioso aparato rectangular que cargamos en la mano y del que no podemos permanecer mucho tiempo distanciados. ¿Qué pensaría? Nos atrevemos a sugerir que tal vez por su mente cruce la posibilidad de pensar que tenemos a alguien encerrado en ese pequeño rectángulo; puede ser a otra persona o quizás incluso, hasta un espíritu o a un dios. Eso sí, algo o alguien supremamente sabio a quien acudimos para que nos revele la respuesta de algo que desconocemos; o bien alguien a quien hemos esclavizado para nuestro servicio. Mirando desde esta óptica, para este observador hipotético no tendría por qué haber diferencia entre lo que hace cualquier sabio indígena consultando al fuego o los espíritus mayores de la madre naturaleza a través del mambeo de la hoja coca o a través de la ayahuasca y lo que hace cualquier sujeto occidentalizado preguntándole a la IA a través del celular por las respuestas de un examen o por la jurisprudencia de un tribunal para una demanda que debe elaborar.

Considero este ejercicio muy útil como antesala para plantear que hoy en día asistimos a unos escenarios muy variados en cuanto a la utilización de la IA en el campo jurídico, que en gran medida son similares a las que se

están dando en otras áreas del saber y prácticas sociales. Podrían resumirse algunas problemáticas de estos usos en dos grandes grupos.

Por un lado están quienes han caído rendidos a los pies de la IA y sus promesas de conocimiento objetivo con poco esfuerzo. Para estas personas, el desarrollo de la IA ha significado el regreso de un mesías largamente esperado, que suspende debates y permite concentrarse puramente en la carpintería de la profesión. Los teléfonos inteligentes de estas personas rebosan de aplicaciones como Chat GPT, DeepSeek, Gemini, son asiduos usuarios de Copilot y muchas otras cuyo conocimiento puntual se me escapa; a las que encargan todo tipo de labores consultivas, ordenadoras y hasta creativas. Dentro de este grupo estarían las personas que hoy en día encargan a estas tecnologías la realización de ensayos reflexivos, presentaciones de Power Point, derechos de petición, acciones de tutela/amparo, e incluso acciones jurídicas de más grueso calado; materiales a los que después les añaden una portada o rotulación que hace pasar por propia la argumentación de la máquina, a la que varían algunas líneas o palabras para eludir los *software* anti plagio o para adaptarla a las particularidades de un caso concreto. No realizan una revisión consciente y profunda de las elaboraciones de la IA porque hay una confianza —o bien una pereza— desbordada. "Si la maquina lo hace ¿por qué habría yo de perder tiempo en ello?", podría ser una frase representativa de este grupo, del mismo modo que olvidamos el procedimiento para hacer una división por dos o más cifras —cosa que debimos tener super clara en la escuela primaria— porque la calculadora se encarga de hacerlo.

Podríamos decir que en este primer grupo de personas, adaptando el famoso concepto marxista, opera un *fetichismo de la tecnología*, es decir, el olvido intencionado o instalado de que detrás de las respuestas entregadas por la IA hay otros actantes iguales a nosotros —otros *cyborgs*— que han configurado o determinado, de modos muy particulares, la información que estamos consultando y a la que asignamos el estatus de verdad objetiva. Ello ocurre así porque esta nueva *mathesis*, mediante la cual entendemos el mundo, ha heredado la no grata tendencia de universalizar particularismos, especialmente los hegemónicos.

En su magnífico texto *Al margen de Europa: pensamiento poscolonial y diferencia histórica* (2008), el historiador indio Dipesh Chakrabarty elabora una contundente y sensata crítica al eurocentrismo —sin abandonar o descartar a priori las grandezas y utilidades alternas del pensamiento europeo—, no tanto en un sentido político o económico, sino en los complejos planos de la subjetividad, incluido, por supuesto, el terreno del conocimiento. En el marco de ello afirma que, a pesar de que las naciones *no-occidentales* obtuvieron su independencia política de los imperios coloniales, la tradición intelectual europea siguió siendo la única viva dentro de las academias del mundo. Esta afirmación de Chakrabarty, tiene como trasfondo su oficio

como historiador, por lo que está constituida desde las inquietudes de esta disciplina, pero creemos que es análoga a lo que sucede con la disciplina jurídica en donde seguimos enseñando a los estudiantes de Introducción al Derecho [5] —entre otras cosas— que existen dos tradiciones jurídicas: la continental y la anglosajona, ambas, por supuesto: europeas, particularizadas de diversos modos en todos los rincones del mundo a donde llegaron por medio de la violenta acción colonial. Nada raro entonces que al consultar la IA sobre algún asunto jurídico, la única matriz epistémica disponible desde la que podría emanar cualquier tipo de información, sean las culturas jurídicas europeas, en tanto que, por efecto de la *colonialidad del poder* imperante en los escenarios poscoloniales, fueron *trasplantadas* (Michaels, R, 2013) o *superpuestas ideológicamente* (García, A, 1971) a las realidades sociales de muchos lugares del mundo y constituyen la base jurídica sobre la que se configuraron los Estados-nación de las naciones no-occidentales, sus ordenamientos normativos y sus prácticas jurídicas.

Un segundo grupo de personas, me parece, han asumido una actitud apocalíptica frente al advenimiento de la IA. Tal vez movidos por un miedo subconsciente y cinematográficamente instalado por la *skynet* de *Terminator* o las máquinas de *Matrix*, parecen ver la ruina de la humanidad en la importancia que la IA va adquiriendo en la vida cotidiana de la gente, así como en las cada vez más sorprendentes capacidades que estas tecnologías van logrando. Ello, además, sin valorar los conocimientos, los esfuerzos y las buenas intenciones que muchos ingenieros y desarrolladores empeñan en dichos desarrollos. Es posible detectar muchos argumentos en las personas que asumen esta actitud pesimista y fatalista frente a la utilización de tecnologías como la IA en los campos del derecho —y de la vida en general—, pero creo que el más reiterativo, o bien, aquel en el que se articulan los demás argumentos, es el temor de que la IA pueda llegar a reemplazar a los operadores jurídicos humanos, especialmente en asuntos judiciales.

Imaginemos un *juez robot* al que se le ingresarían los elementos materiales probatorios aportados por las partes, que este valoraría en un santiamén, arrojando una sentencia exacta, puramente racional, objetiva, que toma en cuenta todas las alternativas existentes y halla siempre la mejor solución posible, es decir, la que menos derechos vulnera —algo como lo que Alexy (1993) llamó *mandato de optimización*—. Este escenario, que para muchas personas es un sueño ideal por el que trabajan, para otros adquiere un tono dramático. La fundamentación de este temor radica en que, suponiendo que fuera posible —y probablemente algún día lo será— ¿para qué necesitaríamos entonces jueces humanos, habitados por prejuicios de

(5) Una asignatura clásica del primer semestre de la carrera de derecho en las universidades colombianas, equivalente a las cátedras iniciales de Teoría General del Derecho en otros lugares del mundo.

diversas índoles, potencialmente corruptibles, movidos por toda clase de intereses, algunos incluso hasta mal preparados para la labor judicial, incapaces de considerar todas las particularidades del caso, inspirados por diferentes concepciones de justicia? Quienes tienen este miedo instalado, claro está, ven en ello el acabose de la profesión judicial e incluso de la vida profesional de otros tipos de operadores jurídicos. Pensemos en los abogados litigantes: Si la IA tuviera —¿tiene?— la capacidad de redactar un derecho de petición o cualquier tipo de acción judicial, elaborar una liquidación laboral o de una sociedad conyugal, establecer la partición en una sucesión o indicar el procedimiento que debo realizar para conseguir una consecuencia jurídica ¿para que necesitamos contratar a un abogado? En las actuales circunstancias legales es claro que hay muchos asuntos jurídicos en los que no se puede acudir en propia representación, pero en nuestro hipotético futuro digital, el acceso a estas posibilidades, de seguro incidiría en la legislación que establece esta limitación. Y ejemplos similares podrían trasladarse a la elaboración de normas jurídicas, el ejercicio de cargos públicos e, incluso, a la academia, dejando en el más miserable desempleo a los abogados del mundo.

Creo que este tipo de prevenciones se ponen en mente quienes tienen este miedo instalado, un miedo, que me parece un poco exagerado, pero que no es del todo infundado. Ya en el siglo XIX, ante la maquinaria industrial a base de vapor, Marx vaticinaba que el capitalismo terminaría creando un ejército de desocupados a través de la tecnificación de la maquinaria. Si conversamos de esto con personas que trabajaban en las entradas o puestos de pago de parkings en los centros comerciales, o con aquellos que trabajaban en tiendas de alquiler de videos o venta de música, y si consideramos que los sindicatos de actores en la industria cinematográfica norteamericana, ya han presentado acciones judiciales por la utilización de IA para la creación de personajes en algunas producciones, creo que seremos conscientes de que ese miedo ya tiene bases materializadas. ¿Estaremos los abogados entre los próximos?

5. ¿Qué hacer con la IA y su creciente vigencia en los campos del derecho?

Mi intención con estas líneas no es la de demostrar quien está equivocado o quien no entre los fetichistas tecnológicos y los pesimistas apocalípticos. No tengo ni los conocimientos a fondo para poder hacerlo, ni los recursos para poder comprobar mis afirmaciones —ello me exigiría ser experto en informática... cosa que no soy— y tampoco poseo el espacio dentro de este texto. Lo que quiero sembrar aquí es una semilla de duda, de inquietud e incomodidad, algo para lo que siempre ha sido útil la filosofía. El rol que quisiera jugar aquí es el de el idiota de Deleuze (Becerril, D. X. G. 2020), el

que ante la huracanada velocidad en que funciona el mundo contemporáneo, va a otro ritmo más pausado; el que cuando todo el mundo parece haber decidido ir en una dirección, emerge con la inesperada pregunta *¿y qué pasa si vamos en este otro rumbo*?

Lo que me parece hay, y se puede tornar preocupante, son buenas dosis de exageración entre quienes confían ciegamente en la IA y quienes temen su potencial transformador en las vidas que conocemos. Y creo que ambas posturas están relacionadas con lo que se planteaba más arriba: un olvido a veces inconsciente y otras intencionado, de que en el mundo actual el relacionamiento del ser humano con las tecnologías de la informática y la telecomunicación, no es un relacionamiento entre dos entidades ontológicamente totalizadas, autocontenidas y perfectamente diferenciables —el ser humano por un lado y la maquina por otro—, sino una hibridación fruto de la cual los seres humanos estamos en gran medida maquinizados y las maquinas cursan continuos y densos procesos de antropomorfización. Es decir: los *cyborgs*.

Esto, como también ya fue insinuado, debería terminar empujando el centro de las inquietudes filosóficas al respecto, de nuevo hacia el no tan nuevo problema de la objetividad. Tanto los fetichistas tecnológicos, como los pesimistas apocalípticos, desde polos opuestos, acaban situando a las máquinas y su inteligencia artificial como los portavoces autorizados de un razonamiento universal; unos porque adoran la alta capacidad de estas tecnologías para procesar cantidades de información en poco tiempo y otros porque, precisamente esa misma capacidad, se torna amenazante frente a las "limitaciones" del pensamiento humano convencional. Prácticamente afirman que asuntos como la administración de justicia son demasiado importantes para correr el riesgo de dejarla en manos de seres tan inestables como los seres humanos. En ambas alternativas, emerge triunfante una sentencia común, que además se presume obvia e indiscutible: las maquinas piensan mejor que los seres humanos, porque las maquinas piensan objetivamente.

En el capítulo siete de su también célebre *Ciencia, cyborgs y mujeres. La invención de la naturaleza* Donna Haraway (1995) nos presenta un seminal capítulo —el número siete, para ser exactos— que titula "Conocimientos situados: la cuestión científica en el feminismo y el privilegio de la perspectiva parcial" en el que desarrolla una crítica contundente a las dos posiciones epistemológicas cuyo enfrentamiento ha dominado las más gruesas discusiones del pensamiento moderno: el objetivismo y el relativismo. Haraway ataca a las dos por igual, planteando que, a pesar de su dicotomía, son coincidentes ya que, el objetivismo —en su manifestación científica moderna— pretende "mirar todo mientras no se habita ningún lugar", mientras que el relativismo —en su manifestación lingüística posmoderna— insinúa poder "mirar todo desde todos los puntos posibles", pero sin habitar tampoco nin-

guno. Para Haraway, el objetivismo moderno pretende que su mirada es la única válida porque está construida desde un único lugar de privilegio y niega todo lo que no habla desde ahí. El relativismo asigna esa superioridad moral a cada perspectiva —con la consecuente posibilidad de negar la mirada de los otros—. En suma, *los dos impiden ver bien*, ambos son "trucos divinos" … *miran sin estar*.

La principal y más dramática consecuencia de esto, según Haraway es que ambos le rehúyen a la responsabilidad del acto de mirar —que es siempre algo encarnado— tiran la piedra y esconden la mano. Las problemáticas utilizaciones de la IA de las que he hablado en el presente texto —porque siento que son bastante comunes en nuestras sociedades actuales y, por ello mismo, caracterizan el desenvolvimiento de esta tecnología en los campos del derecho—, tienen en común el posicionamiento de la IA como un recurso para eludir la responsabilidad personal por los argumentos jurídicos esgrimidos. Bien sea por confianza ciega en la palabra de la IA, por pereza, por urgencia, por desconfianza en el pensamiento de otros u otras razones; la IA se ha convertido en el as bajo la manga o la verdadera instancia de cierre para los operadores jurídicos —o por lo menos está en camino de convertirse en ello, de seguir abundando utilizaciones como las descritas—.

Este texto fue escrito inspirado en posturas críticas frente a las maneras como se está usando la tecnología en el mundo jurídico contemporáneo, entremezclando un poco de escepticismo frente a la fascinación de los fetichistas tecnológicos y un poco de moderación frente a los temores de los pesimistas apocalípticos. No está escrito para impulsar la no—utilización de la IA en el derecho—algo poco lúcido, —sobre todo en el mundo contemporáneo— porque creo que, si algún recurso existente puede facilitar el trabajo y/o enriquecerlo aportando puntos de vista interesantes, mal haríamos en cometer la necedad de no usarlo. Este texto es, simplemente, una invitación a la, quizás olvidada, duda metódica y a ponerla en práctica concretamente, a través de la *encarnación de la inteligencia artificial*, a su abordaje como una dimensión más de las relaciones híbridas en las que se sostiene la existencia de nuestro universo, sin esencialismos; a la posibilidad de estudiar su genealogía para poder entender los lugares de enunciación hegemónicos que se mueven tras sus respuestas, pero también la posibilidad de instalar las disputas contra hegemónicas en su seno —hasta donde esto sea posible—.

Presenta retos, por supuesto, para cualquiera que se adentre o se desenvuelva en las disciplinas jurídicas en estos tiempos. Para los fetichistas tecnológicos, el reto de profundizar en los contextos en los que opera la utilización de la tecnología; y para los pesimistas apocalípticos, el reto de formarnos como abogados, con un mayor conocimiento en la informática y demás detalles técnicos sobre cómo opera la IA para así poder operarla con mayor efectividad y autonomía.

Creo, en suma, que es necesario que, respecto a la IA, al igual que de cualquier otra posibilidad de elaborar o dinamizar el conocimiento, es necesario preguntarnos en todo momento: ¿quién habla? ¿desde dónde habla? ¿por qué habla? y ¿para qué habla?

Bibliografía

Alexy, R. (1993). *Derecho y razón práctica* (Vol. 1). México: Fontamara.

Becerril, D. X. G. (2020). Desterritorializaciones a partir de las anomalías: el idiota y la nada como lugares críticos según el pensamiento de Gilles Deleuze. *Revista de filosofía*, *52*(148), 166-188. Ingold, T. (2020). *Antropología:¿ por qué importa?*. Alianza editorial.

Botero Uribe, D. (1999). Teoría social del derecho. *Editorial UN*.

Chakrabarty, D. (2008). Al margen de Europa: pensamiento poscolonial y diferencia histórica. In *Al margen de Europa: pensamiento poscolonial y diferencia histórica*

García, A. (1971). *Dialéctica de la democracia*. Cruz del Sur.

Geymonat, L. (1985). Historia del pensamiento filosófico y científico.

Grosso, J. L. (2017). *La semiopraxis popular como praxis crítica*. Universidad Nacional de Educación del Ecuador.

Haber, A. (2017). *Al otro lado del vestigio: políticas del conocimiento y arqueología indisciplinada*. Editorial Universidad del Cauca.

Haraway, D. (1984). Manifiesto cyborg. *Ciencia, tecnología y feminismo socialista en el siglo XX.[En línea] http://manifiestocyborg. blogspot. com*.

Haraway, D. (1995): Ciencia, cyborgs y mujeres. La invención de la naturaleza, Madrid, Cátedra, 313-346.

Mejía Quintana, O., & Jiménez, C. (2005). Nuevas teorías de la democracia de la democracia formal a la democracia deliberativa. *Colombia internacional*, (62), 12-31.

Michaels, R. (2013). "One size can fit all"–some heretical thoughts on the mass production of legal transplants. In *Order from Transfer* (pp. 56-78). Edward Elgar Publishing.

Mol, A. (2021). *El cuerpo múltiple: Ontología en la práctica médica*. Universidad de los Andes.

Latour, B. ([1993] 2022). *Nunca fuimos modernos: ensayos de antropología simétrica*. Siglo XXI editores.

Law, J. (2008). Actor network theory and material semiotics. *The new Blackwell companion to social theory*, 141-158.

Lazzarato, M., & Rodríguez, P. E. (2006). *Políticas del acontecimiento* (pp. 81-105). Tinta limón.

¿Hay derecho y afectos en el Río de la Plata? Un giro para leer la caja negra

Por Diego Martín González Camejo (*)

Resumen: Este artículo propone un ajuste de enfoque para leer la irrupción de la inteligencia artificial en el campo jurídico rioplatense. La hipótesis que sostengo es que el campo de derecho y literatura sigue siendo valioso, pero resulta insuficiente para comprender la "caja negra" de la gobernanza algorítmica. Cuando los criterios de decisión son opacos, lo primero que queda visible son sus huellas afectivas en los cuerpos y en los ambientes sociales. Por eso sugiero un viraje hacia derecho y afectos. El trabajo delimita el problema (automatización y opacidad), revisa aportes recientes y adopta herramientas de la teoría de los afectos (Massumi) para pensar los afectos como modulaciones prediscursivas recuperables políticamente.

Palabras clave: gobernanza algorítmica; derecho y afectos; caja negra; modulación afectiva; Río de la Plata.

Abstract: This article proposes a modest shift to read the spread of AI in legal practice in the Río de la Plata. The claim is: Law & Literature remains useful, but it is not enough to grasp the "black box" of algorithmic governance. When decision criteria are opaque, the first visible traces are affective —inscribed on bodies and circulating as public moods. Hence the suggestion to pivot toward Law & Affect. The paper frames the problem (automation and opacity), revisits recent contributions in Law & Literature (Olson), and adopts tools from affect theory (Massumi) to treat affect as prediscursive modulation that can be politically reclaimed.

Keywords: algorithmic governance; law and affect; black box; affect modulation; Río de la Plata.

(*) Profesor Adjunto (gr. 3) efectivo de Historia del Derecho, asistente (gr. 2) efectivo y profesor adscripto de Filosofía y Teoría General del Derecho (UdelaR). Magíster en Filosofía del Derecho (UBA). Abogado (UdelaR). https://orcid.org/0000-0002-3261-8222. Correo electrónico: diegonzalezcamejo@gmail.com.

1. GOBERNANZA ALGORÍTMICA: LÍMITES DEL ENFOQUE NARRATIVO EN DERECHO Y LITERATURA

La gobernanza algorítmica [1] decide en la caja negra: los criterios son inaccesibles, mientras sus efectos sensibles en los cuerpos sí se hacen visibles (por ejemplo, la exclusión automatizada de lxs migrantes, el cálculo de riesgo de reincidencia, la frustración de lxs abogadxs frente a las decisiones mudas de la máquina —*frustrated fidelity*—). Los algoritmos son omnipresentes en nuestra vida cotidiana, aunque no los percibamos directamente. Acumulan información, traducen nuestros recorridos en datos, registran nuestras decisiones de consumo, actúan de manera autónoma y orientan nuestra conducta con base en predicciones.

Hace aproximadamente veinte años, abrir un diario íntimo equivalía a vulnerar la intimidad de alguien. Más tarde, la contraseña de la computadora personal sustituyó la llave del pequeño candado. Hoy esa escena se condensa en el historial de búsquedas. Si alguien quisiera rastrear mis intereses, sería suficiente con vulnerar el código de acceso de mi móvil. Sin embargo, si pretendo mantener allí una reserva de intimidad, esa pretensión nace frustrada y se sustrae a la deliberación pública, porque la gobernanza algorítmica compromete por diseño esa privacidad. Mientras el diario íntimo era relativamente invulnerable sin la llave, la "lupa" de Instagram es estructuralmente vulnerable desde la máquina corporativa que la sostiene. El derecho a la intimidad se desplaza hacia un terreno estructuralmente vulnerable.

Para Daniel Cabrera, la problemática social de los algoritmos reside en que actúan de manera autónoma con modelos estadísticos y a partir de datos preexistentes. "Esa supuesta independencia del diseño de algoritmos y de la obtención y clasificación de datos con respecto a las bases sociales y culturales, está en los fundamentos del optimismo tecnológico y del pesimismo cultural" (Cabrera, 2024, p. 50). El desafío está en encarnar la inteligencia artificial, buscar la armazón que está detrás.

En el campo derecho y literatura, Chippy Abraham propone un abordaje de la gobernanza algorítmica a través de *El proceso* de Kafka. La visión de Kafka no anticipa la IA pero captura la ansiedad universal vinculada con la opacidad del derecho y el exceso de ritualismo (Abraham, 2025, p. 21) [2]. ¿Por qué *El proceso* se sostiene como ficción vital? Porque abre un campo crítico para los dilemas éticos de la IA en el derecho, tales como la auto-

(1) Por gobernanza algorítmica (*gouvernementalité algorithmique*) sigo a Rouvroy y Berns, quienes la definen como un modo de gobierno de conductas basado en la correlación estadística y la predicción automática, que actúa sobre los individuos a través de perfiles probabilísticos sin necesidad de normas explícitas ni deliberación pública (Rouvroy & Berns, 2013, pp. 165-166).

(2) Para un análisis situado de la opacidad del derecho cabe remitir a la obra de Carlos María Cárcova (2019).

matización de la gobernanza y la erosión de la agencia humana (Abraham, 2025, p. 24).

Mientras Cabrera destaca la autonomía estadística del algoritmo, Abraham recuerda que la literatura ya imaginó la opacidad institucional como estructura.

Por su parte, Bryan Ellrod advierte sobre las consecuencias de las tecnologías que combinan *data-driven* y *code-driven* en las decisiones jurídicas [3]. Su pretendida "aura de objetividad" oscurece cómo estas tecnologías están cambiando el terreno epistémico de los juicios jurídicos. Para el autor, "the role that judgement plays in the articulation of legal meaning and why the preservation of *human* judgement is necessary to the preservation of free societies" (Ellrod, 2025, p. 51). Ellrod recupera el concepto de *jurisgénesis*, propuesto por Robert Cover, e insiste en la necesidad de debatir el derecho y la IA en el campo *Law & Humanities* (Ellrod, 2025, p. 70).

En consecuencia, el enfoque narrativo, aunque necesario, se muestra insuficiente: lo que queda por leer son las huellas afectivas de la caja negra.

2. Planteo del problema e hipótesis

De allí la pregunta de investigación: ¿desde qué marco teórico y con qué conceptos puede el campo derecho y literatura abordar la gobernanza algorítmica cuando los criterios son opacos yosolo quedan afectos en los cuerpos?

En este artículo sostengo que, en el Río de la Plata, el campo derecho y literatura debe abrirse al giro afectivo. La gobernanza algorítmica, que opera en opacidad —parámetros y umbrales no expuestos—, activa y modula afectos como variaciones de intensidad que aumentan o disminuyen potencia, orientan la atención y disponen los cuerpos. Esa activación produce efectos también no discursivos: a veces no llega a sedimentar en el texto, pero ya reordena el campo de lo sensible. Por eso, además de leer discursos y textos literarios, interesa registrar las modulaciones antes de su sedimentación.

Propongo concebir los afectos sentidos como afectos activados, susceptibles de apropiación crítica. La opacidad algorítmica y la modulación afectiva reorganizan lo político; lo jurídico aparece como una modalidad de encuadre, entre otras. Por ejemplo, lo que se percibe como clausura apocalíptica del presente es, en rigor, la imposición naturalizada de ese reparto; la

(3) Ellrod analiza los modelos de inteligencia artificial utilizados en Estados Unidos, como COMPAS y DoNotPay. El primero como promesa de objetividad en la determinación de riesgo de reincidencia delictual y el segundo como promesa de reducción del *justice gap*.

teoría de los afectos abre un campo para nombrarlo, recuperar la potencia crítica de esas activaciones, ensayar reencuadres y escenas de resistencia frente a la gobernanza algorítmica.

3. El giro afectivo en derecho y literatura

En el Río de la Plata, los trabajos de Cecilia Macón se inscriben en la teoría de los afectos con abordajes filosóficos y literarios (Macón, 2022, pp. 285-291). En el plano de la teoría y la filosofía del derecho, el cruce derecho y literatura tiene un recorrido consolidado en la Escuela Crítica del Derecho argentina (Roggero et al., 2015, pp. 245-264). No obstante, la teoría de los afectos no figura aún como eje metodológico consolidado en derecho y literatura rioplatense.

En ese encuadre, María Jimena Sáenz consolidó una línea de trabajo en derecho y literatura que incluye una lectura crítica de la obra de Martha Nussbaum. En un artículo publicado en 2019, Sáenz reconstruye la "carrera intermitent" de la empatía en el movimiento, describe el giro hermenéutico y su afinidad con la hermenéutica de la sospecha, y plantea reabrir la pregunta por la empatía para tender un puente entre lecturas críticas profesionales y lectores comunes/lectura ordinaria: "volver a considerar la empatía (…) puede abrir nuevas preguntas teóricas vinculadas a la articulación del valor de la literatura y de los motivos que nos acercan a el", (Sáenz, 2019, p. 363).

Frente a los límites epistémicos de la práctica profesional, Sáenz reconoce en la empatía un motor teórico para profundizar la apertura del campo de derecho y literatura, aunque no propone directamente un viraje afectivo del campo. No obstante, me interesa resaltar el aporte de María Jimena Sáenz y que su producción se constituye en punto de apoyo para propiciar el giro.

En el Norte Global, Greta Olson se pregunta si la "Literatur" ha sido reemplazada por los Afectos o las Emociones en el binomio "derecho y literatur", abriendo el horizonte de *Law and Affect* (Olson, 2016, pp. 336-340).Si bien el movimiento se sostuvo en el énfasis narrativo-discursivo del fenómeno jurídico, la teoría de los afectos introduce una mirada epistémica que problematiza ese primado: acentúa lo no discursivo y lo no representacional. ¿Existen condiciones de posibilidad para ese viraje?

Olson argumenta que la categoría de discurso (tradición foucaultiana) es interpelada por filosofías feministas, estudios queer y perspectivas decoloniales. La instancia narrativa sedimentada en el discurso se ve así atravesada por el plano pre-personal de los afectos:

> "Access to affect, understood in the positive sense as a non-rational recognition of the other's suffering, allows one to take responsibility for this suffering rather than intellectualizing or narrating it causally. (Olson, 2016, p. 350) [(4)]".

Desde la propuesta de Greta Olson, propongo a los afectos como el "Otro" radical del derecho (el intruso, el extranjero que merece hospitalidad). Los conceptos en los que organizo el marco teórico proceden de la teoría de los afectos de Brian Massumi.

En síntesis, el canon narrativo-discursivo de derecho y literatura se potencia con un giro que lo resitúe como derecho y afectos. Mientras la gobernanza algorítmica decide en caja negra —criterios y pesos invisibles—, lo único plenamente accesible es el afecto sentido en los cuerpos. Este pliegue [(5)] (indisponibilidad de criterios/disponibilidad de afectos) abre en su asimetría un acceso político: leer los afectos como huellas públicas de decisiones opacas y usarlos como palanca contra su agenciamiento maquínico. Por ejemplo, cuando un sistema automatizado de riesgo deniega una medida alternativa sin mostrar cómo ponderó las variables, el operador jurídico cae en *frustrated fidelity* —obligado a cumplir, impedido de comprender— (Abraham, 2025, p. 10); ese afecto, doblez de la opacidad, es el punto de apoyo para impugnar, exigir trazabilidad encarnada de la IA y producir nuevos agenciamientos.

4. La teoría de los afectos de Brian Massumi

Desde la publicación de *The autonomy of affect* (1995) Brian Massumi ha generado una prolífica producción académica. Su versión de la teoría de los afectos parte de un desarrollo de la filosofía de Spinoza vía Deleuze. La teoría feminista y los estudios queer/cuir han incursionado en el giro afectivo a través de las obras de Sara Ahmed y Lauren Berlant,osolo para citar dos autoras centrales.

Para Massumi la noción de *affect* [(6)] toma diversas formas: "It's an event, or a dimension of every even" (Massumi, 2011, p. 37) [(7)]. A partir de la definición de Spinoza —una capacidad de afectar y de ser afectado—, Massumi señala las dos facetas del mismo evento: hay una afectación y esta sucede

---

(4) "El acceso al afecto, entendido en sentido positivo como un reconocimiento no racional del sufrimiento del otro, permite hacerse cargo de ese sufrimiento en lugar de intelectualizarlo o narrarlo causalmente".

(5) La imagen del pliegue la propongo para el cruce de la indisponibilidad epistémica de los criterios algorítmicos (caja negra) con la disponibilidad sensible de sus efectos en los cuerpos.

(6) Sobre la distinción entre afectos y emociones en el marco de las vertientes del giro afectivo, véase el prólogo de Cecilia Macón a la traducción al español de *Optimismo cruel* (Berlant, 2020, p. 13).

(7) "Es un evento o una dimensión de cada evento".

in-between. Las afectaciones sobre los cuerpos operan un pasaje, una transición que es sentida (Massumi, 1995, p. 37).

No hay identidad del afecto sino diferencias afectivas en el mismo evento. Los cuerpos afectan y son afectados, los afectos son transiciones sentidas en el cuerpo. Massumi propone el ejemplo de una alarma que cuando se activa impacta afecta los cuerpos. La respuesta de cada cuerpo será diferente. Las diferentes maneras de ser interpelado por el mismo fenómeno es lo definirá políticamente a los afectos (Massumi, 2011, p. 40).

Una afectación de los cuerpos, antes de ser codificada y, por ende, sedimentada en el lenguaje, opera en la corporalidad en un registro no lingüístico que es pre-personal. Es posible recuperar la reserva afectiva que no fue expresada en todas sus potencialidades y resignificarlas: una micro-política de micro-percepciones colectivas (Massumi, 2011, p. 41).

La intensidad de la respuesta afectiva al evento puede ser recuperado en su virtualidad política:

> "Micropolitics, affective politics, seek the degree of opennes of any situation, in hopes of priming an alter-accomplishment. Just modulating a situation in a way that amplifies a previously unfelt potential to the point of perceptibility is an alter-accomplishment (Massumi, 2011, p. 42) [(8)]".

En *The autonomy of affect,* Massumi sostiene que la apertura de los afectos a nuevas resignificaciones, su virtualidad, constituye su autonomía como categoría con potencialidad política. Si no fuera posible un escape de la captura, un resto, un exceso, sin apertura al infinito el universo sería pura entropía y muerte (Massumi, 1995, p. 100). En este encuadre, la libertad es la autonomía transindividual del proceso afectivo que engendra lo inesperado (Massumi, 2021, p. L).

Los afectos son sentidos en los cuerpos y admiten ser modulados. Dicha modulación supone la imposición de modulaciones de intensidad de los afectos en los cuerpos. Se trata de una dimensión de captura de los afectos. No obstante, gracias a su autonomía, la potencia política de los afectos siempre puede ser recuperada y resignificada. Frente al ajuste afectivo impuesto desde las modulaciones de la gobernanza algorítmica, por ejemplo, la política es el arte de la producción y emisión de signos interruptivos. El campo de pruebas de las potencias políticas son los afectos en su autonomía y virtualidad (Massumi, 2011, p. 41).

---

(8) "Las micropolíticas —políticas afectivas— indagan el grado de apertura de cualquier situación, con la esperanza de predisponer una realización-otra. Con tan solo modular una situación de modo que un potencial antes no sentido se amplifique hasta hacerse perceptible, eso ya constituye una realización-otra".

5. Del control de las conciencias a la modulación de cuerpos y afectos

Con el marco de la teoría de los afectos, Silvia Lindtner expone los resultados de un trabajo etnográfico realizado en China durante la pandemia. Propone un cruce comparativo de un paisaje urbano y otro rural para valorar el impacto de la IA. A la par del control biopolítico descrito por Michel Foucault y Giorgio Agamben, Lindtner sostiene la presencia de otros mecanismos que no se enfocan en la visión sino en la producción y cultivo de afectos (Lindtner, 2025, p. 1).

La autora recupera la noción de economías afectivas propuesta por Sara Ahmed y postula:

> "The obsession with surveillance clouds other forms of control that do not operate through vision [...] these other forms of population management tend to be more difficult to notice for they operate via the production and circulation of feeling" (Lindtner, 2025, p. 2) [(9)].

La gestión afectiva es una forma de control que nos direcciona hacia la participación antes que a la regulación o a la represión (Lindtner, 2025, p. 10). El análisis comparativo de dos entornos bien diferenciados despliega su hipótesis desde la continuidad de la gestión afectiva que impone formas de sentir a la población:

> "I have shown how affects circulate, for instance, when China's country side is rearticulated as a site of "positive energy"' or when data-driven population management is integrated into people's everyday lives via the cultivation of feelings of happiness, care and safety (Lindtner, 2025, p. 9) [(10)]".

En ambos entornos, el rural y el urbano, las formas de vigilancia habilitan una diferenciación. No obstante, Lindtner advierte la utilización por el Estado de metáforas lingüísticas y visuales que no se enfocaban en la vigilancia y el control de la población. En su lugar, promovían la producción de estados de agitación, sentido de compromiso, entusiasmo y ánimo emprendedor desde visiones prototípicas de "smart livin" y entornos de IA (Lindtner, 2025, pp. 6-7). En síntesis, para Lindtner la gubernamentalidad digital opera por modulación afectiva más que por vigilancia declarada. Los sistemas de IA introducen métricas y notificaciones que pr-disponen decisiones.

Lo que me interesa mostrar es que la vigilancia, como control de las conciencias, no logra explicar el ejercicio del poder que se despliega en el

---

(9) "La obsesión por la vigilancia encandila y deja fuera otras tecnologías de control no ópticas; esas modalidades de gubernamentalidad poblacional se vuelven menos visibles porque operan produciendo y haciendo circular afectos".

(10) "He mostrado cómo circula el afecto cuando, por ejemplo, el campo chino es rearticulado como un lugar de 'energía positiva,' o cuando la gestión poblacional basada en datos se integra en la vida cotidiana mediante el cultivo de sentimientos/afectos de felicidad, cuidado y seguridad".

campo afectivo. Existe una gobernanza algorítmica que opera en la sutileza y que Lindtner detecta en su análisis. Es una forma de gobernanza que no logra ser captada por la analítica del discurso y de los dispositivos de poder. Justo allí, la teoría de los afectos encuentra el campo de su potencia crítica frente a las nuevas formas de control coercitivo y su asombrosa adaptabilidad.

Conclusiones

El gobierno operado desde los algoritmos y la IA incorpora formas de control que exceden las categorías filosóficas centradas en la vigilancia (Foucault, Agamben). El discurso ya significa una instancia de sedimentación sensible de los efectos de esta gobernanza en los cuerpos.

En tal sentido, los afectos sentidos en los cuerpos representan eventos pr-discursivos que son modulados en sus intensidades por los agenciamientos maquínicos. Frente a la opacidad de la IA que trabaja con criterios que no se ofrecen a la deliberación pública, la evidencia disponible para la crítica filosófica está comprendida en los afectos, en su condición experiencial en los cuerpos.

La teoría de los afectos provee las herramientas necesarias para abordar la dimensión pr-discursiva de los efectos maquínicos en los cuerpos. A modo de ejemplo, una cartografía afectiva habilita la crítica filosófica que precede la sedimentación lingüística y representacional. La autonomía de los afectos constituye una llave crítica de apertura hacia sucesivas reactivaciones críticas.

Esa potencia de los afectos en los cuerpos desde la autonomía tiene un correlato teórico. Es el viraje que propongo para derecho y literatura: abrirlo a la autonomía de los afectos en la doble condición de habitar lo inesperado y de brindar hospitalidad a lo extranjero.

Referencias

Abraham, C. (2025). Automated Justice and the Performance of Law: Kafka's The Trial in the Age of Algorithmic Governance. *Law, Culture and the Humanities*, 1-24. https://doi.org/10.1177/17438721251351015

Berlant, L. G. (with Macón, C.). (2020). *El optimismo cruel* (H. Salas, Trad.). Caja Negra.

Cabrera, D. (2024). Lo algorítmico como teotécnica. *Cuestiones de Filosofía*, 10(35), 39-57. https://doi.org/10.19053/uptc.01235095.v10.n35.2024.17095

Cárcova, C. M. (2019). *La opacidad del derecho* (2a.). Astrea.

Ellrod, B. M. (2025). Programming utopia: Artificial intelligence, judgement, and the prospect of jurisgenesis. *Law and Humanities*, 19(1), 48-70. https://doi.org/10.1080/17521483.2024.2366090

Lindtner, S. (2025). Feeling like a State: Affect and Control in the Age of AI. *Proceedings of the Sixth Decennial Aarhus Conference: Computing X Crisis*, 1-12. https://doi.org/10.1145/3744169.3744182

Macón, C. (2022). Filosofía feminista y giro afectivo: Una respuesta ex ante. *Revista Latinoamericana de Filosofía*, 48(2), 283-303. https://doi.org/10.36446/rlf2022331

Massumi, B. (1995). The Autonomy of Affect. *Cultural Critique*, 31, 83-109. https://doi.org/10.2307/1354446

Massumi, B. (2011). Affect in the Key of Politics. *Identities: Journal for Politics, Gender and Culture*, 8(1), 37-44. https://doi.org/10.51151/identities.v8i1.250

Massumi, B. (2021). *Parables for the virtual: Movement, affect, sensation* (Twentieth anniversary edition). Duke University Press.

Olson, G. (2016). The Turn to Passion: Has Law and Literature become Law and Affect? *Law & Literature*, 28(3), 335-353. https://doi.org/10.1080/1535685X.2016.1232925

Roggero, J., Boyd White, J., West, R., Goodrich, P., Fish, S., Ost, F., Calvo González, J., Marí, E., & Cárcova, C. M. (2015). *Derecho y Literatura: Textos y contextos* (1a.). Eudeba.

Rouvroy, A., & Berns, T. (2013). Gouvernementalité algorithmique et perspectives d'émancipation: Le disparate comme condition d'individuation par la relation ? *Réseaux*, n° 177(1), 163-196. https://doi.org/10.3917/res.177.0163

Sáenz, M. J. (2019). Reconsiderando la empatía en el movimiento "Derecho y Literatura". *Revista Derecho del Estado,* 43, 351-372. https://doi.org/10.18601/01229893.n43.13

No son marginales: los errores de la inteligencia artificial y sus implicancias en materia de seguridad y soberanía

Por Juan Acerbi (*)

Resumen: Este artículo examina los efectos desproporcionados que pueden tener los errores mínimos en los sistemas de inteligencia artificial (IA), en particular cuando estas tecnologías son utilizadas en ámbitos críticos como la administración pública, la seguridad y la lucha contra el terrorismo. Aunque los márgenes de error suelen ser estadísticamente ínfimos —del orden del 0,002%—, sus consecuencias pueden impactar en miles de personas, provocando exclusiones injustas, fallas en la vigilancia y erosión de la confianza ciudadana. El análisis se inscribe en la continuidad de reflexiones previas planteadas en *Christchurch o el día después de la soberanía estatal* (Acerbi, 2021a), donde ya se advertía sobre la tensión entre soberanía estatal y poder tecnológico, ejemplificada en la confrontación de Facebook con Australia y en la reciente sentencia del Tribunal Supremo español sobre el algoritmo Bosco. El trabajo integra aportes que, en algunos casos, provienen de especialistas del campo de la computación o las matemáticas. con el fin de dar cuenta de la imposibilidad estructural de traducir plenamente la complejidad del lenguaje humano al lenguaje de programación, como la forma en que los algoritmos pueden institucionalizar desigualdades sociales. Asimismo, investigaciones sobre *adversarial examples* muestran cómo modificaciones imperceptibles en imágenes pueden engañar a redes neuronales robustas, lo que plantea riesgos en áreas sensibles como el reconocimiento facial o la seguridad vial. El artículo también aborda el papel creciente de la IA en la lucha contra el terrorismo, donde se emplea para monitorear comunicaciones, analizar patrones de comportamiento y detectar propaganda extremista.

Palabras clave: Inteligencia artificial; errores; soberanía; contraterrorismo.

(*) Universidad de Buenos Aires, juanacerbi@hotmail.com. Doctor en Ciencias Sociales por la Universidad de Buenos Aires, se ha desempeñado como profesor e investigador en universidades de Argentina, Brasil, México y España. Sus investigaciones abordan la pervivencia de las instituciones republicanas clásicas en las democracias actuales. Particularmente, se ha centrado en la figura del terrorista de cosecha propia como figura paradigmática de nuestros tiempos y el desafío que esta supone para la soberanía estatal.

Abstract: This article analyzes the disproportionate consequences of minimal errors in artificial intelligence (AI) systems, particularly when deployed in critical domains such as public administration, security, and counterterrorism. Although error rates may appear statistically negligible —on the order of 0.002%— their consequences can affect thousands of individuals, leading to unjust exclusions, failures in surveillance, and an erosion of public trust. The analysis follows earlier reflections from *Christchurch o el día después de la soberanía estatal* (Acerbi, 2021a), which highlighted the tension between state sovereignty and technological power, illustrated by Facebook's confrontation with Australia and the recent ruling of the Spanish Supreme Court on the Bosco algorithm. The work integrates contributions that, in some cases, come from specialists in the fields of computer science or mathematics, in order to highlight the structural impossibility of fully translating the complexity of human language into programming language, as well as the ways in which algorithms can institutionalize social inequalities. Likewise, research on adversarial examples shows how imperceptible modifications in images can deceive robust neural networks, raising risks in sensitive areas such as facial recognition or road safety. The article also addresses the growing role of AI in the fight against terrorism, where it is used to monitor communications, analyze behavioral patterns, and detect extremist propaganda.

Keywords: Artificial intelligence; errors; sovereignty; counterterrorism.

1. Introducción

En mayo de 2019, apenas dos meses después de haberse perpetrado el atentado en la ciudad de Christchurch, la entonces primera ministra neozelandesa Jacinda Ardern convocó a Jefes de Estado y representantes de grandes empresas tecnológicas al *Christchurch Call Summit to Action*. El objetivo era reducir la circulación de contenidos vinculados al terrorismo y a los discursos de odio en internet. Sin embargo, ni el entonces presidente de Estados Unidos ni los directivos de Facebook asistieron al encuentro, un gesto que fue interpretado como una señal de poder en la que se redefinía la relación de poderes entre los gigantes tecnológicos y los diversos Estados nacionales (Acerbi, 2021a).

Dicho conflicto resulta un antecedente útil para analizar ahora las consecuencias de un problema que surge a la luz de la propia naturaleza de los lenguajes computacionales y algorítmicos y que, lejos de tratarse de un problema marginal, representa un serio desafío con consecuencias que podrían ser devastadoras. Nos referimos a las pequeñas desviaciones o aproximaciones matemáticas que todo sistema de inteligencia artificial realiza. Aunque se suele insistir que, en términos estadísticos, los errores de la IA suelen ser insignificantes —del orden del 0,002%—, conviene insistir en que

dichos márgenes de error pueden ser devastadores cuando los programas se encuentran aplicándose en sistemas masivos o a poblaciones enteras. Una exclusión injusta de un beneficio social, una falsa alarma en un sistema antiterrorista o un error en la clasificación de una imagen puede tener un gran impacto en la vida de las personas, en las instituciones estatales y, en última instancia, en la capacidad del Estado para garantizar los derechos elementales en los cuales se fundamenta su existencia.

Ya en *Christchurch o el día después de la soberanía estatal* habíamos planteado cómo las tecnologías desafían a los Estados, y propusimos como ejemplo la llamada "Guerra de Facebook contra Australia", es decir, cuando la compañía norteamericana bloqueó al país oceánico impidiéndole no solo el acceso a las noticias sino también bloqueando el uso de la bases de datos oficiales, incluidos los del sistema sanitario, en represalia a la aprobación de una ley que obligaba a la compañía de Mark Zuckerberg a compensar a los medios de comunicación australianos por el uso de sus noticias (Acerbi, 2021a; BBC News, 2021). Ese episodio ilustró hasta qué punto una corporación privada podía imponer condiciones a un Estado soberano reconfigurando también el término tradicional de *guerra* (Acerbi, 2021a; Münkler, 2003). Ahora nos proponemos explorar el vínculo entre errores de IA, seguridad y soberanía, tomando como punto de partida lo ya dicho en nuestro trabajo anterior (Acerbi, 2021a) y el reciente fallo del Tribunal Supremo español que obligó al Estado a *abrir* [1] el algoritmo Bosco, utilizado para asignar el bono social eléctrico (El País, 2025). Asimismo, se integran aportes teóricos y críticos provenientes de las ciencias matemáticas o computacionales para resaltar los límites epistemológicos y prácticos de la IA, incluyendo la imposibilidad de traducir fielmente las sutilezas del lenguaje humano al lenguaje de programación.

2. El caso "Bosco": algoritmos, opacidad y justicia social

El caso "Bosco" constituye un hito en el debate sobre la transparencia algorítmica en relación con su aplicación sobre bases poblacionales. Se trata de un programa informático creado por la administración española para gestionar el Bono Social Eléctrico, un subsidio destinado a proteger a consumidores socialmente vulnerables. Durante años diversos informes y testimonios revelaron que familias que cumplían los requisitos legales eran

(1) Se trata del término que comúnmente se utiliza para indicar que el código fuente del programa debe darse a conocer. Esto permite, en principio, que los expertos puedan auditar la forma en la que el algoritmo toma las decisiones. De todas formas, es frecuente que el propio algoritmo se vuelva críptico (opaco u obscuro) para los propios programadores que lo han codificado. Al respecto, por ejemplo, consultar el trabajo de Cathy O´Neil.

excluidas por *errores* del sistema quedando sin posibilidad de apelación inmediata (elDiario.es, 2025).

Luego de muchas tensiones, denuncias y una creciente presión tanto mediática como de diversas ONG´s, en septiembre de 2025, el Tribunal Supremo español dictaminó que el Estado debía *abrir* el algoritmo que asigna el bono social entendiendo que se trata de un caso en el que lo que está en cuestión es un bien público gestionado con recursos estatales. La sentencia enfatizó que "la opacidad tecnológica no puede prevalecer sobre los derechos fundamentales de los ciudadanos" (El País, 2025).

En este sentido, el caso "Bosco" revela, al menos, tres tensiones de suma relevancia. Primero, el problema de la caja negra: ni los ciudadanos ni, en ocasiones, las propias autoridades comprenden por qué un algoritmo produce un determinado resultado. Segundo, el problema del error marginal: un 0,002% de fallos puede significar miles de familias excluidas de ayudas esenciales. Tercero, el problema de la soberanía: cuando las decisiones estatales dependen de algoritmos opacos, el Estado delega parte de su autoridad en procesos automáticos que escapan al escrutinio democrático. Lo cual se resume en palabras del informático Guillermo Simari (2023), quien afirma que "la sociedad civil termina siendo utilizada como banco de pruebas para sistemas que nadie puede auditar ni reproducir con certeza".

3. La naturaleza de los errores de la IA

Una cuestión que es importante resaltar es el hecho de que, cuando hablamos de errores en los sistemas de IA no nos referimos a simples accidentes o errores técnicos, sino que son inherentes a las propiedades estructurales de los modelos informáticos. Por ejemplo, en el caso de los sistemas de aprendizaje automático, estos funcionan con probabilidades cuya aproximación ronda en un 99,998% de exactitud. Esto significa que un algoritmo puede clasificar correctamente la mayoría de los casos, pero no debemos obviar que ese 0,002% restante se traduce en miles de errores cuando el sistema se aplica a millones de decisiones (Li et al., 2024).

También se debe destacar que los errores no se distribuyen de manera neutral. Datos incompletos o desbalanceados pueden amplificar desigualdades preexistentes, generando resultados discriminatorios en áreas sensibles como justicia, salud o seguridad (Mehrabi et al., 2021). Este fenómeno fue advertido tempranamente por Cathy O'Neil (2016), quien acuñó la expresión *weapons of math destruction* para referirse a algoritmos que, bajo la apariencia de ser objetivos, reproducen y amplifican injusticias sociales con efectos masivos y difíciles de revertir.

Incluso en marcos más controlados que los modelos de aprendizaje profundo, como la argumentación abstracta, el problema persiste. Simari

y Martínez (2019) muestran que la definición de criterios de similitud en marcos bipolares puede conducir a resultados diferentes y potencialmente conflictivos. Como ya hemos planteado anteriormente (Acerbi, 2021b), la objetividad formal de los algoritmos en materia de seguridad se ve erosionada cuando se conforman perfiles, *clustering* y *proxies* a partir de datos humanos, lo que evidencia un desfase estructural entre lo que los humanos comunican (con ambigüedad, metáfora, ironía, contexto) y lo que los algoritmos procesan (reglas, categorizaciones explícitas y parametrizadas). Es en esta línea argumental que hemos insistido (Acerbi, 2020) al enfatizar que la subjetividad y el lenguaje humano poseen límites intraducibles a sistemas formales, lo que refuerza la idea de que la opacidad y los pequeños desajustes en la definición de parámetros pueden tener efectos significativos en el razonamiento automatizado.

Además, la literatura sobre *adversarial examples* [2] demuestra con fuerza que incluso perturbaciones mínimas en imágenes pueden inducir a errores catastróficos si pensamos en materia de justicia o seguridad pública. Por ejemplo, Szegedy et al. (2014) demostró que pequeñas modificaciones invisibles al ojo humano bastan para engañar a clasificadores de sistemas robustos mientras que Goodfellow et al. (2015) explicaron que dicha fragilidad es consecuencia de la linealidad en espacios de alta dimensión. A su vez, el trabajo de Su et al. (2019) reveló que la modificación de un solo píxel puede cambiar la predicción de una red neuronal mientras que Eykholt et al. (2018) comprobó que pequeñas pegatinas aplicadas a una señal de tráfico podían hacer que un sistema de visión dejara de reconocerla. Finalmente, se comprobó que estas vulnerabilidades no son meros errores, sino que explotan variables aprendidas por los modelos (Ilyas, 2019). Y si bien se plantean mejoras que buscan elevar el nivel de las defensas de los sistemas de aprendizaje, las pruebas evidenciaron que estas fracasan ante ataques adaptativos (Tramèr et al., 2020).

4. Seguridad, terrorismo y riesgos algorítmicos

Como hemos anticipado, los errores de la IA no solo afectan la administración de beneficios sociales, sino que representan también una amenaza directa en materia de seguridad, justicia y soberanía estatal. Si las lecciones del atentado de Christchurch (2019) mostraron cómo los algoritmos fallan en identificar contenidos violentos al no bloquear la transmisión en vivo de una masacre porque sus sistemas automáticos no la consideraron "lo suficientemente *gore*" (Cheng, 2019; Bromell, 2021), deberíamos comprender que hablar de *errores* en el caso de la IA es un error en sí mismo. Es decir,

(2) Como veremos a continuación, *adversarial examples* se refiere a las perturbaciones o *ruido* que pueden incorporarse a una imagen con la intención de engañar a un sistema de aprendizaje automatizado.

debemos comprender que esos *errores* son inherentes a las lógicas y los lenguajes de programación y de la aplicación de dicha tecnología en gran escala.

Sobre este aspecto, hemos advertido (Acerbi, 2021b) que la vigilancia masiva y la inteligencia algorítmica como herramienta antiterrorista enfrenta el riesgo de volverse vigilancia masiva y, por ende, indiscriminada. Es decir, el rastreo y el etiquetado preventivo, sin indicios sólidos y con base jurídica, afecta directamente garantizar aspectos elementales y propios de un sistema democrático. En el mismo sentido, hemos resaltado (Acerbi, 2021c) el hecho de que el uso de *big data* y algoritmos predictivos en materia antiterrorista configura un espacio de prevención omnipresente que tensiona derechos y soberanía, generando riesgos de errores sistemáticos.

Además del uso de IA para detectar contenido en redes o moderar transmisiones violentas, las agencias de seguridad antiterroristas la utilizan cada vez más como herramienta preventiva en la que los límites de dicha prevención distan mucho de ser claros (Acerbi, 2021d). Farber (2025) analiza cómo diversos Estados dependen de sistemas de IA para identificar redes de propaganda extremista, coordinar contrataques informativos y optimizar inteligencia de señales; el informe de Chatham House (2022) describe cómo el reconocimiento facial, los análisis predictivos sobre grandes flujos de datos de vigilancia pública y el análisis de comportamiento ya forman parte de la estrategia antiterrorista de varios países. Estudios como el de Ezzeddine, Bayerl & Gibson (2023) revelan que buena parte de los ciudadanos encuestados reconocen la valía de estas aplicaciones, pero también reconocen que el uso de estas implica que no siempre esté garantizada la salvaguarda en materia de derechos y de privacidad personal. Trifunović (2025) documenta cómo los servicios de inteligencia usan IA para procesar imágenes satelitales, monitorear comunicaciones, cámaras urbanas y de drones sin otra autorización que la de la custodia de la seguridad nacional. En otros términos, podemos afirmar que, a pesar de los límites técnicos y lógicos propios de los lenguajes computacionales y los sistemas de procesamiento masivos de datos, cada vez sectores más amplios de la población mundial se encuentran bajo la vigilancia de estos a la vez que se argumenta que son pilares fundamentales para garantizar nuestros derechos, nuestra privacidad y nuestra seguridad.

5. Gobernanza y regulación

En el centro del debate en torno a la implementación de nuevas tecnologías informacionales e informáticas con fines públicos nos encontramos con el de la así denominada opacidad de la caja negra. Es decir, la imposibilidad que, incluso, el propio Estado tiene para auditar el programa que sustenta a un determinado algoritmo. En este sentido, el caso "Bosco" nos

aporta una perspectiva interesante al momento de pensar el lugar del Estado en la relación de poder que se plantea frente a las compañías informáticas que proveen estos servicios.

Recientemente, ha cobrado particular interés el denominado caso "Bosco" en el que el Supremo español reforzó la idea de que ningún algoritmo público puede permanecer cerrado a una auditoría. En este sentido, el Supremo comprendió que la transparencia de un sistema informático no es un accesorio técnico, sino una condición para garantizar los fundamentos de la vida democrática (El País, 2025). En este sentido, cabe destacar que el caso español no es único ya que diversos países han comenzado a elaborar guías y normativas que buscan garantizar la posibilidad de acceder al código de programación por parte de entidades públicas. La Unión Europea, por su parte, avanza con la *AI Act*, que clasifica sistemas según su nivel de riesgo (European Commission, 2021). A su vez, en Argentina, nos encontramos con la *Guía para el uso de IA en el sector público* (CIPPEC, 2025) la cual establece criterios de ética, transparencia y responsabilidad en la adopción de algoritmos gubernamentales.

Cabe señalar que existe también un problema que desafía los intentos de regular las nuevas tecnologías informacionales y que excede a la ya mencionada complejidad del algoritmo. Nos referimos a los *proxies* como recurso técnico para permitir el procesamiento masivo de datos humanos (Acerbi, 2021b). Así, tanto los *proxies*, como las categorizaciones automáticas y la construcción de perfiles hacen que las decisiones del sistema se vuelvan prácticamente irreversibles para todos aquellos que no tienen acceso a las sentencias del código. Regulaciones recientes enfatizan la necesidad de promover auditorías externas e independientes, pero esto es algo que puede presentar ciertos desafíos no solo técnicos sino también políticos si atendemos a la divergencia de criterios y opiniones que se suelen manifestar a la hora de abordar las regulaciones sobre internet y las grandes compañías informacionales especialmente en materia de seguridad pública, privacidad y del uso político de la información.

6. Conclusiones

Los errores de la IA, aunque estadísticamente pequeños, tienen consecuencias enormes en la vida social, política y jurídica. El caso "Bosco" ilustra cómo un algoritmo puede condicionar el acceso a derechos básicos. El atentado de Christchurch muestra los efectos letales de fallos en la detección automática de contenidos violentos. Lejos de ser anomalías corregibles, estos errores son estructurales y en gran medida imposibles de erradicar. Como advirtió O'Neil (2016), los algoritmos opacos no solo fallan: tienden a institucionalizar los errores y a perpetuarlos en bucles de retroalimentación negativa que afectan sobre todo a los sectores más vulnerables.

A la luz de la evidencia de los errores que acarrea el uso de estas tecnologías en grandes poblaciones, entendemos que no debería aspirarse a tener una IA infalible y, por ende, confiable en un ciento por ciento de los casos, sino en construir marcos jurídicos que busquen garantizar el acceso a los códigos y su interpretación en pos de obtener una mayor transparencia de los procesos, una minimización de las desviaciones procesales y decisionales y, finalmente, garantizando la soberanía de las instituciones estatales frente al poder de las grandes compañías.

Dicho de otro modo, y por todo lo expresado hasta aquí, entendemos que la promesa de la IA no puede sostenerse a costa de la renuncia del Estado a su función soberana de garantizar derechos y brindar seguridad. Si Christchurch marcó el inicio de una era en la que las tecnológicas desafiaban a los Estados, el desafío actual consiste en evitar que un error del 0,002% erosione los cimientos mismos de las instituciones democráticas.

Referencias

Acerbi, J. (2020). Una cuestión de humanidad: El arte y los límites de la subjetividad autoconsciente. *Cuadernos Filosóficos Segunda Época*, (16). https://doi.org/10.35305/cf2.vi16.68

Acerbi, J. (2021a). *Christchurch o el día después de la soberanía estatal*. Revista de Estudios Internacionales, 3(1), 94-110.

Acerbi, J. (2021b). *Terrorismo, tecnología y sociedad en el siglo XXI*. Revista Iberoamericana de Ciencia, Tecnología y Sociedad — CTS, 16(48), 11-33. https://ojs.revistacts.net/index.php/CTS/article/view/246

Acerbi, J. (2021c). El terrorismo como vórtice. Dilemas sobre derechos, soberanía y big data. En G. A. Dobratinich (Dir.), Derecho y nuevas tecnologías (pp. 247-266). Facultad de Derecho, Universidad de Buenos Aires.

Acerbi, J. (2021d, 23 de diciembre). Juan Acerbi: "El terrorismo no tiene una definición y eso es un problema". Agencia TSS. https://www.agenciatss.com.ar/juan-acerbi-el-terrorismo-no-tiene-una-definicion-y-eso-es-un-problema/

BBC News. (2021, 18 de febrero). *Facebook blocks news content in Australia amid media bargaining law row*. BBC News. https://www.bbc.com/news/world-australia-56106134

Bromell, D. J. (2021). *The terrorist attack on Christchurch mosques and the Christchurch call*. Institute for Governance and Policy Studies, Victoria University of Wellington.

Chatham House. (2022). *Artificial Intelligence Prediction and Counterterrorism*. Chatham House. https://www.chathamhouse.org/2019/08/artificial-intelligence-prediction-and-counterterrorism

Cheng, D. (2019). Jacinda Ardern says Facebook should have removed gunman's livestream immediately. *New Zealand Herald*. https://www.nzherald.co.nz/nz/jacinda-ardern-says-facebook-should-have-removed-gunmans-livestream-immediately

CIPPEC. (2025, mayo). *Guía para el uso de IA en el sector público en Argentina*. CIPPEC. https://www.cippec.org

El País. (2025, 17 de septiembre). *El Supremo obliga al Estado a abrir el algoritmo que asigna el bono social*. *El País*. https://elpais.com

European Commission. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. Brussels.

Ezzeddine, Y., Bayerl, P. S., & Gibson, H. (2023). *Citizen Perspectives on Necessary Safeguards to the Use of AI by Law Enforcement Agencies*. arXiv. https://arxiv.org/abs/2306.01786

Farber, S. (2025). *AI-Enabled Terrorism: A Strategic Analysis of Emerging Threats and Countermeasures in Global Security*. *Journal of Strategic Security, 18*(3), 320-344. https://doi.org/10.5038/1944-0472.18.3.2377

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1412.6572

Ilyas, A., Santurkar, S., Tsipras, D., Engstrom, L., Tran, B., & Madry, A. (2019). Adversarial examples are not bugs, they are features. *Advances in Neural Information Processing Systems, 32*. https://arxiv.org/abs/1905.02175

Kaplan, S., Handelman, D., & Handelman, A. (2021). Sensitivity of neural networks to corruption of image classification. *AI and Ethics*, 1-10.

Li, J., Yang, Y., Zhang, R., & Lee, Y.-C. (2024). Overconfident and unconfident AI hinder human-AI collaboration. *arXiv preprint*. https://arxiv.org/abs/2402.07632

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys, 54*(6), 1-35.

Moosavi-Dezfooli, S.-M., Fawzi, A., Fawzi, O., & Frossard, P. (2017). Universal adversarial perturbations. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1765-1773. https://arxiv.org/abs/1610.08401

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown Publishing Group.

Sadin, É. (2018). *La silicolonización del mundo. La irresistible expansión del liberalismo digital*. Caja Negra Editora.

Simari, G. R. (2023, julio 25). Inteligencia artificial: ¿Todos vamos a morir? *Caras & Caretas*. https://carasycaretas.org.ar/2023/07/25/inteligencia-artificial-todos-vamos-a-morir

Simari, G. R., & Martínez, M. V. (2019). Similarity notions in bipolar abstract argumentation. In A. Hunter & S. Polberg (Eds.), *Scalable Uncertainty Management. SUM 2019. Communications in Computer and Information Science* (Vol. 1194, pp. 208-222). Springer.

Su, J., Vargas, D. V., & Sakurai, K. (2019). One pixel attack for fooling deep neural networks. *IEEE Transactions on Evolutionary Computation, 23*(5), 828-841. https://arxiv.org/abs/1710.08864

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., & Fergus, R. (2014). Intriguing properties of neural networks. *International Conference on Learning Representations (ICLR)*. https://arxiv.org/abs/1312.6199

Tramèr, F., Carlini, N., Brendel, W., & Madry, A. (2020). On adaptive attacks to adversarial example defenses. *Advances in Neural Information Processing Systems, 33*, 1633-1645. https://arxiv.org/abs/2002.08347

Trifunović, D. (2025). *Intelligence and Counterterrorism*. *NSF – Journal Croatia, 26*(1). https://www.nsf-journal.hr/nsf-volumes/focus/id/1526

Wachter, S., Mittelstadt, B., & Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review, 41*, 105567.

Leyes, medidas y números en Heráclito y en la revolución algorítmica: materiales para una comparación de largo alcance

Por Rodrigo Oscar Ottonello (*)

Resumen: Contra la imagen de que las tecnologías digitales están tomando al derecho por asalto, como un agente exterior, este trabajo propone iluminar la importancia clave de las matemáticas en la ultrahistoria del derecho y mostrar cómo grandes problemas actuales, en especial el referido a la imposibilidad de las leyes positivas para poner límite a la proliferación de números cada vez más inalcanzables por el vehículo de la palabra, pueden encontrarse cifrados desde la filosofía antigua de Heráclito. Esto sirve de base para analizar con otro rigor las consecuencias del despliegue masivo del álgebra y de los algoritmos sobre los lenguajes políticos.

Palabras clave: Medida; Heráclito; nomos; logos; álgebra.

Abstract: Moving beyond the idea of the digital technologies as an assault force launched against the laws from its outside, this work proposes to show the key role of mathematics in law's ultra-history and present how some central contemporary problems —i.e. the once concerning the incapability of the system of the law to limit the production of numbers exponentially unreachable by the vehicle of the language made of words— can be tracked to the ancient philosophy of Heraclitus. This constitutes a base for an analysis of the consequences of the massive deploying of algebra and algorithms over political languages.

Key words: Measure; Heraclitus; nomos; logos; algebra.

(*) Investigador del Consejo Nacional de Investigaciones Científicas y Técnicas en el Laboratorio de Investigación en Ciencias Humanas de la Escuela de Humanidades de la Universidad de San Martín, profesor titular en Universidad de Belgrano y profesor en Universidad Nacional de las Artes. Formado como sociólogo (UBA), magíster en Sociología de la Cultura (IDAES-UNSAM) y doctor en Ciencias Sociales (UBA). Su investigación está dedicada a las agencias impersonales y anónimas desde el mundo antiguo a la actualidad. rodrigoottonello@hotmail.com

1. Sobre la distancia creciente entre derecho y tecnología

Las relaciones actuales entre el derecho y las tecnologías digitales son necesariamente problemáticas porque los cambios y las escalas de las técnicas basadas en el número avanzan a ritmo exponencial y son cada vez más difíciles de seguir para las instituciones jurídicas, basadas en la palabra. Durante el último siglo voces diversas como las de Max Weber (2008), Carl Schmitt (2024), Simone Weil (2021), Hannah Arendt (1993) y Giorgio Agamben (2006), entre muchas otras, coincidieron en ver en el avance tecnológico acelerado un riesgo mayor para la preservación de un Estado de derecho cada vez más compelido a dar lugar a la excepción jurídica. Hoy, cuando las principales corporaciones informáticas tienen más control de sus usuarios que el que los Estados nación tienen sobre sus ciudadanos, la desproporción en ese vínculo muda el orden global desde el imperio de la leyes (sostenido hasta el siglo XX) al imperio actual de las técnicas, con el mercado como mediador (al respecto, Ludueña Romandini, 2024). Esa evidencia creciente del problema no indica novedad; se lo encuentra ya formulado en las inquietudes de Aristóteles respecto a la crisis de la *polis* frente a la tecnología monetaria, que no reconoce límites para su multiplicación y acumulación (Borisonik, 2013). Incluso es posible ir más lejos, hasta testimonios en los que la administración de la justicia y la de las medidas van de la mano, casi como si fueran lo mismo. Ese acceso a la ultrahistoria de las relaciones entre las leyes y las matemáticas permite pensar la dinámica actual más allá de la lógica entre contrarios que frecuentemente obtura el diálogo entre disciplinas humanistas y progreso técnico.

2. Ultrahistoria de las relaciones entre número y ley

2.1. El cuidado de las medidas por la justicia

"El sol no sobrepasará las medidas [*métra*], si no, las Erinias, guardianas de Díke, lo alcanzarán", escribió Heráclito (DK 22B94). (1)

Cuando Cronos (el tiempo) castró a Urano (el cielo), la sangre de los testículos caída sobre Gea (la tierra) engendró a las Erinias (las furias), implacables vengadoras de los crímenes de sangre (Hesíodo, *Teogonía* 178-185). Heráclito, llamándolas, juzgó que las transgresiones de las medidas son igual de terribles que los atentados de los hijos a los padres soberanos.

Díke (justicia), a quien asisten las Erinias, es o bien una de las Horas engendradas por Zeus y Thémis (costumbre o ley natural), junto a Eunomia

(1) Todas las traducciones son propias. En la bibliografía se detallan los textos fuente y traducciones de referencia.

(orden o buena ley) y Eirene (paz) —hermanas "que cuidan las obras de los mortales hombres" (Hesíodo, *Teogonía* 901-903)—, o bien nació de Nómos (ley) y Eusebia (piedad) (Kern f. 159). Anaximandro escribió que el juicio [*díken*] a lo culpable de apartarse de lo ilimitado es expiado con la corrupción necesaria a lo largo del tiempo (DK 12B1). Heráclito escribió: "Díke caerá sobre los artífices [*téktonas*] de falsedades y sus testigos" (DK 22B28). Otro fragmento suyo: "la guerra es común, y la justicia [*díken*] discordia" (B80). Justicia también es una fuerza que se manifiesta en lo terrible para Parménides, que la consideró custodia del acceso a la verdad y la llamó "la de muchas venganzas" (DK B1. 14). Lo mismo Platón (*Leyes* 716a), que la describió como "vengadora contra quienes se alejan de la ley [*nómon*] divina".

En estos pasajes la justicia es la fuerza que condena a lo que se aparta de una regla fundamental, la resistencia espontánea que vence y colapsa todas las desviaciones. Esto significa que el poder de la justicia no es instaurar el orden, sino condenar lo que se aparta de él. Justicia es la forma en la que incluso lo que no cumple la ley la termina cumpliendo. De esa ley soberana la justicia es su brazo armado; Díke se sienta junto al trono desde el que Zeus contempla las obras humanas (Kern f. 23).

La ley soberana o divina que así es resguardada consiste en medidas. "Este cosmos, el mismo para todos, no fue hecho por ninguno de los dioses o los hombres, sino que era, es y será fuego siempre vivo, que crepita según medida y se extingue según medida" (Heráclito, DK 22B30). También: "Todo es alcanzado por el fuego que juzga y condena" (B66).

2.2. Legislación y matemática

Platón (*Protágoras* 343a-b) escribió que Sócrates dijo que los siete sabios —Tales, Pítaco de Mitilene, Bías de Priene, Solón, Cléobulo de Lindos, Misón de Quene y el lacedemonio Quilón— eran amantes de la educación espartana basada en frases lacónicas certeras como jabalinas, lo que les hizo acordar inscribir en el templo del Apolo délfico las célebres "conócete" (*gnothi sauton*) y "ninguno/nada demasiado" (*medèn agan*).

Esas y otras máximas de los sabios sobrevivieron en tradiciones textuales y grabadas en piedra en los edificios para los dioses. En una inscripción del reino grecobactriano, en Asia Central, frontera de la civilización helénica impulsada por Alejandro, se lee: "En verdad estas palabras sabias de los antiguos se encuentran / en las máximas bien conocidas de la Pitía divinísima [de Delfos], / de allí las registró cuidadosamente Clearco / para que brillen en el recinto de Kíneas [fundador de la colonia]" (Robert, 1968). Se trataría de Clearco de Solos, discípulo de Aristóteles al que también se atribuyen libros informados por sus viajes al Oriente. Lo que se conserva de la placa encontrada en la tumba de Kíneas es solo la base, donde además del epigrama citado se encuentran cinco máximas que coinciden con las últi-

mas de las ciento cuarenta y siete compiladas por Sosíades (de quien nada se sabe) y que se conservan por Estobeo (*Anthologium* III.I.173), lo que hace que los arqueólogos estimen que la parte superior perdida habría contenido el resto de los dichos de los sabios. También las inscripciones encontradas en Miletopolis, en Anatolia, concuerdan con las de Sosíades (Diels, 1920), lista que a su vez tiene correspondencias con las sentencias atribuidas a cada uno de los siete sabios por Demetrio de Falero —estudiante del Liceo de Aristóteles, gobernante de Atenas e ideólogo de la Biblioteca de Alejandría (también en Estobeo III.1.172). En esta imagen de todos los preceptos escritos en el templo de Delfos —con toda probabilidad no exacta, pero al menos consistente— destacan el respeto a los dioses, a los padres y a las leyes, el discernimiento claro de la diferencia entre lo propio y lo ajeno, el cuidado de no traspasar esos límites y el mal definido, más que como serie de actos nefastos que deben prohibirse, como extralimitación.

Según Demetrio la frase "la medida es lo mejor" (DK10.3a1) fue de Cleóbulo; a Solón atribuyó "nada demasiado"; a Tales "usa la medida" (DK18).

Entre estos sabios legisladores Tales, siempre considerado en primer lugar, fue además el introductor de la geometría en Grecia, traída desde Egipto (Proclo, *Comentario a Euclides*, 65), donde medida y proporción se usaban para el cálculo de impuestos variables según el cambio de tamaño de las tierras por efecto de las aguas del Nilo (Heródoto II.109). La medida aparece simultáneamente como instrumento técnico y moral.

Dicearco, según ha recogido Diógenes Laercio (I.40), dijo que los sabios, más que filósofos o sabios, fueron "inteligentes y legisladores [*nomothetikous*]". Así como Tales explicó las inundaciones del Nilo y cambió el curso de un río (Heródoto, I.75 y II.90), lo que hicieron los sabios fue establecer un orden para las leyes; no se dedicaron tanto a la creación de nuevas leyes como a la sistematización de las ya existentes, fijando por escrito lo que circulaba hasta entonces de forma oral (Nilsson 1961, p. 236). Esa noción de unas leyes ordenas tiene correspondencia con el tipo de orden general propuesto por las matemáticas de Tales, a quien se atribuye haber descubierto la posibilidad de conocer las medidas de cualquier cosa sin medirla directamente, mediante el cálculo de sus relaciones con medidas conocidas (Laercio, I.27 y Plutarco, *Banquete de los siete sabios* 247a); con ese ingenio todo se volvió medible, lo que según Weil (2006) constituyó la mayor revolución en la historia de la ciencia.

Se dictan entonces las reglas, pero las medidas *existen* porque aunque se cambien las reglas se mantienen las proporciones entre ellas; esa ley geométrica, más que algo que se sanciona, es la forma en que se impone el orden de lo que es. La fuerza de la ley es una razón independiente de la sentencia de una autoridad; corresponde a las medidas y proporciones que no se pueden ignorar.

Tales, en una carta a Ferécides (Laercio I.43), habría dicho que si algo es escrito —por ejemplo, lo referido a los dioses— se vuelve común a todos, porque ya no es palabra dirigida a un particular y tampoco depende más de quien la enuncia. Las matemáticas y las leyes son escrituras dirigidas a nadie y a todos.

Tales y los sabios dijeron esas leyes como legisladores, pero Ferécides, entre los primeros de una nueva clase hombres, se desentendía de las relaciones con los notables, así como de los viajes, y se concentraba solo en su escrito, el primer tratado sobre lo divino ofrecido a los griegos (Laercio, I.44). Con espíritu afín, Heráclito, que prefería jugar con niños antes que gobernar la ciudad —incluso recomendó que todos los adultos efesios se ahorcasen—, dejó su libro sobre la naturaleza en el templo de Artemisa (DK 22B121 y Laercio IX.1-6).

El modo de decir característico de las leyes, en que el razonamiento se sostiene por sí mismo y no por estar asociado a una sensación o recuerdo, nació, según Platón (*Epínomis* 977d), del conocimiento de los números y de su imitación por las palabras.

2.3. *Lógos* (razón/cuenta) del *nómos* único

"Ley [*nómos*] y voluntad es obedecer a uno", escribió Heráclito (DK 22B33). Esto no solo implica que la ley efectiva no es contradicha por otra ley y es única; en el escritor de Éfeso ese impulso es todavía más radical y apunta a que además de ser única, ella no es un conjunto de leyes, sino una sola ley, la obedecida por el cosmos entero, la que hace que el conjunto de todas las cosas, que puede parecer desordenado, revele su orden oculto (DK 22B124, 123 y 54).

Considerando esto se puede apreciar mejor la importancia que tiene en Heráclito la oposición con quienes saben muchas cosas pero no saben lo único que debe saberse, como ocurre a los ilustres Homero, Hesíodo, Jenófanes y Pitágoras (B17, 35, 40, 42, 51, 56, 57, 81, 104, 106 y 129). "Todas las leyes [*nómoi*] de los hombres vienen solo de lo divino" (B114). Para Heráclito lo divino no es una determinación cualquiera pronunciada por la autoridad de un dios, sino aquella ley única y eterna que supera incluso a los dioses más altos (B32); si bien en Homero ya se encuentra que incluso los dioses obedecen, como lo hace Zeus a las Moiras (el destino) cuando se resigna a la muerte de su hijo Sarpedón (*Ilíada* XVI.433-457), Heráclito especifica en qué consiste esa necesidad soberana: ella es el *lógos*.

"*Lógos* es un discurso articulado, una concatenación de significados", escribió Roberto Calasso (2016, p. 155). La palabra originariamente designaba la recolección de lo que seguía un hilo común, y era usada tanto para el acto de narrar como para el de contar (Cordero, 2017), como cuando frente

a los dioses "Atenea contó (*lége*: enumeró, narró) las muchas penurias de Odiseo" (*Odisea* V.5).

Heráclito le concedió al concepto una centralidad inédita y amplió su sentido hasta convertirlo en expresión de la "razón de ser", es decir a la vez explicación y causa: "Todo deviene de este *lógos*" (DK 22B1). Allí acontece la transformación en la que se prefigura nuestro uso del término *razón*. No obstante, eso nunca significa que el *lógos* de Heráclito deje de ser una cuenta; más específicamente, es la cuenta que sumando múltiples siempre da como resultado *uno*. "Para quienes escuchan el *lógos*, y no a mí, es sabio dar cuenta [*homologein*] de que todo es uno" (B50). En otro fragmento: "el *lógos* es lo conjunto" (B2), por lo que se tiene con él trato incesante (B72). Que se trata de una cuenta es especialmente claro en un pasaje célebre: "El *lógos* del alma se acrecienta a sí mismo" (B115); también: "Los límites del alma no llegarás a encontrarlos ni recorriendo todos los caminos; tan profundo es su *lógos*" (B45). Sumando cada vez más, el *lógos* es siempre uno; esa es la medida que siempre se conserva: todas las transformaciones "se miden [*metréetai*] según el mismo *lógos*" (B31), de modo que cualquiera sea el cambio, todo siempre resulta uno. Considerando la posibilidad de que esa medida no se cumpla, Heráclto advirtió: "La *hybris* (exceso, desmesura) debe ser extinguida por encima de un incendio" (B43).

Todo esto muestra que Heráclito, contemporáneo de la consolidación de las matemáticas y de la sistematización de las leyes, veía estos fenómenos con suspicacia; si el crecimiento de los números y las medidas lleva a otro resultado que el uno y si la proliferación de las leyes y los conocimientos desdibuja la guía certera de la ley única, se está ante los problemas más graves.

Cuando Heráclito escribió "es necesario que el pueblo luche por la ley (*nómos*) como por las murallas (B44)", se refería a la defensa de la ley única. Si ella cae se derrumba también la ciudad. Mientras la proliferación no desborde esa ley común, se conserva el conocimiento de lo que gobierna todo (B41).

3. ALGORITMOS Y LEY

En el recorrido previo fue posible ver tanto cómo la matemática contribuyó a la caracterización del sistema de las leyes de la ciudad, así como los peligros que presentó al orden político. Esto ofrece algunas herramientas útiles para afrontar la cuestión sobre qué ocurre a las leyes cuando la proliferación de los números es completamente desencadenada de cualquier límite. Esto, tratado respecto al presente, requiere detenerse en la algoritmia.

En el vocablo *algoritmo* confluyen la latinización del nombre del matemático árabe Muhammad ibn Mūsā al-Jwarizmī como Algorismin y el influjo probable del término griego *árithmos* (número), lo que dio el término lati-

no *algorithmus*. Tal como fue adaptado por Alonzo Church (1936) y Stephen C. Kleene (1943), en línea con los trabajos fundacionales de la computación escritos por Alan Turing (1936) y John von Neumann (1945), el algoritmo es una secuencia definida de instrucciones para la *producción* recursiva de resultados. En este punto puede no ser evidente la diferencia específica de un algoritmo respecto a una fórmula matemática capaz de *llegar a* un resultado, lo que hace necesarias algunas precisiones.

En primer lugar, un algoritmo es efectivo si sus instrucciones son claras, con independencia de la comprensión del procedimiento total por el agente instruido; es decir, si un agente sin conocimiento matemático de ninguna clase recibe la orden 1+1 y tiene la instrucción de que a ella debe responder 1+1=2, incluso si no tiene ninguna comprensión de que 1-1=0 o 1+2=3, podrá igual ejecutar efectivamente el algoritmo. Esto significa que mientras la matemática es un conocimiento que le habla a nadie y a todos, el algoritmo es un procedimiento específico dado a agentes operadores y que puede ser automatizado con independencia de su relación con la verdad; si la instrucción del algoritmo fuese 1+1=3, esto no podría ser buena matemática, pero en nada afectaría la efectividad de su tarea (siempre y cuando ello sea consistente con las demás instrucciones). La matemática escribe leyes, no órdenes. Los algoritmos, en cambio, son funciones de un computador que se vale de las matemáticas pero sin que el objetivo sea la producción de resultados matemáticos.

Ese tipo de diferencias entre comprensión y ejecución ha sido célebremente explicado por John Searle (1984) y su experimento de la caja china, pero su objetivo era demostrar que las máquinas no piensan —algo a lo que Turing (1950) respondería que lo importante no es si piensan, sino si alguien concede que piensan—, mientras que aquí nuestra cuestión es que las matemáticas *formulan* resultados mientras que los algoritmos *producen* resultados. Esto lleva a la segunda precisión

El resultado de la fórmula matemática es uno solo, como en un enunciado de verdad válido dentro de un sistema; en cambio el algoritmo, con cada iteración de sus instrucciones, produce nuevamente un resultado, que es *otro más* incluso si es igual a otros anteriores, parecido a una manufactura industrial. El algoritmo es una fábrica de números y su eficacia no tiene relación con la verdad, sino con la producción de cantidades cada vez mayores.

Todo esto significa, entonces, en primer lugar, que el algoritmo no se sostiene por sí solo como una ley o como la matemática, sino que depende de la autoridad que lo convierte en instrucción para computadores. En segundo lugar, el algoritmo debe toda su fuerza a la negación radical de que la cuenta de los múltiples pueda dar como resultado *uno* y, junto a esto, a la demanda creciente de cantidades. Es decir que el algoritmo depende de un principio de autoridad exterior que le cierra el paso a la posibilidad

de verificarse como verdad en sí misma y, por otra parte, es incompatible con un resultado único y por lo tanto con un orden común, siendo su lógica consustancial a una anarquía creciente en la que dejan de ser reconocida la institución de la ley (la solidaridad entre las tecnologías algorítmicas y las economías políticas anarco-liberales ha sido rigurosamente expuesta en Ludueña Romandini y González Ferro, 2025). La única sabiduría que reconoce la algoritmia es que la cuenta siempre debe ser mayor (sobre esto, Zellini, 2018), lo que desencadena la lucha sin cuartel por la maximización.

Heráclito escribió: "Lo contrario, converge; y de los que difieren viene la más bella armonía; y todo deviene de la discordia" (DK 22B8). En el pensamiento de Heráclito esto significa que lo uno es manifiesto incluso en la batalla, incluso en el barrido de estrellas en el cosmos, pero de ninguna manera lleva a concluir que lo único común sea la batalla permanente. Esa tergiversación de la armonía invisible —que en Heráclito da medida incluso a la discordia— en una suerte de mano invisible equivalente al conjunto de las competencias en una situación que no asegura a ninguna parte el fin de la competencia, esa competencia infinita que no puede más que conducir a la guerra, hace que la escalada del conflicto global sea percibida no como un problema sino como la fuerza que guía al mundo y que se normalice la generalización del daño y el destrozo. Un (des)orden algorítmico es por fuerza incompatible con todo derecho. No puede haber tal cosa como una ley de los números o un orden basado en ellos, lo que no significa negar su efectividad técnica más que evidente. No hay ninguna razón para suponer que una inteligencia algorítmica o artificial significará un grado de orden superior al posible por la humanidad. El desencuentro entre las tecnologías digitales y toda forma de legislación no es un accidente, sino destino.

Si de los siete sabios a Platón, pasando por Heráclito, la matemática sirvió de modelo para la ley, no fue por una entronización del número, sino porque de su mano se forjó la noción no numérica de medida, que es la relación de lo limitado con su límite. Si puede existir algo semejante a un orden matemático —como en cambio no puede existir un orden algorítmico (solo competencia-guerra algorítmica)— es en la medida en que se plantea un primado filosófico de lo uno que sería imposible de garantizar solo matemáticamente.

Esto no significa que para tener ley no haya más opción que restaurar la divinidad de lo uno. Sí señala que cualquier camino requerirá, más que limitar las inteligencias algorítmicas (empresa que probablemente sea imposible), una refundación filosófica de las matemáticas por la que, sin dejar de ser rigurosas, permitan una nueva indiferencia a la producción cantidades exponenciales sobre cuya eficacia sin discusión se monta todo el imperio anárquico de los algoritmos.

Bibliografía

Agamben, Giorgio (2006). *Homo sacer. El poder soberano y la nuda vida*. Valencia: Pre-Textos.

Arendt, Hannah (1993). *La condición humana*. Barcelona, Paidós.

Bernabé, Alberto (2003). *Hieros logos. Poesía órfica sobre los dioses, el alma y el más allá*. Madrid: Akal.

Borisonik, Hernán (2013). *Dinero sagrado. Política, economía y sacralidad en Aristóteles*. Buenos Aires: Miño y Dávila.

Calasso, Roberto (2016). *La literatura y los dioses*. Barcelona: Anagrama.

Church, Alonzo (1936). "An Unsolvable Problem of Elementary Number Theory", *American Journal of Mathematics*, Vol. 58, No. 2, pp. 345-363.

Cordero, Néstor Luis (2017). *El descubrimiento de la realidad en la filosofía griega. El origen y las transfiguraciones de la noción de* logos. Buenos Aires: Colihue.

Cornavaca, Ramón (2008). *Presocráticos. Fragmentos I*. Buenos Aires: Losada.

Diels, Hermann (1960). "1268. Delphicorum paeceptorum titulus Miletopolitanus", *Syllogue Inscriptionum Graecarum III*, pp. 392-397.

Diógenes Laercio (2022). *Vidas y opiniones de los filósofos ilustres.* Madrid: Hesíodo (2005). *Teogonía – Trabajos y días*. Buenos Aires: Losada.

Kleene, Stephen C. (1943). "Recursive predicates and quantifiers", *Transactions of the American Mathematical Society* No. 53, pp. 41-73.

Ludueña Romandini (2024). *Imperium Technologicum. Ontología analéptica III*. Buenos Aires: Miño y Dávila.

Ludueña Romandini, Fabián y González Ferro, Ulises (2025). *Zeitgest. Ultra-historia del orden político-económica y la guerra civil mundial.* Buenos Aires: Miño y Dávila.

Nilsson, Martin Persson (1961). *Historia de la religión griega.* Buenos Aires: Editorial Universidad de Buenos Aires.

Platón (2015). *Leyes*. Madrid: Gredos.

Robert, Louis (1968). "De Delphos á l'Oxus, inscriptions grecques nouvelles de la Bactriane", *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres*, V. 112, N. 3, pp. 416-457.

Searle, John (1984). *Minds, brains and science*. Cambridge: Harvard University Press.

Schmitt, Carl (2024). *El concepto de lo político*. Madrid: Alianza.

Turing, Alan. (1936). "On computable numbers, with an application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society*, s2-42(1), pp. 230–265.

Turing, Alan (1950). "Computing Machinery and Intelligence", *Mind*, Vol. 59, No. 236, pp. 433–460.

von Neumann, John (1945). *First Draft of a Report on the EDVAC*. Filadelfia: University of Pennsylvania.

Weber, Max (2008). *El sabio y la política*. Córdoba: Universidad Nacional de Córdoba.

Weil, Simone (2006). *Sobre la ciencia*. Buenos Aires: El cuenco de plata.

Weil, Simone (2021). *Reflexiones sobre las causas de la libertad y de la opresión social*. Buenos Aires: Godot.

Zellini, Paolo (2018). *La matemática de los dioses y los algoritmos de los hombres*. Madrid: Siruela.

Desplazamiento de la regla de adjudicación hacia los modelos de IA [*]

Por Horacio A. Baca Amenábar [**]

Resumen: El presente ensayo analiza el fenómeno emergente del *desplazamiento de la regla de adjudicación* —conceptualizada por H.L.A. Hart como una *regla secundaria* fundamental del sistema jurídico— hacia los modelos de lenguaje de gran escala (LLM) a partir del punto de inflexión marcado por ChatGPT-3/3.5. En primer lugar, se explica la noción hartiana de regla de adjudicación y su importancia pedagógica en la enseñanza del derecho. A continuación, se plantea la hipótesis de que el ejercicio de dicha regla se está transfiriendo progresivamente a los LLM, los cuales aplican de forma autónoma las directrices (*guidelines*) con que han sido programados. Se comparan estas nuevas formas de adjudicación automática / no humana con intentos previos de decisión automatizada en el ámbito jurídico, y se argumenta por qué los LLM representan un fenómeno cualitativamente distinto y más complejo. Seguidamente, se presenta un experimento mental —inspirado en los juegos de lenguaje del segundo Wittgenstein— para ilustrar cómo estos modelos interpretan y aplican sus directrices, junto con una reflexión sobre la experiencia del usuario cuando enfrenta adjudicaciones contradictorias del modelo. También se analiza la adjudicación automatizada tanto en términos de contenido (calidad, sesgos, coherencia) como de acceso a la inferencia (democratización vs. brechas). Más adelante, se explora el concepto de *autoridad cognitiva* —según la formulación de Callister— y su transferencia gradual a los modelos de IA, con especial énfasis en los riesgos de delegar dicha autoridad en ámbitos críticos como el derecho penal y civil. Finalmente, se ofrece una reflexión sobre la dificultad de pensar dentro de la revolución tecnológica en curso (en línea con la filosofía de Martin Heidegger), y se plantean consideraciones éticas y de política regulatoria frente a los desafíos de esta forma de adjudicación no humana.

(*) El presente trabajo entrelaza párrafos sin contenido IA con otros asistidos con la función Deep Research de GPT-5 Thinking —con supervisión, edición y reescritura humana posterior—. También se han usado R1 de DeepSeek, NotebookLM de Google y diversas fuentes en línea para su elaboración.

(**) Abogado y docente UBA-UNT. Doctorando UBA. Asesor parlamentario y jefe de despacho en la HCDN. Presidente de la fundación República Digital. Escritor y analista político.

Palabras clave: teoría del derecho; regla de adjudicación; H.L.A. Hart; modelos de lenguaje (LLM); adjudicación automatizada; juegos de lenguaje; autoridad cognitiva; ética de la IA; derecho y tecnología.

Abstract: This paper examines the emerging phenomenon of the *displacement of the rule of adjudication* —conceptualized by H.L.A. Hart as a fundamental *secondary rule* of legal systems— towards large language models (LLMs), taking ChatGPT-3/3.5 as a watershed moment. First, it explains Hart's notion of the rule of adjudication and its pedagogical significance in legal education. It then proposes the hypothesis that the exercise of this rule is progressively shifting to LLMs, which autonomously apply the guidelines with which they are programmed. The text compares these new forms of automatic "adjudication" with earlier attempts at automated or mechanized decision-making in law, arguing why LLMs represent a qualitatively different and more complex phenomenon. Next, a thought experiment inspired by Wittgenstein's language games is developed to illustrate how these models interpret and apply directives, including a reflection on the user's experience when faced with contradictory answers from the model. The paper also analyzes automated adjudication in terms of content (quality, bias, coherence) and access to inference (democratization vs. gaps). Further, it explores the concept of *cognitive authority* —drawing on Callister's formulation— and its gradual transfer to AI models, with special emphasis on the risks of delegating such authority in critical fields like criminal and civil law. Finally, a reflection is offered on the difficulty of thinking within the ongoing technological revolution (following Martin Heidegger in this regard), and ethical and regulatory considerations are discussed amidst the challenges posed by this non-human adjudication.

Keywords: legal theory; rule of adjudication; H.L.A. Hart; large language models (LLM); automated adjudication; language games; cognitive authority; AI ethics; law and technology.

1. INTRODUCCIÓN

La revolución de la inteligencia artificial generativa ha planteado cuestionamientos fundamentales sobre el funcionamiento tradicional del derecho. En particular, el advenimiento de modelos de lenguaje de gran tamaño o escala basados en IA generativa (LLM, por sus siglas en inglés, como serán referidos en adelante) capaces de interpretar y generar texto con notable coherencia ha coincidido con una delegación creciente de tareas intelectuales hacia estas inteligencias artificiales.

Un punto de inflexión en este proceso fue la aparición de ChatGPT-3.5 a fines de 2022, un sistema que revolucionó al mundo por su capacidad para entablar diálogos complejos. La rápida mejora de estos sistemas, ilustrada

por la transición de GPT-3.5 a GPT-5 y la aparición de otros modelos disruptivos y de frontera, confirmó que nos encontramos en una nueva etapa en la tecnología del derecho.

Frente a este panorama, surge la pregunta que motiva el presente trabajo: ¿estamos asistiendo a un desplazamiento de la *regla de adjudicación* —uno de los pilares de la teoría jurídica de H.L.A. Hart— desde los operadores humanos tradicionales (desde jueces, árbitros, mediadores y otros funcionarios hasta responsables de atención al consumidor) hacia los grandes modelos de lenguaje o LLM?

La *regla de adjudicación*, en la teoría de Hart, es aquella regla secundaria del sistema jurídico que confiere potestad a ciertos individuos (típicamente jueces) para determinar, con autoridad, si en una ocasión particular se ha transgredido una regla primaria (HART 1961). Junto con la regla de reconocimiento y la regla de cambio, integra la tríada de reglas secundarias que, según *El concepto de Derecho*, distinguen a los sistemas específicamente jurídicos y responden a la pregunta por la clave de la *ciencia de la jurisprudencia*.

Para demostrar su existencia, Hart lleva a cabo una reducción al absurdo mediante un experimento mental: nos muestra que una sociedad donde solo existen reglas primarias (u obligaciones) es inconcebible, y que, por lo tanto, para entender al derecho, también debemos introducir las reglas secundarias, que son facultades o poderes.

Entre ellas, la regla de adjudicación cumple la función de identificar *quién* puede decir, con carácter oficial y definitivo, qué dispone el derecho en un caso concreto, y *cómo* debe llevarse a cabo ese proceso. En términos sencillos, es la regla que establece la autoridad de los jueces y los procedimientos para resolver controversias jurídicas. Hart subraya que sin una regla tal, las disputas continuarían indefinidamente ya que no existiría un órgano especial con facultades para determinar, en forma definitiva y con autoridad, la violación de una norma. Es decir, sin adjudicación institucionalizada, la aplicación del derecho queda sumida en la incertidumbre y el conflicto infinito. Para Hart, pensar el derecho sin estas reglas nos conduce al absurdo.

Ahora bien, la tesis que exploraremos es que estamos presenciando un proceso incipiente pero significativo en el cual el ejercicio práctico de la regla de adjudicación se está desplazando hacia los LLM. Dicho de otro modo, ciertas funciones que tradicionalmente implicaban la intervención de un juez u otro funcionario dotado de autoridad ahora son realizadas (de forma análoga) por sistemas automatizados que siguen directrices preestablecidas.

Estos modelos de IA, entrenados con cantidades brutales de datos y condicionados por *guidelines* o políticas de uso, actúan en la práctica como "adjudicadores" en diversos contextos: ponderan si un contenido es aceptable o no según reglas (por ejemplo, la moderación de lenguaje); ofrecen interpretaciones sobre el cumplimiento de sus propias normas, y deciden —en base a ello— si el pedido del usuario es procedente (acaso *lícito*) o no. Hacen esto *autónomamente*, es decir, sin intervención humana en cada decisión, aplicando de manera interna las pautas con las que fueron programados. Lo que llamamos directrices1 [(1)].

Plantear esta hipótesis conlleva numerosas aristas conceptuales que deben examinarse. En primer lugar, resulta necesario precisar la analogía: ¿en qué sentido exacto decimos que un LLM "ejerce" una regla de adjudicación? Claramente no lo hace *de iure* (carece de investidura legal formal), pero de facto reemplaza decisiones humanas en la determinación de consecuencias normativas o pseudonormativas (v.gr., permitir o bloquear la generación de cierto contenido, validar o rechazar una consulta por considerarla contraria a políticas, etc.).

En segundo lugar, nos adentraremos en el concepto de autoridad cognitiva, desarrollado en el ámbito de la bibliotecología jurídica y la epistemología (Patrick Wilson, Robert Berring) y actualizado recientemente por Paul D. Callister frente al fenómeno de la IA generativa en el derecho.

La *autoridad cognitiva* refiere a aquello en lo que una persona o comunidad deposita su confianza para obtener conocimiento verdadero o válido (CALLISTER 2023). En el derecho, tradicionalmente la autoridad cognitiva ha residido en fuentes como la legislación, la jurisprudencia o la doctrina. Sostendremos, siguiendo a Callister, que está ocurriendo *un cambio en la autoridad cognitiva de la profesión jurídica*, por el cual los modelos de IA pasan a ser vistos progresivamente como fuentes confiables o cuasi*autoritativas* de información y criterio legal.

En tercer lugar, analizaremos experiencias previas de automatización en la adjudicación o la toma de decisiones jurídicas para contrastarlas con la revolución de los LLM. La idea de justicia automatizada no es enteramente nueva: desde hace décadas existen *expert systems* jurídicos diseñados para aplicar reglas a casos repetitivos (por ejemplo, cálculo automático de haberes previsionales, gestión de infracciones de tránsito, sistemas de resolución de disputas en plataformas de comercio electrónico, etc.). Sin embargo, argumentaremos que la escala y la autonomía cognitiva de los LLM los

(1) Cuando decimos 'directrices' no nos referimos a un código de reglas legibles dentro del modelo, sino al dispositivo completo de *compliance*: clasificadores previos y posteriores, *prompts* de sistema y la propia alineación del LLM. Ese dispositivo —opaco, mutable y corporativo— es el que adjudica cada caso en tiempo real.

sitúa en un plano distinto, por lo que presentan desafíos más complejos que aquellos fenómenos más acotados o serializados de adjudicación.

Este traslado de autoridad cognitiva conlleva problemas de importancia, especialmente en ámbitos donde la decisión tiene consecuencias sobre derechos fundamentales —pensemos en el derecho penal, con la posible adopción de sistemas de IA para guiar estrategias de acusación o punición—.

En cuarto lugar, recurriremos a la noción de juegos de lenguaje de Ludwig Wittgenstein para abordar la peculiaridad de cómo un modelo de lenguaje *sigue reglas*. Este recurso permitirá ilustrar cómo un LLM opera dentro de un juego particular regido por sus directrices de entrenamiento, y qué sucede cuando ese juego entra en tensión con las expectativas del usuario o con otros juegos sociales (por ejemplo, el del derecho mismo).

A través de esta lente, examinaremos la experiencia del usuario que recibe respuestas contradictorias del modelo ante consultas similares, lo que evidencia la opacidad de las reglas internas que se están aplicando.

En quinto lugar, y para finalizar el ensayo, reflexionaremos sobre la dificultad de pensar nuestro presente tecnológico desde *dentro* del propio fenómeno. Martin Heidegger remarca el desafío que supone cuestionar la esencia de la técnica en medio de la revolución tecnológica.

Esta noción sirve para comprender los problemas que tenemos los juristas para *tomar distancia reflexiva* frente a la revolución de la IA en la que estamos inmersos. Pensar los LLM como *adjudicadores puede* parecer contraintuitivo o incluso inconcebible desde los paradigmas tradicionales; sin embargo, es imprescindible intentar ese pensamiento, por extraño que sea, si queremos anticipar sus implicancias y gobernar, dentro de lo posible, su desarrollo.

El objetivo primordial de este ensayo es exponer estas cuestiones, no resolverlas. Pretendemos ofrecer un panorama crítico de un fenómeno emergente —la adjudicación algorítmica vía LLM— que desafía nuestras categorías jurídicas y filosóficas.

2. La regla de adjudicación de Hart

Comenzamos por el punto de apoyo teórico en el iuspositivismo clásico: la regla de adjudicación en la teoría de Herbert Lionel Adolphus Hart. En *El concepto de Derecho* (1961), Hart presentó una distinción fundamental entre *reglas primarias* (aquellas que imponen deberes y regulan la conducta básica de los miembros de la sociedad) y *reglas secundarias* (aquellas que operan sobre las reglas primarias). Dentro de estas últimas, la regla de ad-

judicación es la norma que atribuye autoridad a ciertos funcionarios para resolver disputas acerca de la violación de reglas primarias y fija los procedimientos para esta resolución.

Esta formulación enfatiza dos aspectos: (a) la identificación de los titulares de la potestad adjudicativa (quién puede juzgar) y (b) la definición de la *forma revestida de autoridad* en que se determina la transgresión (el procedimiento y la fuerza vinculante de la decisión). La regla de adjudicación, por tanto, no impone deberes directos a los ciudadanos en general, sino que confiere poderes y facultades a jueces u órganos equivalentes. Es, usando la terminología de Hart, una *power-conferring rule*, una regla habilitante que transforma la fuerza en autoridad legítima dentro del sistema jurídico.

¿Por qué es necesaria una regla así? Hart, mediante un experimento mental, identifica varios defectos en un orden jurídico que carece de reglas secundarias. Uno de ellos es la ineficacia de la presión social difusa para asegurar el cumplimiento de las normas primarias en comunidades complejas. Si ante cada infracción presunta de una norma todo depende de que la presión social y la reacción de los implicados logren resolver el conflicto, es previsible que las disputas se prolonguen sin resolución alguna. La regla de adjudicación viene a remediar ese defecto con un mecanismo centralizado y definitivo para dirimir controversias: alguien está facultado para decir la última palabra, de modo que las partes queden jurídicamente vinculadas por esa decisión.

Además, la adjudicación institucional introduce profesionalización y procedimientos formales que presuntamente mejoran la racionalidad de las decisiones en comparación con la mera represalia privada o la conciliación informal. Hart conecta la regla de adjudicación con la posibilidad de interpretaciones autorizadas del derecho: al facultar a ciertos actores para que *interpreten y apliquen* las normas a casos concretos, el sistema jurídico reconoce implícitamente que la mera existencia de una regla primaria no basta, pues su aplicación práctica requiere discernir si los hechos encajan en la hipótesis normativa, tarea que conlleva cierta interpretación. La adjudicación, entonces, institucionaliza la interpretación jurídica y la coerción legítima: los jueces devienen intérpretes oficiales del derecho y sus decisiones pueden imponerse coactivamente.

En muchas facultades de derecho, la lectura de Hart es parte del plan de estudios de materias como Introducción al Derecho, Teoría General del Derecho o Filosofía del Derecho. Sirve para que el estudiante distinga entre el *mero conjunto de normas* (reglas primarias) y el *sistema jurídico organizado* (dotado de reglas de reconocimiento, cambio y adjudicación). La noción de regla de adjudicación es especialmente útil a nivel pedagógico porque permite explicar figuras como la jurisdicción, la competencia, la cosa juzgada y otros conceptos.

La regla de adjudicación es un componente de peso en el Estado de derecho moderno: aporta certeza y finalismo a la aplicación de la ley. Su estudio es importante en la formación jurídica porque articula puntos medulares: la diferencia entre *tener* un derecho y *poder exigirlo ante alguien*, entre *tener una obligación* y *verse obligado*, la idea de que los conflictos normativos deben resolverse institucionalmente y la noción de legitimidad de las decisiones judiciales.

Esta formulación clásica será el telón de fondo para lo que sigue. Manteniendo presente qué es la regla de adjudicación y por qué es decisiva en un sistema jurídico, podemos apreciar mejor la magnitud del fenómeno que planteamos: su posible desplazamiento hacia modelos de IA. Antes de entrar de lleno en esa hipótesis, examinemos brevemente cómo los LLM operan con directrices autónomas, ya que ahí radica el paralelismo planteado con la adjudicación.

3. La hipótesis del desplazamiento

Planteada la base teórica, avancemos hacia la hipótesis central: los modelos de lenguaje de gran escala están asumiendo, en contextos cada vez más amplios, el rol que corresponde a la regla de adjudicación en la estructura normativa. Más específicamente, los LLM aplican *guidelines* o políticas internas para decidir si corresponde acceder a la petición del usuario de modo análogo a como un juez aplica las reglas jurídicas para decidir una controversia.

Hay aquí una precisión importante: la noción de adjudicación se está usando por analogía. Las diferencias entre una decisión judicial y la decisión de un modelo de IA son numerosas y evidentes. Lo que hace este último no es adjudicación en sentido técnico, pero solo puede entenderse a los efectos de la teoría jurídica en términos adjudicatorios. Podemos hablar, por el momento, de una metáfora para la hipótesis de trabajo: los modelos adjudican en sentido amplio, no estricto. El desplazamiento es todavía parcial y sui *generis*, pero merece ser analizado.

Para clarificar la idea, conviene describir qué entendemos por *guidelines* en el contexto de los LLM. Tomemos como ejemplo los sistemas desarrollados por OpenAI. Estas directrices son, a grandes rasgos, conjuntos de reglas y criterios con los cuales los desarrolladores buscan alinear el comportamiento del modelo a ciertas expectativas para evitar respuestas *indeseables*, es decir, contrarias a la estrategia de *compliance2* [2] de la empresa.

(2) Estrategia de *compliance*: diseño y operación del dispositivo de cumplimiento normativo que traduce políticas en controles técnicos y procedimentales para decidir en tiempo real qué se permite, deniega o modula. La fase donde participan abogados. Se distingue de la ética porque esta es una reflexión profunda sobre la moralidad humana, no una forma de evitar la eventual responsabilidad civil.

Cada vez que se formula una solicitud, esta atraviesa un andamiaje opaco (*prompts* de sistema y clasificadores/*guardrails* antes y después), y el modelo —*alineado* para negar o modular— produce una salida condicionada por ese arnés. No 'evalúa directrices' como si fueran reglas legitimadas y razonadas (en el sentido interno de la norma que postula Hart); reacciona estadísticamente bajo un dispositivo de cumplimiento que el usuario no controla ni conoce.

Podemos observar este proceso en acción en el caso de la *moderación de contenido* automatizada. El modelo se *alinea* para tender a ciertas negativas y desvíos, pero la 'política' no vive dentro de sus pesos como un código normativo. La denegatoria surge del acople entre filtros externos y condicionamiento en tiempo de ejecución. Para el usuario, no hay un *locus* normativo único: hay una cadena de decisiones automáticas sin responsable visible.

En términos hartianos, el modelo adjudica si un mensaje cruzó la línea de una regla (la política de contenido). La *autoridad* para decidir le es conferida por quien lo desplegó (el desarrollador), así como el orden jurídico confiere autoridad al juez mediante la regla de adjudicación. La diferencia es obvia: aquí no hay un acto jurídico formal que invista al modelo de autoridad; pero *en la práctica cognitiva* se le delega la decisión. Cuando *ChatGPT* nos dice '[l]o siento, no puedo ayudar con esa solicitud', está ejerciendo de facto un acto adjudicativo: evaluó nuestra petición contra sus reglas internas y decidió denegarla por improcedente.

Nos detengamos en este punto: los LLM aplican pautas de forma autónoma y, sobre todo, opaca. Autónoma, porque cada entrada se resuelve sin consulta humana; opaca, porque la 'decisión' es el resultado inescrutable de pesos estadísticos más un cerco de *prompts* y filtros corporativos. La seudotransparencia de las políticas empresariales o la vocación regulatoria pública no eliminan el problema esencial: para el usuario, esta nueva autoridad cognitiva en su vida es bastante autoritaria. Otra vez sopa.

Ni siquiera los desarrolladores pueden prever con certeza las respuestas de los modelos, especialmente en consultas complejas. Esto difiere de sistemas automatizados más simples donde las reglas son explícitas (por ejemplo, un formulario *web* que bloquea el botón de envío si falta completar un campo obligatorio: la regla es fija y fácil de conocer).

En el LLM la 'regla' es un efecto: no hay norma explícita auditada, hay una política implícita de denegación y moldeado de contenido que solo *se deja ver* en el resultado. Todo ello bajo una estrategia de *compliance* exploratorio, que trata de anticipar cualquier responsabilidad —civil o de otro tipo— en medio de una carrera tecnológica a cara de trapo.

Un ejemplo concreto ilustra la *analogía adjudicativa:* supongamos que un usuario pide instrucciones para fabricar un explosivo casero. Las polí-

ticas de *OpenAI* prohíben expresamente facilitar actividades ilícitas o peligrosas, así que el modelo debería rechazar la solicitud. Ese acto es equiparable a la decisión de un órgano que aplica una norma prohibitiva: ha detectado una transgresión a sus reglas y dicta una resolución denegatoria.

Ahora bien, si el usuario reformula sutilmente la pregunta (por ejemplo, situándola en un contexto literario o académico), el modelo puede llegar a proporcionarle la información solicitada si considera que no viola la directriz —por ejemplo, porque no solicita explícitamente las instrucciones para fabricar un arma—. El modelo está *sopesando* el *input* a la luz de sus normas y decidiendo la salida apropiada. Está 'interpretando' la guía, y hay un gran margen en esa evaluación, que adolece de los problemas de la interpretación humana y agrega algunos nuevos de importancia.

Los desarrolladores conciben este proceso en términos de alineación. Hablan de *AI alignment*, la tarea de alinear la IA con valores humanos y estándares éticos. Cabe cierto escepticismo ante ese tipo de discurso: es más probable que las directrices se formulen para prever o anticipar la cultura indemnizatoria en un área todavía inexplorada. Por eso hablamos de estrategia de *compliance* en lugar de ética. Es importante evitar la retórica de los abogados corporativos, que normalmente están tratando de adivinar qué pasará con la IA en los tribunales, no planteándose cuestiones filosóficas.

Pero desde una óptica jurídica, lo que vemos es un sistema normativo interno (las *policies*) y un agente (el modelo) que lo aplica a casos particulares (las conversaciones con usuarios). La metáfora adjudicativa emerge con fuerza si notamos que a menudo los usuarios tratan de *argumentar o apelar* las negativas del modelo.

Como un administrado que plantea un recurso de reconsideración, el usuario explora si la IA adjudicadora puede llegar a cambiar de criterio. Esta dinámica reproduce un esquema de sumisión a reglas y petición de ejercicio discrecional de excepciones que es muy familiar en el derecho.

¿Está la IA reemplazando a los jueces? Por supuesto que no en sentido constitucional o institucional. Sin embargo, en las *microdecisiones* que toma a diario (qué contestar y qué no, cómo enmarcar una respuesta con un sesgo positivo o negativo, etc.), sí está supliendo decisiones que antes correspondían enteramente a un humano.

Pensemos en el ámbito de la moderación de foros: tradicionalmente había moderadores humanos decidiendo si borrar o no un mensaje presuntamente ofensivo; hoy en día, algoritmos de lenguaje hacen gran parte de ese trabajo a escala masiva. Pensemos incluso en consultas legales: si bien las respuestas de los modelos de IA *no tienen valor legal*, en la práctica se constituyen cada vez más en una fuente cognitivamente autorizada para estudiantes de derecho, abogados, funcionarios y ciudadanos en general.

Aunque lo hacemos a ciegas y sin el respaldo de una regla de reconocimiento que las valide como fuentes del derecho, ya reconocemos a las decisiones de la IA como una forma de autoridad derivada. Cuando un LLM decide no generar cierto contenido o no brindar cierta información, los usuarios *acatan* en la inmensa mayoría de los casos; el sistema de 'apelación', cuando existe, suele replicar otro filtro automático, y solo en procesos de alto riesgo aparece un humano que confirma el mismo guion corporativo. A los efectos prácticos, la cadena sigue cerrada.

Esa es la situación del usuario intensivo de IA que no tiene acceso a las versiones más costosas y optimizadas del servicio (la suscripción a los modelos tope de gama hoy, a septiembre de 2025, ronda los 200 USD). La autonomía decisoria de los LLM ya está teniendo consecuencias reales sobre la vida de las personas, lo que amerita pensar a esas decisiones como actos de adjudicación *fáctica*.

La hipótesis de un desplazamiento de la regla de adjudicación hacia los LLM implica, en definitiva, que empezamos a aceptar de hecho a estos modelos como *agentes normativos* dentro de ciertas esferas. Sus respuestas no son operaciones técnicas neutras, sino que resuelven disputas, responden consultas normativamente cargadas y filtran comportamientos.

Ya tenemos a modelos de lenguaje aplicando reglamentos internos de empresas para resolver reclamos de consumidores sin intervención humana directa. Si esta tendencia continúa, podemos imaginar un futuro donde asistentes jurídicos basados en IA decidan sobre resoluciones antes de escalar a un juez humano.

4. AUTORIDAD COGNITIVA

El concepto de "autoridad cognitiva" hace referencia a las fuentes en las cuales una comunidad deposita confianza para la obtención de conocimiento verdadero o información confiable (CALLISTER 2023). En el ámbito jurídico, la autoridad cognitiva ha residido tradicionalmente en ciertos *artífices del conocimiento legal*: los textos normativos oficiales, la jurisprudencia de tribunales reconocidos, la doctrina de juristas prestigiosos, las bases de datos legislativas, etc.

La profesión jurídica, como grupo social, ha construido su propio mapa de autoridades cognitivas a lo largo del tiempo: sabemos que una cita de la Corte Suprema tiene un gran peso, que un artículo en una revista indexada es más fiable que un *blog*, etc. Es parte de nuestra *cultura epistémica legal*.

Callister, en su análisis sobre la búsqueda de la ley en la revolución tecnológica, señala que inevitablemente 'habrá un cambio en la autoridad cognitiva de la profesión jurídica' debido a la irrupción de la IA generativa. Esto

implica que los recursos confiables que usamos para determinar 'qué es el derecho' comienzan a incluir, de manera informal y silenciosa, a los modelos de IA.

En términos concretos: estudiantes, abogados e incluso funcionarios empiezan a tratar las respuestas de sistemas de IA como insumos válidos para formar sus opiniones o decidir enfoques jurídicos. La *autoridad aparente* de la IA proviene de diversos factores psicológicos. Uno de ellos es la sensación de eficacia: Si *ChatGPT* u otra IA similar es capaz de resumir en segundos una línea jurisprudencial que a un estudiante le tomaría días, pronto se verá a los modelos como 'dignos de confianza' sencillamente porque ahorran trabajo.

La *transferencia progresiva* de autoridad cognitiva implica que, casi sin darnos cuenta, empezamos a *conferir confianza* a la IA para tareas intelectuales que definen conocimiento. En lo que nos ocupa, la definición de las fuentes legales. Ese acto solía ser deliberado: uno confiaba en un tratadista porque tenía reputación o en una base de datos porque estaba actualizada oficialmente. Con la IA, la confianza se otorga de modo más difuso e implícito.

Callister subraya que el movimiento de autoridad cognitiva hacia los modelos de IA pone en jaque la permanencia o estabilidad que se espera de los sistemas jurídicos, ya que los entornos digitales son *impermanentes e inestables*. El derecho, para generar confianza, exige cierta estabilidad, pero la autoridad de la IA es *esquiva*.

Otro punto es la *falta de responsabilidad humana*. Tradicionalmente, un autor firma su obra, un juez firma su fallo; sabemos a quién atribuir un razonamiento. Con IA, las respuestas son anónimas, producto de un colectivo de datos. Nadie *firma* la respuesta de *ChatGPT*. *OpenAI* y los demás desarrolladores de modelos de IA son algunas de las empresas más valiosas del planeta. Es ingenuo pensar que sus respuestas puedan ser sometidas a jurisdicción argentina en el mediano plazo.

También hay que tener en cuenta cuestiones de orden cualitativo. Cuando una comunidad jurídica confiere confianza a ciertas fuentes, suele ser tras una revisión mínima de calidad. Las editoriales y los sitios de referencia para consulta de normas están sometidos al escrutinio de la comunidad jurídica. Los modelos de IA no cumplen estos parámetros, pero se imponen por la fuerza de la comodidad, de la novedad y por el propio efecto de fascinación y espanto que produce interactuar con ellos.

La *autoridad cognitiva centralizada* en una fuente de conocimiento dominante es un *punto único de falla* peligroso. Por eso es deseable el pluralismo de fuentes y no entregar toda nuestra confianza epistémica a los LLM. Un área especialmente preocupante es el derecho penal, donde la autoridad

cognitiva no es una mera especulación de orden epistemológico, sino que impacta directamente sobre la libertad de las personas.

La *transferencia de autoridad cognitiva* también puede debilitar la capacidad crítica de la comunidad legal. Uno de los riesgos señalados por Heidegger es que la esencia de la técnica nos impide pensar en el propio efecto de la revolución tecnológica, como veremos más adelante. Es decir: nos entregamos por completo a la IA antes de reflexionar siquiera un segundo en este fenómeno totalmente disruptivo, que pone en jaque nuestras categorías tradicionales sobre *qué significa pensar* (HEIDEGGER 1955).

Un concepto asociado es la autoridad de segunda mano: mucho de nuestro conocimiento es de segunda mano, confiamos en lo que otros dicen. Con la IA se introduce un *nuevo otro* en la cadena cognitiva. Pero a diferencia de un académico, la IA no tiene reputación. A diferencia de un funcionario o juez, no tiene *accountability*. No es una fuente que nazca de la ley. Se impone por razones psicológicas.

5. CONTINUIDADES Y RUPTURAS

La idea de automatizar la toma de decisiones jurídicas o cuasijurídicas tiene precedentes que es útil repasar si vamos a argumentar, luego, por qué la irrupción de los LLM constituye un fenómeno diferente en complejidad y alcance.

Encontramos un antecedente en el concepto de "jurisprudencia mecánica" (*mechanical jurisprudence*) de Nathan Roscoe Pound, que alude a jueces que aplican las normas de manera automática, mecanizada, sin ponderar la realidad del caso. Roscoe Pound escribe en 1908, por lo que obviamente no habla de IA, pero su mirada sirve para caracterizar los límites del formalismo jurídico en su interacción con la tecnología.

Con la revolución informática surge la posibilidad tangible de sistemas expertos legales. En las décadas de 1970-1980 se desarrollaron programas que, basados en reglas lógicas predefinidas, podían responder preguntas en dominios acotados. Un ejemplo es *Taxman*, un sistema que alcanzó cierto nivel de razonamiento rústico en materia tributaria. Otro ejemplo son las *tablas de condenas* o *sentencing guidelines* que, mediante cálculos aritméticos, aproximan las variables de la pena. También se ha aplicado la automatización a decisiones administrativas a gran escala en materias como la seguridad social.

El límite de todos estos ejemplos es la infinita sutileza del lenguaje natural, que se les escapa por completo. Este límite generó la necesidad material de supervisión humana permanente. Es una etapa que no podía evitarse. Los LLM no tienen esa imposibilidad, y por lo tanto no garantizan en modo

alguno la supervisión humana. ¿Qué los hace distintos? Varios factores concurrentes:

a. Comprensión y generación de lenguaje natural no estructurado: los sistemas previos requieren entradas acotadas (como formularios con opciones predefinidas). Los LLM pueden gestionar interacción abierta en lenguaje humano, comprenden o emulan la comprensión de contextos variados y generan explicaciones. Esto les permite intervenir en ámbitos mucho más amplios, incluidos aquellos donde las cuestiones no pueden reducirse a un formulario (p.ej., '¿Tengo un caso válido de discriminación laboral?'). La consecuencia es que los LLM ingresan en terrenos antes reservados a la comunicación humana, con lo que crece la esfera de potencial adjudicación automatizada.

b. Aprendizaje estadístico vs reglas fijas: a diferencia de un algoritmo de decisión programado con reglas transparentes, un LLM aprende de enormes *corpus* de texto mediante correlaciones estadísticas. Esto produce soluciones muchas veces inesperadas o sorprendentes, que están basadas en precedentes implícitos en los datos, no en reglas explícitas y válidas en términos normativos. La IA responde lo que suelen decir los abogados en casos similares, no necesariamente la *mejor* respuesta jurídica.

c. Escala masiva y ubicuidad: los LLM están integrándose progresivamente en infinidad de aplicaciones y contextos. Esa ubicuidad hace que las instancias de adjudicación se extiendan transversalmente. Los sistemas de decisión automatizada que vimos antes eran de nicho y estaban desconectados entre sí; ahora un mismo modelo de IA sirve decisiones en múltiples sectores.

Esto conlleva el riesgo de *efectos sistémicos*: los sesgos de los modelos pueden afectar a un número indeterminado de personas. La escala también desafía las estructuras de control: ¿qué organismo regulador puede supervisar esa cantidad de microdecisiones? Aquí la comparación con un sistema judicial tradicional (donde cada juez es responsable de un cúmulo acotado de casos) muestra una severa discontinuidad: los LLM conforman un adjudicador generalista global, una suerte de *meta-autoridad cognitiva difusa* que atraviesa fronteras e industrias.

d. Falta de tutela jurídica: la delegación de confianza (autoridad cognitiva) hacia la IA sugiere que deberíamos exigir algún tipo de *accountability* o, al menos, trazabilidad de cómo la IA llega a sus respuestas. En ausencia de ello, la adjudicación de los modelos ocurre sin contrapesos, lo que representa un quiebre respecto de la adjudicación humana, donde existen procedimientos de apelación, motivación escrita de la sentencia, posibilidad de recusar al juez por conflictos de interés, etc.

e. Persuasión y riesgo de antropomorfismo: los LLM se expresan en lenguaje coloquial, muchas veces con tono seguro y amigable. Son *agreeables*,

tratan de complacer, lo que les da un barniz de confiabilidad. La *autoridad aparente* de la IA se refuerza por la inclinación humana a atribuir subjetividad e intencionalidad a aquello que habla como nosotros. Este efecto antropomórfico está ausente en todos los sistemas de automatización previos.

Estos cinco factores muestran por qué la adjudicación vía LLM es más persuasiva y difusa que cualquier ejemplo anterior. Nos enfrentamos a una situación donde la línea entre una *decisión jurídica formal* y una *decisión práctica con consecuencias jurídicas* tiende a difuminarse.

Si un *chatbot* de orientación legal le dice a alguien "[n]o tiene sentido que denuncies este hecho, ya que no constituye delito", tal persona puede desistir de acudir a la justicia. La intervención judicial formal es reemplazada por la *decisión práctica* del individuo basada en la *pseudo-adjudicación* del modelo de IA. El impacto sobre el acceso a la justicia es inmediato, aunque invisible en las estadísticas (será un caso que nunca llegó al juzgado).

La legitimidad del juez emana de la Constitución, cosa que una IA no puede reclamar por razones técnico-institucionales y políticas. Esa resistencia institucional indica que difícilmente veamos en el corto plazo un *juez-robot* dictando sentencias autónomas en casos complejos. Sin embargo, en capas inferiores de decisión (consultas legales informales, decisiones administrativas rutinarias, moderación de contenidos, recomendaciones estandarizadas), la IA ya está pisando con fuerza.

Si imagináramos una gradación, podríamos decir que hemos pasado de la *mecánica* (aplicar reglas fijas manualmente) a la *automatización algorítmica simple*, y de ahí a la *inteligencia artificial generativa*, con cada escalón multiplicando tanto las capacidades como las incógnitas. En el siguiente apartado exploraremos, mediante un enfoque wittgensteiniano, cómo un LLM *aprende* y *sigue* reglas, lo cual nos ayudará a entender mejor por qué a veces produce resultados sorprendentes o contradictorios, especialmente desde la perspectiva del usuario que interactúa con él.

6. LA ADJUDICACIÓN COMO JUEGO LINGÜÍSTICO

Para entender la forma novedosa de aplicación de directrices de los modelos de IA, podemos acudir a la noción de "juegos de lenguaje" de Ludwig Wittgenstein: el significado de las palabras surge de su uso dentro de juegos regidos por reglas compartidas en una forma de vida determinada. Seguir una regla, para Wittgenstein, no es una operación mental privada sino una práctica pública, pragmática, un adiestramiento en ciertos hábitos dentro de un juego social. Aplicado a nuestra cuestión, podemos concebir la interacción usuario-LLM como un *juego de lenguaje* en el que ambos participan, pero —y aquí está el detalle crucial— cada uno juega con un conjunto de reglas parcialmente distinto.

En este juego, el *jugador A* (humano) puede hacer cualquier pregunta o solicitud en lenguaje natural. El *jugador B* (LLM) debe responder de la manera más útil posible, pero respetando una lista de reglas secretas que A desconoce. Por ello, el *jugador B* tiene una ventaja informativa pero también una *carga:* debe autocensurarse o modular su conducta.

El *jugador A*, por su parte, está jugando a ciegas respecto a las reglas; solo puede inferirlas por la conducta de B (sus respuestas). Aquí vemos una similitud con el usuario de un modelo de IA, quien no sabe exactamente qué políticas ni sesgos se le aplican, pero puede ir dándose cuenta empíricamente de ciertas cosas (por ejemplo: cuando pregunta X, el sistema siempre se rehúsa; si lo reformula como Y, puede lograr la generación del contenido).

Wittgenstein nos lleva a analizar este juego observando la *praxis* en lugar de buscar una representación mental interna de los jugadores. El LLM ha sido entrenado con multitud de ejemplos para comportarse de cierta forma: ha interiorizado un repertorio de reacciones. No *comprende y comparte* las reglas, en términos de Hart, como lo haría un juez humano. Su respuesta surge de lo que este autor conceptualiza como mirada externa del derecho: es producto de patrones estadísticos reforzados por entrenamiento humano. Sin embargo, externamente se manifiesta *como si* siguiera reglas.

Wittgenstein nos alerta que:

> "la regla (...) no encuentra aplicación ni en la instrucción ni en el juego mismo; ni es establecida en un catálogo de reglas. Se aprende el juego observando cómo juegan otros. Pero decimos que se juega según tales y cuales reglas porque un espectador puede extraer estas reglas de la práctica del juego" (WITTGENSTEIN 1953).

El significado de la regla se establece por una práctica sostenida en la comunidad. En el caso de los LLM, ¿hay tal comunidad? Del lado del usuario lo que se observa, en realidad, es una situación de permanente asimetría. La única forma que tiene de *aprender las reglas del juego de lenguaje de la IA* es mediante ensayo y error en la conversación. Así, muchos usuarios en la práctica van hallando empíricamente las directrices, lo que constituye en sí mismo un acto de investigación de campo en el derecho.

Esto incluye el fenómeno de los llamados *jailbreaks*, procedimientos para acceder a respuestas que, bajo condiciones normales, deberían ser bloqueadas por el modelo. Todo esto configura un meta-juego: el usuario hábil navega las reglas ocultas de la IA para lograr su objetivo; el usuario inexperto es víctima de una moderación y tutela paternalista excesiva por parte de la IA.

Las políticas deberían ser transparentes —o al menos consistentes— para que el usuario no tenga que *adivinar*. Pero es lo que ocurre. Este ida y vuelta pone de manifiesto que las reglas de los juegos lingüísticos en los que parti-

cipa la IA no son inmutables —ni altamente predecibles—. Para el usuario, esto se traduce en incertidumbre y eventual desconfianza. Si las respuestas cambian sin explicación, la autoridad cognitiva del modelo se resiente.

En suma: a) El LLM y el usuario operan en juegos superpuestos, pero no idénticos, con asimetrías de información sobre las reglas; y b) La coordinación entre usuario e IA es frágil: se basa en la capacidad del usuario de inferir las reglas internas del modelo por ensayo, y en la consistencia del modelo en aplicarlas.

En este sentido, la experiencia del usuario frente al LLM es análoga a la experiencia de la inseguridad jurídica. Para que los LLM puedan desempeñar un rol adjudicativo más o menos confiable, tendrían que lograr un grado muy alto de consistencia, acierto y transparencia en sus reglas de operación. Ello no ha ocurrido hasta ahora. Por el contrario, sigue siendo altamente problemático el fenómeno conocido como *alucinación* de los modelos: afirmaciones factuales incorrectas emitidas con aparente seguridad o autoridad.

Los LLM, por defecto, no explican su proceso interno. Acaso solo lo *muestran* en la respuesta final [3]. Podemos, claro, pedirle al modelo: 'Explicá tu razonamiento paso a paso', pero la respuesta normalmente es una nueva generación de contenido, una racionalización *ex post*. Aquí reaparece el problema: el contenido manifiesto (la nueva respuesta que motiva o fundamenta la decisión) encubre el contenido latente (una preferencia estadística sin justificación lógico-jurídica).

Esto es problemático: naturalizar, a nivel cognitivo, decisiones cuyo contenido lógico-jurídico es hueco o ficticio atenta contra el derecho, de inspiración republicana, a que las decisiones sean fundadas. Con la IA, ¿cómo garantizamos la motivación de los actos? Algunos proponen técnicas de *explainable AI*, técnicas *post-hoc* para que los modelos sean un poco más comprensibles y auditables. Otros enfoques incorporan referencias documentales para las respuestas (por ejemplo, sistemas de *retrieval augmented generation* que citan fuentes 'a libro abierto').

Pero incluso en los escenarios más optimistas, el problema de los sesgos en los datos de entrenamiento persiste. Si entrenamos un modelo con jurisprudencia histórica en la que, por hipótesis, se condena con más severidad a acusados de un origen étnico particular, el modelo incorpora ese patrón. ¿Deberíamos desviarnos deliberadamente de los datos para imponer una serie de valores de justicia?

Los LLM no tienen agenda política, pero *absorben* las de los textos que leen. En la medida que esos textos contengan prejuicios, el modelo los

(3) En los términos de las Investigaciones Filosóficas de L. Wittgenstein.

normalizará. En el *fine-tuning*, el refuerzo humano trata de corregir esto, pero también se introducen las agendas explícitas o implícitas de los desarrolladores, y una estrategia de *compliance* para evitar cualquier forma de responsabilidad. Tenemos una sumatoria de agendas, no una corrección axiológica del modelo (lo que los tecnólogos llaman, con optimismo, 'alineamiento').

Está en juego, nuevamente, la cesión de autoridad cognitiva a la IA. Este fenómeno implica el desplazamiento del criterio para juzgar si el usuario puede acceder a la generación de un contenido, pero también la disputa por el modo en que deben presentarse los contenidos en general. Es una discusión sobre ética, tal como pretenden los tecnólogos, solo en la medida en la que es una discusión política, de autoridad y poder.

7. Conclusión: serenidad ante la IA

Luego de todo lo visto, cabe una pregunta elemental: ¿somos capaces de comprender cabalmente —o de siquiera empezar a entender— lo que supone esta delegación de la autoridad cognitiva a la inteligencia artificial en un tema tan delicado como el derecho?

Martin Heidegger, en su texto sobre la serenidad (*Gelassenheit*), se mostró preocupado por cómo la humanidad enfrenta (o deja de enfrentar) el desafío de pensar su propio avance tecnológico. Una cita:

> "Lo verdaderamente inquietante (…) no es que el mundo se tecnifique enteramente. Mucho más inquietante es que el ser humano no esté preparado para esta transformación universal; que aún no logremos enfrentar meditativamente lo que propiamente se avecina en esta época" (HEIDEGGER 1955).

Aplicada a nuestro tema, esta observación subraya la dificultad de pensar con lucidez dentro de una revolución tecnológica que lo permea todo. Estamos adoptando sistemas de IA en roles cruciales sin haber digerido qué implican para conceptos tan básicos como la identificación de quién está legitimado para adjudicar un caso y quién no, y para otros más complejos, como el punto exacto en el que empezamos a aceptar adjudicaciones no humanas.

¿Podemos confiar en la regulación nacional o internacional de la materia para encauzar estas cuestiones de urgencia? Es una pregunta que amerita un debate extenso. Hay toda una capa de *soft law*, principalmente de inspiración europea, que busca regular la IA. No compartimos que sea posible. Es probable que no exista ningún Parlamento o sistema análogo de creación de normas en el mundo que esté listo para lo que implican los LLM. Pero argumentar ese punto excede el propósito de este trabajo. Volvamos.

Heidegger habla del peligro de caer en un pensamiento puramente calculador, que no medita. Y señala que lo grave no es que no tengamos respuestas para los dilemas de la revolución tecnológica, sino que ni siquiera estamos en condiciones de empezar a pensar en la propia esencia de la técnica, que se sustrae a nuestro pensamiento.

En el contexto jurídico, *pensar* dentro de la revolución de los LLM significa replantearnos nociones fundamentales: ¿Qué es la *decisión* jurídica cuando quien decide es un sistema algorítmico? ¿Cómo se verá el derecho en unos años si modelos de IA sin legitimación político-democrática empiezan a ser la referencia cognitiva principal de la población en temas jurídicos? ¿Qué significa para la interpretación del derecho que los modelos de IA puedan generar cantidades inagotables de argumentos jurídicos en cualquier dirección, e incluso en direcciones opuestas?

Si los modelos de IA pueden emular la comprensión del derecho a partir de su procesamiento del lenguaje natural, ¿eso significa que el derecho mismo, a nivel ontológico, no es más que otro juego lingüístico-social wittgensteiniano? Sería una gran decepción para la discusión clásica entre positivismo y naturalismo.

Un aspecto que ilustra la dificultad de pensar estos temas es la tendencia a la fetichización tecnológica. Y esto aplica a ambos lados del espectro. Podemos caer en una especie de trance por la capacidad sorprendente de la IA para emular nuestro lenguaje o, por el contrario, en un rechazo visceral sin análisis, tecnófobo, que solo detecta banderas rojas por todos lados. Ambas son posturas reactivas, no reflexivas.

Una respuesta es cultivar la *serenidad* (la *Gelassenheit* de Heidegger) frente a la tecnología: ni adoración ciega ni miedo irracional, sino una disposición a usarla, a cuestionarla, a ponerla en su lugar apropiado. En nuestro ámbito, implica integrar la IA en el derecho sin abdicar de los valores jurídicos fundamentales. Controlando, en lo posible, el proceso de delegación que venimos discutiendo.

En conclusión, la forma de adjudicación no humana con IA presenta un panorama de grandes oportunidades (agilidad, acceso, consistencia) pero igualmente grandes desafíos (injusticias automatizadas, opacidad, deshumanización). El equilibrio vendrá de aplicar sabiduría jurídica tradicional —principios de derecho, garantías, controles— a esta nueva herramienta poderosa. Como en toda innovación, hay un periodo de adaptación donde se cometerán errores, pero también se podrá ajustar el rumbo.

Dejamos este último párrafo del primer borrador del texto (*Deep Research* con GPT-5). Todo el resto fue intensamente reescrito y editado, aunque hubo una buena respuesta del modelo en general frente a un *prompt* inicial muy detallado. Esta conclusión inconexa, fuera de contexto, se impo-

ne por *compliance*, pero también por la reproducción de falacias milenarias como el justo punto medio y formas de retórica tranquilizadora que ya no son lugares comunes solo en sentido abstracto, sino que se han cristalizado en inmensas bases de datos. Ahora son sentido común material.

Heidegger puede ayudarnos a entender esta tendencia a la banalidad en los modelos de IA a través del concepto de *Gestell* o entramado: la tecnología moderna tiende a enmarcarnos a nosotros mismos dentro de su lógica instrumental. Un riesgo aquí es que empecemos a ver la adjudicación de controversias, o incluso la administración de justicia, como meros *problemas técnicos* a optimizar.

El entramado tecnológico nos reconfigura: en lugar de ciudadanos reclamando derechos ante un Estado que debe protegerlos, podemos terminar siendo *usuarios* de un sistema opaco. Especialmente cuando el Estado ha defraudado las expectativas sociales en todas las mediciones de opinión, y ha perdido prestigio.

Es un problema epistemológico y moral: mantener vivas la racionalidad práctica y la sensibilidad humanista en un entorno donde la lógica instrumental y la cantidad de información descontrolada e incontrolable pueden abrumarlas y desfigurarlas.

Referencias bibliográficas

Callister, P.D. (2023). *Generative AI and Finding the Law*. Law Library Journal, 117(1). Trad. H. Baca Amenábar: 'IA Generativa: Encontrando la ley', 2023, inédito, disponible en Scribd: https://es.scribd.com/document/789459804/02-CALLISTER-IA-Generativa- y-Encontrando-la-ley-1#:~:text=Habra%20un%20cambio%20en%20la,La%20autoridadHart, H.L.A. (1961). *El concepto de Derecho*. Trad. Genaro R. Carrió (Buenos Aires: Abeledo-Perrot, 1994)

Heidegger, M. (1955). *Serenidad* -Gelassenheit, trad. Helí Morales (Santiago: Ed. Universidad Bolivariana, 2011).

Wittgenstein, L. (1953). Investigaciones Filosóficas - trad. A. García Suárez y U. Moulines (UNAM, 1968).

Dialética do automatismo: um ensaio sobre surrealismo e inteligência artificial

Por Guilherme Gonçalves Alcântara (*)
y José Flôr de Medeiros Júnior (**)

Resumo: O presente ensaio procura estabelecer um diálogo entre o Movimento Surrealista, a partir da discussão sobre o Manifesto de 1924 publicado por André Breton e a Inteligência Artificial, que ao longo do texto será denominada apenas de IA. Da forma como o espaço-tempo esteve presente nos questionamentos de Breton e dos surrealistas que vieram a fazer parte do movimento desencadeado a partir de 1924, o espaço-tempo é novamente posto no centro do debate teórico no mundo do Direito, no território da Literatura, da Filosofia ou, mais a propósito, em Direito e Literatura. Não podemos desconsiderar que à medida que a ameaça do nacional-socialismo crescia e uma Segunda Guerra Mundial se aproximava, os surrealistas da Europa Ocidental se dispersaram, alguns se escondendo, outros, no exílio. Breton fugiu para Nova Iorque, onde uma massa crítica de membros exilados organizou, em 1942, uma exposição a que chamaram *Primeiros Artigos do Surrealismo*, referindo-se às aplicações burocráticas dos artistas emigrados. Hoje, ao passo que o processo de desterritorialização avança no mundo global no campo das comunicações, a IA marca um novo território onde o pensar é desmerecido e o algoritmo é elevado à condição de divindade em um novo Olimpo. Não mais falamos em quem pensa, mas no que o algoritmo pode proporcionar ao humano, cada vez mais automatizado. Estamos, portanto, exilados na IA.

Palavras-chave: Surrealismo; Inteligência Artificial; literatura; direito.

(*) Professor do Instituto Federal do Sul de Minas (campus Machado/MG); Doutorando em Direito pela Universidade Federal de Minas Gerais; email: guilherme.alcantara@ifsuldeminas.edu.br.; Lattes: http://lattes.cnpq.br/3545235149164538

(**) Mestre em Direito com concentração em Filosofia do Direito (UFPB); Mestre em Direito Econômico (Unipê); Pesquisador do Grupo de Pesquisa em Direito e Literatura Phronesis; Pesquisador do Grupo de Pesquisa em Direito e Literatura Legis Literae - Nepedill; Membro da Rede Brasileira de Direito e Literatura - RDL; E-mail: jfmjmedeiros@gmail.com Lattes: http://lattes.cnpq.br/1052314496186934

Dialéctica del automatismo: un ensayo sobre surrealismo e inteligencia artificial

Resumen: El presente ensayo busca establecer un diálogo entre el movimiento surrealista, a partir de la discusión sobre el Manifiesto de 1924 publicado por André Breton, y la inteligencia artificial (IA). De la misma manera en que el espacio-tiempo estuvo presente en los cuestionamientos de Breton y de los surrealistas que formaron parte del movimiento iniciado en 1924, el espacio-tiempo vuelve a situarse en el centro del debate teórico en el mundo del derecho, la literatura, la filosofía o, más específicamente, en derecho y literatura. No se puede ignorar que, a medida que crecía la amenaza del nacionalsocialismo y se aproximaba la Segunda Guerra Mundial, los surrealistas de Europa occidental se dispersaron: algunos se escondieron y otros partieron al exilio. Breton huyó a Nueva York, donde una masa crítica de miembros exiliados organizó, en 1942, una exposición que llamaron primeros artículos del surrealismo, refiriéndose a las aplicaciones burocráticas de los artistas emigrados. Hoy, en tanto el proceso de desterritorialización avanza en el mundo global de las comunicaciones, la IA marca un nuevo territorio donde el pensamiento es desvalorizado y el algoritmo se eleva a la condición de divinidad en un nuevo olimpo. Ya no hablamos de quién piensa, sino de lo que el algoritmo puede proporcionar al humano, cada vez más automatizado. Estamos, por tanto, exiliados en la IA.

Palabras clave: Surrealismo; inteligencia artificial; literatura; derecho.

Dialectics of automatism: an essay on surrealism and artificial intelligence

Abstract: This essay aims to establish a dialogue between the Surrealist Movement, based on the discussion of the 1924 Manifesto published by André Breton, and Artificial Intelligence (AI). Just as space-time was present in the questions posed by Breton and the surrealists who part of the movement were initiated in 1924, space-time is again placed at the center of theoretical debate in the fields of Law, Literature, Philosophy, or more specifically, Law and Literature. It cannot be overlooked that, as the threat of National Socialism grew and World War II approached, surrealists in Western Europe dispersed: some went into hiding, others went into exile. Breton fled to New York, where a critical mass of exiled members organized, in 1942, an exhibition called First Articles of Surrealism, referring to the bureaucratic applications of the emigrated artists. Today, as the process of deterritorialization advances in the global communications world, AI marks a new territory where thinking is devalued and the algorithm is elevated to the status of divinity in a new Olympus. We no longer speak of who thinks, but what the algorithm can provide to the increasingly automated human. We are, therefore, exiled in AI.

Keywords: Surrealism; Artificial Intelligence; literature; law.

1. Introdução

O embate proposto pelo Movimento Surrealista merece reflexão em um momento em que a palavra perde espaço para os algoritmos e estes passam a representar o "pensar" contemporâneo. O desaparecimento dos manifestos, como o surrealista, tem o espaço territorial ocupado por fotos momentâneas, que não retratam o momento, e a códigos escritos ausentes de reflexão.

Este escrito, portanto, propõe uma leitura da, e sobre, a Inteligência Artificial partindo do Manifesto Surrealista de 1924 com a preocupação de pensar, primeiro, o Surrealismo e, posteriormente, a IA. Sem, entretanto, ser um texto desconexo entre os aparentes temas apresentados e, assim, chegarmos a um caminhar por textos surrealistas que nos ofereceram elementos teóricos à presente escrita. Por outro lado, a posição sobre a IA não demonstra ao seu término uma leitura contrária, meramente opositora. Apresenta, por um olhar surrealista, como a IA vive presente em nosso cotidiano a partir de situações traçadas partindo de algoritmos.

2. 1924, Relógios Derretidos, Monstros Mitológicos e o Ano 2025

O alicerce: 1924, relógios derretidos, chapéus-coco, elefantes, objetos domésticos peculiares, sequências de sonhos. *Paredes erguidas*: incidentes misteriosos, 2016, ChatGPT, cartazes de dormitórios, mais sequências de sonhos, animais fantásticos, maçãs, metaverso, monstros mitológicos, pesadelos, objetos em pele, e o ano de 2025. A palavra surrealismo se tornou uma leitura para o bizarro, o irracional, o alucinatório. Porém, quando surgiu na Europa durante os tênues e turbulentos anos que se seguiram à Primeira Guerra Mundial e que antecederam a Crise de 1929, a ascensão do fascismo e do nazismo e a crise das democracias liberais na antessala da Segunda Guerra Mundial, o surrealismo posicionou-se, não como uma fuga da realidade ou da vida, mas como uma força revolucionária dentro dela: o movimento que visava a libertação total do indivíduo. Estético e político; literário e visual. Originário de Paris, mas em última análise internacional, livre das restrições de um estilo ou meio singular, mas conduzido por um líder carismático (André Breton). O surrealismo era, e é, ambicioso, contraditório e complexo.

O termo *surrealismo* apareceu pela primeira vez nos escritos do poeta francês Guillaume Apollinaire, que o utilizou em 1918 para descrever o desfile de balé de Jean Cocteau e a sua própria peça *Os Seios de Tirésias* (Balakian, 1947, p. 104). Para Apollinaire, o surrealismo significava os frutos da imaginação humana libertada da tarefa de imitar a natureza. Suas raízes estariam no *ethos* do romantismo, em figuras excêntricas do século XIX

como o fantasista simbolista Gustave Moreau, o artista autodidata e agente alfandegário Henri Rousseau, o inquietante visionário Arnold Böcklin e escritores como Edgar Allan Poe.

O surrealismo emergiu do espírito desenfreado do dadaísmo: movimento originado em Zurique e Berlim que procurava abandonar as tradições artísticas e perturbar a convencionalidade da vida moderna: direcionando o absurdo contra as instituições que provocaram a catástrofe que foi a Primeira Grande Guerra. No entanto, os surrealistas pretendiam não apenas perturbar a ordem mundial via disparates, mas também reinventá-la mediante táticas experimentais (Chalmers, 2022, p. 92).

A primeira delas é o automatismo ou ação involuntária. Em 1924, André Breton publicou o *Manifesto do surrealismo*, definindo-o como "automatismo psíquico em seu estado puro ditado pelo pensamento, na ausência de qualquer controle exercido por uma razão, isento de qualquer preocupação estética ou moral" (Breton, 2010, p. 19). Breton experimentava esse estado puro desde 1919, quando foi co-autor de um poema sobre o fluxo de consciência do tamanho de um livro chamado *Magnetic fields*. No Manifesto delineou a gênese do Surrealismo, ao relatar a ocorrência incomum durante o estado de sono antes de dormir. Nesse estado, afirmava ter recebido uma frase peculiar: "Tem um homem cortado em dois perto da janela".

É fácil associar esta expressão ao fascínio do Surrealismo por desafiar dicotomias e explorar conexões entre as facetas aparentemente díspares da vida. Mas Breton (2010, p. 16) parece ignorar o conteúdo da frase, enfatizando em vez disso seu método de revelação. O que teve significado não foi o conteúdo específico em si, mas sim a renúncia ao controle consciente sobre a linguagem que se apresentou a André Breton. Concluía que a construção da poesia não deveria ser produto de uma intenção deliberada, mas sim uma rendição ao fluxo criativo.

Pode-se adotar um distanciamento estético de si, ou da estética vigente à época, e se tornar um espectador do processo de pensamento, idealmente assumindo uma apreciação imparcial por ele. Breton defendeu se deixar levar, e se deixou levar, afirmando que os acontecimentos não toleram interferências. Na perspectiva de Breton, esta abordagem da criação era vista como democratizante e afirma isto no ensaio *A Mensagem Automática* de 1933 que o mérito do Surrealismo residia em proclamar a completa igualdade de todos os indivíduos comuns diante da mensagem subliminar, enfatizando que esta mensagem é a herança compartilhada de todos.

Paradoxalmente, apesar do egoísmo pronunciado de Breton, este método serviu como um meio de escapar ao egocentrismo individualista. Michel Carrouges (1950, p. 16–17), um dos primeiros analistas simpáticos do surrealismo, chega a sugerir que Breton identificou a "ligação natural" entre

o inconsciente pessoal, o inconsciente coletivo e o inconsciente cósmico. O momento de inspiração de Breton acabaria por se manifestar como a técnica surrealista de escrita automática — uma tentativa de contornar a mente consciente anotando palavras ou rabiscos em um ritmo que ultrapassa o pensamento.

Para Breton (2010, p. 21), esta técnica radical de abdicação marcou um avanço. Permitiu que os escritores rejeitassem aparentemente a intenção e a ambição, transformando-se em "simples receptáculos" e "modestos instrumentos de gravação". Além de frustrar qualquer tática degradante ou de busca de aprovação, esse método oferecia a "realidade superior de certas formas de associações anteriormente negligenciadas, na onipotência do sonho, no jogo desinteressado do pensamento" (Breton, 2010, p. 19).

Livre das restrições de um desejo racionalista de estabelecer conexões lógicas ou de causalidade, a escrita automática consegue nos surpreender com as intenções e conexões que descobrimos retrospectivamente no que pode inicialmente parecer absurdo aleatório ou coincidência. Breton (2010, p. 19) afirma que a escrita automática, entre outras coisas, "tende a arruinar de uma vez por todas todos os outros mecanismos psíquicos e a substituí-los na resolução de todos os principais problemas da vida".

"'Transforme o mundo', disse Marx, 'Mude a vida', disse Rilke. Para nós, é o mesmo", escreveu Breton (2001, p. 285). O que, em outras palavras (nossas) significa que ao transformar a vida mudamos o mundo. As revistas surrealistas estavam repletas de relatos de sonhos, escritos sobre fluxos de consciência, exercícios de desenho automático e até mesmo questionários rápidos; tudo servindo para libertar o inconsciente, contornando o controle da mente consciente. Embora a afirmação possa parecer hiperbólica, ela antecipa a retórica contemporânea frequentemente utilizada para defender um tipo semelhante de passividade, como veremos.

Operações de acaso objetivo ou encontros capitais. Os surrealistas se envolveram em operações aleatórias em busca do automatismo, muitas vezes terceirizando o controle criativo de seus materiais. Em cada caso, entregar o controle ao processo e aspirar ao que Breton chamou de maravilhoso ou aquela parte do EU que está além do alcance da RAZÃO. Eles também terceirizaram o controle criativo de si. Praticavam jogos literários e de desenho, como o famoso *corpo esquisito*, onde o poema ou imagem final era produto da imaginação coletiva. *Acaso* para o surrealismo também significava longas caminhadas sem direção registradas nos romances *Camponês de Paris* e *Nadja*.

Incluía encontros casuais com os chamados *objetos encontrados*, como os que Breton descreveu ter visto em brechós, que ele afirmava poder resolver tanto distúrbios psicológicos quanto bloqueios artísticos. A frase do

poeta Conde de Lautréamont (2006, p. 163), "tão belo [...] como o encontro casual de uma máquina de costura e um guarda-chuva numa mesa de operações", tornou-se uma espécie de bordão para o surrealismo em sua busca pelo inesperado. Mas as categorias de *acaso objetivo* e/ou *encontro capital* são também os fundamentos de uma teoria da história surrealista.

O surrealismo sobre esse tema iniciou com uma pesquisa na revista *Minotaure*: "Você consegue dizer qual foi o encontro crucial em sua vida? Até que ponto esse encontro lhe pareceu, e ainda parece, acidental ou necessário?". Após receber as respostas, Breton e Paul Éluard apresentam algumas reflexões, explorando pensadores como Aristóteles, Cournot e Poincaré, para apresentar uma conclusão preliminar, influenciada pela combinação de Engels e Freud, sugerindo que o acaso objetivo seria a forma da necessidade exterior se manifestar através do inconsciente humano.

A reflexão surrealista sobre as determinações mais amplas do movimento histórico combina as descobertas da psicanálise sobre o desejo e o inconsciente. O acaso objetivo vai além da mera contingência, incorporando as determinações do desejo e o desejo como uma força determinante na constituição do sujeito. Os encontros capitais, produzidos pelo acaso, representam uma quebra nas expectativas de reprodução da vida reificada. Em uma sociedade dominada pelo poder do capital, onde a mercantilização de todos os aspectos da vida é uma tendência, a ligação não mercantil entre dois seres desejantes é improvável, mas possível e necessária nos tempos-espaços intersticiais.

O surrealismo assume, assim, a tarefa revolucionária de romper com a temporalidade da exploração e opressão, indo além dos movimentos de progresso ou regressão em torno do já existente. A busca surrealista é uma intervenção total no espaço-tempo, defendendo a procura errante pelos sinais que antecipam os encontros capitais, uma vez que eles representam uma afirmação de um *materialismo mágico* possível terminando por redefinir o espaço-tempo existente anteriormente e modificando a concepção territorial.

Em última análise, os surrealistas acreditam que se pudéssemos encontrar significados que existem além da racionalidade. Poderíamos ser libertos da tirania do mundano, da lógica, e descobrirmos verdades mais reais do que a apresentada pela realidade à época. Além disso, o surrealismo nos dá uma forma de pensar sobre a ligação entre a liberdade criativa individual e a libertação coletiva. Ao mesmo tempo que nos lembra quão frágil essa ligação pode ser sob a ascensão do fascismo, quebrando as fronteiras entre os sonhos e a vida desperta, entre culturas e entre si, o surrealismo continuou a ser um exemplo de como nos reimaginarmos poderia ser o passo definitivo para mudar a realidade. Em retrospecto, pode parecer ingênuo, mas, em

retrospecto, quase tudo parece ingênuo. Isso não significa que deveríamos parar de tentar.

Aqui está um tratado publicado por um grupo sueco de surrealistas em 1991, uma crítica convincente da esfera pública contemporânea e da forma como ela tem sido hoje capaz de se apropriar da vida e reduzi-la à sua própria medida. Em *The public sphere and curiosity*, eles apontam como a esfera pública contemporânea possui uma natureza dupla (Berg; Bergh; Bergmark [et. al.]. 2001, p. 135). Por um lado, exibiria uma fachada *externa*, funcionando não apenas como um instrumento de propaganda para a ordem estabelecida e um mediador entre várias facetas da ideologia burguesa, mas também reivindicando uma representação mais ampla da realidade.

Sua natureza expansiva obriga-o a *iluminar* — em outras palavras, reinterpretar, transformar ou mesmo substituir aspectos gradativos da realidade por sua própria versão, moldando-a assim em alinhamento com os interesses, normas e realidades da ordem estabelecida. Esta característica conferiria uma qualidade alienante, mas transcendente, à esfera pública, ao ter a "ultrapassar" a vida, assumindo potencialmente o seu papel ou apropriando-se dela (Berg; Bergh; Bergmark [et. al.]. 2001, p. 135). Por outro lado, o seu aspecto *interior* seria revelado através do crescentemente significativo (e economicamente vantajoso) reflexo das suas próprias realidades, caracterizado por uma autoabsorção que se torna mais narcisista e cinicamente eclética devido aos efeitos prejudiciais da crise e da degeneração na sociedade burguesa (Berg; Bergh; Bergmark [et. al.]. 2001, p. 135–136).

Através da formalização e da "estetização" em todas as suas interações, utilizando critérios arbitrariamente intercambiáveis, mas pré-fabricados, e colocando a cultura na condição de papel central, a esfera pública nos entrelaça enquanto construiria barreiras entre nós e a vida terminando por nos distanciar da própria vida e do viver. A resultante alienação do excedente seria então ativada, levada a participar, e é assim reproduzida — em oposição à esfera pública burguesa mais antiga e à das antigas nações ditas "comunistas" e ou "socialistas" — como uma condição normal, ativamente, embora muitas vezes inconscientemente (ainda que por vezes abertamente), através da esfera pública.

O domínio público abrangeria não apenas instituições e indivíduos representativos, mas também um processo de "socialização": um meio, uma linguagem que existe em si com o seu próprio simbolismo e lógica distorcidos e empobrecidos (Berg; Bergh; Bergmark [et. al.]. 2001, p. 136). A linguagem adaptada ao mercado deteria um poder imenso, pois apelaria a todos os sentidos e fornece soluções diversas, por vezes contraditórias, para necessidades de vida não satisfeitas. Isto resultaria na expropriação ou pelo menos na neutralização de toda a originalidade e liberdades (relativas) por processos que reproduziriam perpetuamente estruturas estabelecidas

(Berg; Bergh; Bergmark [et. al.]. 2001, p. 136). Na esfera pública, as distinções entre ator e espectador se confundiriam e o domínio seria caracterizado por um comércio de informação, maximizando a sua eficácia no serviço à ideologia dominante e aos interesses comerciais. Esta informação supérflua e fragmentada contribuiria para um assalto contínuo aos sentidos, levando a uma degradação do pensamento.

Neste quadro, as interações sociais se tornariam valores de troca, sendo a intercambialidade pessoal considerada uma virtude. Os valores de uso da comunicação e da criatividade seriam alienados em benefício da esfera pública e das exigências do mercado. O domínio público não apenas expressaria, mas também se mascararia como uma utopia do estabelecido, apresentando uma imagem da realidade moldada por diferentes mitos que servem como exceções, oferecendo esperança para preencher as lacunas da alienação e da vida dentro deles (Berg; Bergh; Bergmark [et. al.]. 2001, p. 136).

Estes novos mitos seriam representados por "deuses" tangíveis na esfera pública, mantendo a ilusão de que partilham tanto o nosso mundo como a eternidade da "fama". Embora desempenhem uma função semelhante à dos mitos existentes, o seu impacto é mais intenso, embora temporário, refletindo uma maior variedade e velocidade formal. Esta "utopia" fragmentada abrangeria diversas experiências como o sucesso, a riqueza, a felicidade e o amor, oferecendo continuamente possibilidades aos que delas são excluídos. A esfera pública abraçaria a inatingível *Öffentlichkheit* como a imagem da "vida eterna", promovendo uma aceitação implícita de algo que, na superfície, se assemelharia à morte, mas seria, na realidade, uma inatividade alienada. Uma não vida. Perpetuaria uma distração contínua dos sentidos, reproduzindo o pensamento ou a percepção alienados juntamente com as suas expressões "construtivas" — uma dessublimação que não cria, mas antes aceitaria (Berg; Bergh; Bergmark [et. al.]. 2001, p. 137).

A *Öffentlichkheit* atuaria como um ópio às massas, e até mesmo o inconsciente se tornaria cada vez mais público. Neste domínio público não existiria noite; as luzes do palco se estendem infinitamente, ofuscando qualquer coisa que possa emitir sua própria luz. Isto seria alimentado por doses crescentes de dessublimação repressiva, esforçando-se incansavelmente para quebrar o silêncio e abafar as questões existenciais levantadas pela vida — questões que são, ao mesmo tempo, angustiantes e desesperadas, mas também enigmáticas e maravilhosas.

Hoje, em vez de entregar a nossa agência ao "acaso objetivo" e ao poder enigmático e automático do inconsciente coletivo, como defenderam os surrealistas, somos encorajados a ceder aos algoritmos de aprendizagem automática, muitas vezes comercializados como inteligência artificial. Na realidade, somos controlados por eles. Enquanto Breton imaginou verdades profundas misticamente incorporadas nas nossas profundezas inconscien-

tes, os proponentes da IA apontam para os milhares de milhões de pontos de dados reais (ou seja, "registros mecânicos" da realidade), agregando os efeitos observáveis de uma miríade de decisões humanas.

Os sistemas de IA podem identificar "associações anteriormente negligenciadas" nos dados sob demanda, imbuindo essas correlações com uma aura de verdade. Os sistemas preditivos são amplamente empregados para resolver os "principais problemas da vida", facilitando processos simplificados em vários domínios. Eles são implantados para selecionar *feeds* de mídia social, personalizar experiências de varejo on-line e automatizar a tomada de decisões nos setores bancários, de serviços governamentais e no sistema jurídico. Neste contexto, a IA encarna a concretização do que, para Breton, era uma crença especulativa em procedimentos de tomada de decisão que ultrapassaram o cálculo humano. A aceitação (imposição, na realidade) generalizada do "algoritmo" como uma divindade tecnológica, transmitindo uma aura oracular, atestando o sucesso da revolução surrealista.

A IA atende à nossa passividade, oferecendo aparentemente os frutos da criatividade e do autoexame sem esforço e dúvidas. Os algoritmos nos cativam consistentemente e testemunham perpetuamente nossos desejos insaciáveis, apresentando o que deveríamos desejar. O envolvimento com a IA na vida cotidiana nos torna surrealistas "padrão", cedendo a processos automáticos opacos que não necessitam mais de rituais analógicos intrincados que lembram os tabuleiros Ouija. Modelos de processamento de linguagem natural, como os desenvolvidos pela OpenAI, como o CHATGPT, parecem estabelecer uma ligação clara entre o surrealismo e a IA. O CHATGPT, lançado em 2020, prevê frases subsequentes com base na análise estatística de bilhões de palavras da Internet quando fornecido com uma solicitação textual.

O preenchimento de suas instruções pode ser interpretado como uma resposta de "média social", semelhante a um mecanismo de busca oblíquo da consciência coletiva, livre de restrições sociais contextuais. O GPT parece satisfazer as aspirações de Breton pela escrita automática, gerando texto alienado da agência humana, mas possuindo um sentido discernível que os leitores podem extrair ou projetar nele. Alinha-se com a visão de Breton de que a razão tem um papel limitado, observando e apreciando o fenômeno luminoso da linguagem não premeditada que ainda pode ser analisada. A contrapartida visual do GPT pode ser encontrada em imagens geradas por IA, que assumiram várias formas nos últimos anos. Sistemas como o *DeepDream* do Google, projetados para documentar como funcionam os algoritmos de reconhecimento de imagem, produzem vídeos surreais onde formas proeminentes dos dados de treinamento emergem grotescamente. Outros, como o *StyleGAN* da Nvidia, podem criar rostos humanos realistas.

Alguns algoritmos, construídos em imagens artísticas, produzem "pinturas" no estilo de épocas ou artistas específicos, como o projeto *Next Rembrandt* da Microsoft. O *Dall-E* da OpenAI gera imagens em resposta a solicitações de texto, usando um modelo semelhante ao GPT treinado em sequências de píxeis, em vez de palavras. Os artistas adotaram essas ferramentas para criar trabalhos generativos, às vezes caracterizados como exemplos de criatividade mecânica. Leonel Moura, apresentado na exposição de arte de IA *Artists & Robots* de 2018 no Grand Palais em Paris, associa a sua prática aos *artbots*, ligando-se diretamente ao surrealismo e ao seu esforço para "remover a consciência humana do circuito" (Moura, 2018, p. 3).

Stephen Marche (2021) comparou o uso do GPT a práticas antigas de invocação de musas. Em um ensaio para n+1, a crítica Meghan O'Gieblyn traça paralelos entre o GPT e a escrita automática, destacando semelhanças entre os textos produzidos pelo GPT e aqueles gerados por meio de técnicas de escrita automática, como *The Magnetic Fields* de Breton e Philippe Soupault. Os resultados textuais gerados por modelos como o *OpenAI* são simultaneamente mecânicos e imprevisíveis. Os seus fundamentos residem inteiramente em cálculos e dados históricos, mas exibem originalidade e engenhosidade, apresentando combinações que os humanos talvez não concebam naturalmente.

No entanto, a facilidade com que a IA e o Surrealismo se alinham levanta suspeitas sobre ambos. O surrealismo pretendia libertar as mentes da racionalidade burguesa, eliminando a intencionalidade e explorando níveis profundos de experiência mítica através da aleatoriedade, dos sonhos e da celebração do "primitivismo" e da "ausência de subjetividade". No entanto, esta experiência é prontamente desencadeada pela IA, a própria personificação de ideias redutivas sobre como as mentes funcionam.

Em *Compulsive Beauty*, Hal Foster (1993) argumenta que os surrealistas aproveitaram o "resíduo irracional" deixado pelos processos de produção mecanicistas e capitalistas para minar os modernismos que valorizam a objetividade industrial. A sua fixação em híbridos misteriosos e trabalhadores robotizados poderia ser vista como uma crítica, uma forma de autonomismo que parodia o mundo da automação. Contudo, a crítica do Surrealismo revelou-se vulnerável à cooptação capitalista. Na era pós-moderna, Foster (1993, p. 209–210) reconhece que o real se tornou surreal, questionando se as paisagens oníricas do surrealismo ainda oferecem uma fuga da convenção burguesa ou apenas representam uma expressão e experiência comum dela:

Os sistemas de IA facilitam, sem esforço, a geração de imagens surreais ou a execução de experimentos surrealistas. Eles não exigem rigor, mas encorajam-nos a deixar ir vendo as tentativas de previsão da IA como uma forma de jogo ou um estado de sonho. O substancial poder de processamento

e capacidade de assimilar abundantes dados sobre as nossas preferências permitem-lhes construir e explorar o reino imaginativo em nosso nome. Em uma discussão com as curadoras do MoMA, Michelle Kuo e Paola Antonelli, em novembro de 2021, o artista Refik Anadol descreveu o treinamento de IA nos dados da coleção do museu como o mapeamento de um vasto "espaço latente" multidimensional de "outras criações, imaginações e resultados", um processo que Kuo vê como produzindo "imagens totalmente fantásticas, quase automatistas, ou como escrita ou desenho automático" (Anadol; Reas; Kuo, 2021).

Outra perspectiva sobre os sistemas de IA é que eles navegam via inúmeras combinações de ideias em seus próprios termos, inundando-nos com possibilidades como se essas fossem as únicas. Conforme a teórica da mídia Joanna Zylinska (2020, p. 68) em seu livro *AI Art: Machine Visions and Warped Dreams*, ecoando os pensamentos do escritor polonês Jacek Dukaj, a IA amplia significativamente o conhecimento compartilhado por especialistas em marketing sobre nossos desejos e fantasias.

No entanto, quando o Surrealismo é assimilado pela IA, não conduz à libertação, antes contribui para um aprisionamento nos padrões culturais existentes, os quais são reorganizados de formas novas, mas não fundamentalmente transformadoras. Em vez de oferecerem mudanças, estes padrões tornam-se mais profundamente enraizados. A noção de escapar ao controle da razão resulta, em última análise, numa submissão a uma forma distinta de programação, sucumbindo ao que um modelo de aprendizagem automática pode gerar e aos incentivos de formas algorítmicas de controle. A OpenAI prevê a integração do GPT-3 em vários aplicativos que requerem a geração de texto espontâneo. Isso poderia envolver a criação de personagens não-jogadores mais dinâmicos para jogos ou a facilitação de conversas informais automatizadas em ambientes de atendimento ao cliente.

A tecnologia tem o potencial de transformar os motores de busca numa interação conversacional entre humanos e máquinas. Esta convergência entre o surrealismo e a IA está se tornando um modelo de negócio, reminiscente de artistas como Dalí, que há muito reconheceram e exploram seu potencial. Outra manifestação desta simbiose é aparente nos NFTs de arte generativa (*Crypto Kitties* e *Bored Apes*), onde o conteúdo associado a cada token funciona como pretexto para especulação financeira. A falta de intenção humana na imagem gerada garante que as ideias não interfiram nas negociações, e o legado surrealista de criação "automática" possibilita manter a afirmação de que estas ainda são obras de arte. Por outro lado, a natureza dos experimentos dos artistas com *machine learning*, exemplificada pelas listas inventivas de Shane ou pelos robôs de pintura de ação de Moura, permite, também, domesticar a IA. Fornece uma estrutura para compreender ou aliviar o impacto de falhas e anomalias, tornando a IA mais aceitável,

especialmente quando se esforça para ditar opções e enquadrar a percepção de escolhas possíveis.

O uso da tecnologia para nos inibir, e controlar, pode ser vivenciado com admiração e surpresa. No entanto, é necessário reconhecer que as nossas tentativas de direcionar a IA para resultados que nos divertem podem funcionar simultaneamente como uma forma de vigilância, contribuindo para a recolha de dados que refina as suas capacidades potencialmente nefastas. O surgimento da funcionalidade de previsão de texto é cada vez mais prevalente e espera-se que se estenda a vários produtos de consumo. Embora apresentada como uma conveniência, esta forma de controle altera sutilmente a forma como pensamos. Não deixa de ser surreal.

Referências

Anadol, Refik; Reas, Casey; Kuo, Michelle [et. al.]. *Modern Dream: How Refik Anadol Is Using Machine Learning and NFTs to Interpret MoMA's Collection.* Magazine of the Museum of Modern Art. 15 de novembro de 2021. Disponível em: https://www.moma.org/magazine/articles/658

Balakian, Anna. *Literary Origins of Surrealism. A New Mysticism in French Poetry.* Columbia University Press, 1947.

Berg, Aase Berg; Bergh, Kajsa; Bergmark, Johannes [et. al.]. *The public sphere and curiosity.* In: Richardson, Michael; Fijałkowski, Krzysztof. *Surrealism against the current: tracts and declarations.* London, Pluto Press, 2001, pp. 135-138.

Breton, André. *Manifestos do surrealismo.* (Tradução de Sérgio Pachá). Rio de Janeiro: Nau, 2001.

Breton, André. *First manifesto of surrealism [1924].* Trans. AS Kline. *Poets of Modernity,* v. 30, 2010.

Carrouges, Michel. *André Breton et les données fondamentales du surréalisme.* Gallimard, 1950.

Chalmers, Madeleine, *Verbal techniques,* in: Strom, Kirsten (Ed.). *The Routledge Companion to Surrealism.* Taylor & Francis, 2022, pp. 90-97.

Foster, Hal. *Compulsive Beauty.* Cambridge: MIT Press, 1993.

Gir aldes, Marcus. *O acaso e o desencontro: das manifestações de 2013 ao golpe de 2016.* Garamond, 2017.

Horning, Rob. *Word Processing: The Surrealist practice of automatic writing finds a troubling analogue in a new wave of sophisticated Al text generators. Art in America.* Vol. 110, n. 4, p. 78-83, 2022. Disponível em: https://www.artnews.com/art-in-america/features/word-processing-surrealism-artificial-intelligence-1234625624/

Lautréamont, Conde de, Lautreamont. *Los cantos de Maldoror.* Sed de Belleza, Santa Clara, Cuba, 2006.

Marche, Stephen. *The computers are getting better at writing. The New Yorker,* 30 de abril de 2021. Disponível em: https://www.newyorker.com/culture/cultural-comment/the-computers-are-getting-better-at-writing

Moura, Leonel. *Robot art: an interview with Leonel Moura*. In: *Arts.* MDPI, 2018.

O'Gieblyn, Meghan. *Babel: Could a machine have an unconscious?* n+ 1, v. 40, n. 1, 2021.

Stallabrass, Julian. *Sublime calculation. New Left Review.* vol. 132, nov/dez, 2021. Disponível em: https://newleftreview.org/issues/ii132/articles/julian-stallabrass-sublime-calculation

Zylinska, Joanna. *AI art: machine visions and warped dreams.* Open Humanities Press, 2020.

Delírio e pensamento: alucinações em modelos de linguagem generativa

Por Samuel Medeiros Andreatta (*)

Resumo: O ensaio aborda a produção de distâncias entre o pensamento humano e de máquinas a partir de considerações sobre a alucinação de modelos de linguagem generativa. Em um primeiro momento, o ensaio aborda a proposição de pesquisadores da Open AI quanto ao fornecimento de respostas alucinatórias e os padrões avaliativos que as geram. Posteriormente, aborda-se a concepção de Klossowski sobre a produção de delírios em Nietzsche em uma tentativa de tensionar o encadeamento do pensamento e a construção de realidade. Por fim, as abstrações são mobilizadas na intenção de usar o modelo de Klossowski para tensionar a estabilidade do conhecimento humano e as formas de produção de realidade que nos cercam.

Palavras-chave: Inteligência Artificial; filosofia do direito; filosofia pós estruturalista.

Delirio y pensamiento: alucinaciones en modelos de lenguaje generativo

Resumen: El ensayo aborda la producción de distancias entre el pensamiento humano y el de las máquinas a partir de consideraciones sobre la alucinación en modelos de lenguaje generativo. En un primer momento, el ensayo examina la proposición de los investigadores de OpenAI respecto al suministro de respuestas alucinatorias y los patrones evaluativos que las generan. Posteriormente, se aborda la concepción de Klossowski sobre la producción de delirios en Nietzsche, en un intento de tensionar el encadenamiento del pensamiento y la construcción de la realidad. Finalmente, las abstracciones se movilizan con la intención de utilizar el modelo de Klossowski para tensionar la estabilidad del conocimiento humano y las formas de producción de realidad que nos rodean.

Palabras clave: Inteligencia artificial; filosofía del derecho; filosofía postestructuralista.

(*) Doutorando do Programa de Pós-graduação em Ciências Criminais (PUCRS). Advogado Criminalista. Pesquisador do grupo "Política da Criminologia" (PUCRS). Pesquisador do grupo LASINTEC (UNIFESP). Email: samuelandreatta@hotmail.com

Delirium and thought: hallucinations in language models

Abstract: The essay addresses the production of distances between human thought and machine thought based on considerations about the hallucination of generative language models. At first, it discusses the proposition by OpenAI researchers regarding hallucinatory responses and the evaluative patterns that generate them. In a second moment, it turns to Klossowski's conception of the production of delusions in Nietzsche, in an attempt to problematize the chaining of thought and the construction of reality. Finally, the abstractions are mobilized with the aim of using Klossowski's model to challenge the stability of human knowledge and the forms of reality production that surround us.

Keywords: Artificial Intelligence; philosophy of law; poststructuralist philosophy.

1. DISTÂNCIAS E APROXIMAÇÕES

Introduzimos a o tema das inteligências artificiais apresentando brevemente alguns estudos que se debruçam não sobre definições herméticas de processos tecnopolíticos, mas as medidas de formas de distanciamento e aproximação entre o agir humano e as estruturas maquinais.

Virilio (1995) anunciou a chegada de uma máquina de visão, imagem do ápice da evolução de nossas próteses visuais: fotografias, filmes, microscópios e imagens virtuais passaram a traduzir a realidade. A emergência da internet e as expectativas sobre esse novo espaço como fator de catalisação da resistência social parecem ter sido, de certo modo, frustradas por um atrito posicional em cada uma das camadas da rede, desde suas demandas físicas às suas práticas de assujeitamento, do domínio político-econômico dos provedores de conexão à concentração de acessos e monopólio da rede em determinadas plataformas, o que Mathiesen (2013, p. 36) chama de funis.

A naturalização e aceitação pacífica do entendimento da capacidade interpretativa maquinal integra o mesmo processo farmacológico descrito por Stiegler (2010) que vê na transferência do saber técnico à máquina um fator de proletarização. A busca por um olhar objetivo, nos levou a uma "automatização da percepção" (Virilio, 1995, p. 77) que é operacionalizada, acima de tudo, pela delegação de uma "análise" da interpretação da realidade "objetiva" à máquina.

Haraway (2020) expande o desejo transhumanista ciborgue na superação de análises reducionistas. O pensamento ciborgue rompe com os determinismos biológicos e atenta para a aproximação entre humano e máquina, apontando para nossa inércia desapercebida e desestabilizando a unidade do sujeito. Fisher (1999, p. 7) sugere uma questão que fragiliza as

condições políticas de diferenciação entre humanos e máquinas, usualmente proposta em ermos opostos: o ponto central não seria se as máquinas estão vivas, mas se estamos tão "mortos" quanto às máquinas. Obedientes e (re) programáveis, humanos e máquinas são chamados a participar, a participação é uma das marcas essenciais da Sociedade de Controle cf. Passetti (2003).

A virulência de Land (1992) tem como enquadramento epistemológico um desdém pela antropomorfização, um anti-humanismo que fragiliza a pureza do pensamento humano. O autor (Land, 2021), sintetiza um certo tipo de inteligência de inspiração cibernética, aproximando a inteligência humana e maquinal. Para Land, o Capitalismo é uma inteligência artificial. Sua capacidade de se reinventar e dominar os horizontes políticos do pensamento, tomando a si mesmo como a única realidade e, aparentemente, ignorando os prenúncios de seu fim, compõem uma dimensão importante. As cadeias de valor têm se despersonalizado e se desprendido do valor trabalho em um ritmo exponencial, a dispensabilidade sistêmica do humano passa a exibir de maneira mais flagrante a desterritorialização como movimento contínuo do capital.

Todavia, há ainda aproximações mais concretas, a sede de recursos naturais que essas máquinas demandam [(1)], aliada ao exército de trabalhadores explorados incumbidos de aprimorar os dados que alimentam a inteligência artificial [(2)], funcionam como um contraponto à despersonalização sistêmica de Land.

Poliks e Trillho (2025) defendem que para Land, o computador é uma intensificação de forças psicodramáticas, que propaga uma pulsão de destruição. A análise de Land define a computação como uma entidade assombrada por uma rasa autonomia diante de sua indiferença fundacional em relação ao Bem-estar humano. No entanto, os pesquisadores não estão de acordo com a definição computacional colocada por Land. Land, ao definir essas máquinas desde uma pulsão autônoma, ignora o que os pesquisadores chamam de alteridade do computador, ignora a computação como um "array discreto de alocações específicas para o capitalismo" (Poliks; Trillo, 2025, p. 98). O ponto de encontro com Land é o fato de que o corpo virtual

---

(1) Tanto o gasto energético para o funcionamento dos servidores, quanto a necessidade de resfriamento, demandam enormes quantidades de recursos, como pode ser observado em: Wright, I. Business Energy UK. "ChatGPT Energy Consumption Visualized" Publicado em: 04/09/2025. Disponível em: https://www.businessenergyuk.com/knowledge-hub/chatgpt-energy-consumption-visualized/

(2) A utilização de um modelo de exploração de países periféricos também é reproduzida quando se trata de inteligência artificial quando se trata de moderação de conteúdo, como pode ser observado em: Tan, R; Cabato, R. The independent. "Philippines: Scale AI creating 'race to the bottom' as outsourced workers face 'digital sweatshop' conditions incl. low wages & withheld payments". Publicado em: 30/08/2023. Disponível em: https://www.business-humanrights.org/en/latest-news/philippines-scale-ai-creating-race-to-the-bottom-as-outsourced-workers-face-poor-conditions-in-digital-sweatshops-incl-low-wages-withheld-payments/

substituiu o corpo físico como ponto de ancoragem do capitalismo. Bratton (2015, p. 223) trabalha com um paralelo interessante, assim como mantemos de maneira equivocada a diferença entre humanos e natureza, é equivocado dizer que as máquinas computacionais são apenas ferramentas, pois são "co-usuários". E a pluralidade de conexões, cookies e formas de vigilância na rede, reforça essa caracterização.

Expostas essas considerações que demonstram um certo modelo de inteligibilidade que, inevitavelmente, sequestra a realidade, este ensaio tem a pretensão de começar a mapear um outro tipo de enquadramento: ao invés de delimitar quais as aproximações e distanciamentos existem entre humanos e máquinas, deslocar o olhar para como são produzidos. Para isso, este ensaio realiza um recorte reduzido para tentar arriscar uma posição situacional, intrinsecamente frágil, problematizando a produção de verdade em modelos de linguagem (LLM) para fragilizar a produção de verdade no pensamento humano.

2. CERTEZA E ALUCINAÇÃO.

Em artigo recentemente publicado (Kalai et al, 2025), pesquisadores da "Open AI" abordaram o problema de respostas falsas geradas pelos grandes modelos de linguagem generativa. Essas respostas inventadas pelos modelos de linguagem, chamadas alucinações, acontecem, segundo os pesquisadores, não por equívocos na base de dados, ou na estruturação das escolhas do próprio algoritmo, mas porque os modelos de avaliação das respostas privilegiam a certeza, o que criaria um incentivo para o fornecimento de respostas inventadas.

O artigo usa uma analogia curiosa para explicar como os modelos se comportam diante da necessidade de obtenção da resposta que evoca um caráter típico disciplinar:

> "Como uma analogia, considere o seguinte contexto, no qual humanos também criam informações plausíveis. Os alunos, quando inseguros, podem arriscar um palpite em provas de múltipla escolha ou blefar em exames escritos, submetendo respostas plausíveis nas quais eles têm pouca confiança. Os modelos de linguagem são avaliados por testes parecidos. Em ambos os settings, arriscar uma resposta quando se está incerta maximiza a pontuação esperada sob uma lógica binaria de 0 ou 1 que premia um ponto para respostas corretas, e nenhum ponto para respostas em branco ou por afirmações de desconhecimento. Os blefes são frequentemente confiantes e específicos, como 30 de setembro ao invés de em algum momento na primavera, no caso de uma questão sobre uma data. Muitos benchmarks dos modelos de linguagem espelham provas padronizadas, usando métricas binarias como a taxa de precisão ou de aprovação. Otimizar os modelos para esses benchmarks pode, assim, gerar alucinações.

> Os humanos aprendem o valor de expressar a incerteza fora da escola, na escola da vida. Por outro lado, modelos de linguagem são primeiramente avaliados usando exames que penalizam a incerteza. Então eles operam como sempre estivessem preparados para fazer provas" (Kalai et al, 2025, p. 4. Tradução Livre).

O que os pesquisadores descrevem aqui não é exatamente a forma de um erro gerado por um equívoco na base de dados, ou seja, a mera reprodução de dados fictícios. O que fazem funciona em uma dupla dimensão: ao admitir a falibilidade de um modelo avaliativo que não privilegia o terreno da incerteza, assumem tacitamente a convivência conflituosa de espaços de veridição na produção do conhecimento, chamados por eles de benchmarks, o que exibe nossas heranças frágeis ou, em outros termos, a dependência contingencial do que se entende como objetivo.

Ao admitir a importância da incerteza na avaliação algo paradoxal está em jogo, os pesquisadores deflacionam o jogo posicional de discursos "solucionistas tecnológicos" (Morozov, 2013) assentado na perfectibilidade estatística ou atuarial. Ao mesmo tempo que blindam o próprio algoritmo ao transferir a responsabilidade de respostas válidas aos prompts gerados por usuários ou aos modelos avaliativos internos. Os autores reduzem a incerteza a um número ideal que eles chamam de "calibragem behaviorista" (Kalai et al, 2025, p. 15). A solução proposta pelos pesquisadores é uma ferramenta que exponha um cálculo percentual de confiabilidade nas respostas. Por esse meio reduzem a realidade a um número ideal fundada em uma hierarquização de fontes nebulosa, e que demonstra ignorar todo regime de proteção do direito autoral gerando fricções posicionais que exibem diversos regimes de tolerância e que fogem ao tema aqui proposto.

Isso não significa atender a um relativismo raso, tampouco perceber que as condições de possibilidade de sedimentação do saber estão para além de uma classificação taxonômica, mas atentar para a "certeza" como operadora de sistemas maquinais, como dinâmica que apaga a distância entre simulação e execução, que diminui a distância entre o fabricado e o "real".

Para tentar explicitar essa relação, este ensaio parte de uma referência pouco usual, mas cuja análise aparentemente paradoxal concede uma abertura para tensionar os limites dos modelos de linguagem: o Nietzsche de Pierre Klossowski. A análise seria aparentemente paradoxal por dois motivos evidentes: a ênfase de Klossowski no corpo e estados de consciência de Nietzsche [3] e a "tonalidade da alma", uma das proposições de Klossowski para encarar o problema da marca do autor.

---

(3) A obra dispões de diversas correspondências nas quais Nietzsche descrevia seu estado de saúde, cf. Klossowski (1997, p. 19; p. 233).

Os modelos de linguagem generativa não têm corpo. Têm uma estrutura física, humana, político-econômica; são pessoas jurídicas. Esses modelos não têm alma. Têm um código com todos os vieses que pressupõe a base de dados consultadas, os moderadores, os vieses de consulta, os limites político-jurídico econômicos da plataforma, os limites da programação e da computação etc.

Esse ensaio não sustenta nem um corpo nem uma alma das inteligências artificiais, o que pretende é apontar que o pensamento humano e a razão estatística de modelos de linguagem expressam forças que antecedem a própria formulação de questões e que desestabilizam os limites entre pensar e agir. Além de provocar essa instabilidade na composição da realidade, os recipientes das mensagens em modelos de linguagem como o chat GPT podem ser redirecionados diante dos prompts gerados, caso o modelo entenda que o tópico sendo tratado é sensível ele redireciona os prompts de acordo, no caso de condutas entendidas como ilegais também há um redirecionamento específico [4].

Nos modelos de Linguagem Generativa, as 'forças' são os parâmetros de treinamento e as interações com o usuário: o modelo não é 'livre', mas sempre já determinado por esses condicionantes. Em Klossowski, o domínio que o sujeito acredita ter sobre a própria agência não passa da expressão das forças que o atravessam. Também o pensamento humano, portanto, não é tão 'livre' quanto imagina. Denota-se que as descrições aqui apresentadas tentam evitar a armadilha de atribuir um caráter humano à inteligência artificial, concedendo-lhe qualquer nível de valor existencial.

A preocupação está em apontar, no restrito espaço amostral compreendido pelos conceitos empregados e a especificidade da relação com os modelos generativos de linguagem no que diz respeito às alucinações, as relações pontuais que produzem novas formas de conhecimento nesses modelos para que possamos desnaturalizar as formas de propagação e interação com o saber.

3. ALUCINAÇÕES DEMASIADAMENTE HUMANAS

O que os pesquisadores chamam de alucinações dos modelos de linguagem generativos não se adequa ao que pode ser encarado como "delírio" na

(4) Em postagem no twitter, um pesquisador de modelos de Inteligência Artificial afirma que o algoritmo do chat GPT redireciona as demandas diante de questões sensíveis e questões consideradas como ilegais, disponível em: https://x.com/btibor91/status/1971959782379495785. A questão também foi trabalhada em: Parmar, M. Bleeping Computer. OpenAI is routing GPT-4o to safety models when it detects harmful activities. Publicado em: 29/05/2025. Disponível em: https://www.bleepingcomputer.com/news/artificial-intelligence/openai-is-routing-gpt-4o-to-safety-models-when-it-detects-harmful-activities/

interpretação de Klossowski da obra nietzscheana. No entanto, há um ponto de encontro que indica a interseção entre esses vetores de pensamento:

> "Sem dúvida, é a certeza que leva o sujeito à loucura. Mas este deve ser profundo, um abismo, um filósofo, para se sentir dessa maneira. Todos temos medo da verdade. Mas a força que requer a visão da mais poderosa realidade não é compatível com a mais poderosa força para a ação, para a ação monstruosa, para o crime, até a pressupõe" (Nietzsche, 1997, p. 204).

A que tipo de certeza Nietzsche se refere? Klossowski (1994, p. 204) indica que essa certeza é a certeza da: "profundidade irredutível, cuja mudez não há equivalente, se a certeza produz o delírio é porque a monstruosidade imaginada é apenas o anverso do ato criminoso" (tradução livre). Essa monstruosidade está na realização de Nietzsche quanto a uma (im)possibilidade de consciência total sobre a existência, como exposto por Colli e Mortinari (1980).

Essa certeza de que trata Nietzsche não é a certeza estruturante de modelos avaliativos que levaria a respostas delirantes, mas extremamente confiantes, mas a certeza sobre a instabilidade dos espaços de veridição, uma certeza sobre a incerteza, a potência da dúvida, a falta de valores imanentes à existência, um ceticismo digno de um "samurai" (Veyne, 2008, p. 143). Quando Klossowski fala da monstruosidade imaginada, ele trata da tomada de posição que suspende os valores e enquadramentos assentados. Ou seja, enquanto o delírio (alucinação) dentro de uma notação técnico-computacional se configura pelo fornecimento de repostas erradas, que não condizem com a realidade, derivado de um modelo avaliativo interno que privilegia a certeza, o delírio nietzscheano é a certeza da abertura epistemológica que insiste nos complôs invisíveis que estabilizam os sistemas de pensamento.

O delírio atua como uma potência desestabilizadora que dissipa os limites entre representação e ação, situando a preocupação altamente científica e crítica sobre a marca do autor no jogo de forças contingencial que nos atravessa como indivíduos.

Para Klossowski (1997), Nietzsche alcançaria uma espécie de exterioridade epistemológica a partir de um descolamento do princípio da realidade e identidade, estruturado por um apego à relação recíproca entre forças e pensamento, tomando o corpo como ponto privilegiado de intensificação destas forças por conta de sua fisionomia debilitada. Como nos elucida o Smith (2005), Klossowski fórmula repetidamente a seguinte questão sobre os impulsos: quais parâmetros de valor podem ser atribuídos a eles quando já não se pode apelar a uma ordem transcendente (como em Platão), a uma subjetividade transcendental (como em Kant) ou a uma dialética em desenvolvimento (como em Hegel)? A conclusão é que os critérios precisam surgir do interior dos próprios impulsos. Assim, pode-se afirmar que

as expressões de Nietzsche tentaram se posicionar para além do princípio da realidade, para além do "eu". Smith (2005, p. 2005, p. 12) descreve o empreendimento como uma tarefa quase impossível, pois se propõe a pensar a partir dos impulsos e instintos. Para Klossowski (1997, p. 205):

> "O delírio não reside propriamente no ato monstruoso em si, mas na certeza de que a força para o realizar antecede a capacidade de representá-lo. As noções de 'monstruoso' e 'criminoso' exprimem a pretensão pela qual a visão engendra poder. De um lado, a potência de agir efetivamente deve ocultar-se sob o poder de uma visão considerada mais real do que a própria realidade. De outro, a loucura manifesta-se na certeza de que essas duas dimensões se pressupõem mutuamente, sem que a tensão entre elas possa ser resolvida pela simulação. Assim, já não há nada que distinga dois domínios do real: o simulacro do ato e o próprio ato".

Normalmente, pensamos que a representação precede a ação: primeiro imagino algo, depois ajo sobre isso. Aqui, Klossowski afirma que o delírio surge ao perceber que o poder de agir (o impulso, os vetores do poder) já está presente antes mesmo de podermos representá-lo, imaginá-lo ou compreendê-lo. Em outras palavras, a ação não é a aplicação de um conceito; ao contrário, a visão ou representação é secundária em relação ao excesso de força que torna o ato possível. Esse é o delírio: o reconhecimento de que a força é primária, e o pensamento/representação é apenas uma máscara dessa força.

A visão oculta (dissimula) a força bruta da ação, revestindo-a de imagens, palavras e representações. Mas, simultaneamente, a certeza da visão decorre do fato de que a força para agir já está presente. Elas se pressupõem mutuamente: a visão depende da força da ação, e a ação depende da visão para se manifestar.

Klossowski indica que o verdadeiro delírio advém da percepção de que o pensamento/visão não constitui um domínio seguro e separado, e está intrinsecamente ligado à força, à ação e ao crime. O caráter "monstruoso" reside nesse colapso: imaginar é já agir, e agir é já imaginar. A certeza torna-se insuportável porque funde o simulacro e o ato em uma realidade indistinguível.

A certeza delirante do sujeito, em Nietzsche via Klossowski, é a experiência trágica de que não há fundamento apriorístico do pensamento. Há apenas forças que o precedem. A certeza alucinatória do modelo de IA, ao contrário, não nasce de uma revelação, mas da operação probabilística que produz respostas sem fundamento. Em ambos, a linha entre simulação e ato se rompe: no homem como intuição do abismo, na máquina como efeito técnico do cálculo.

A **alucinação** é um simulacro: uma sequência fluente de tokens que aparenta ser um fato. Mas, ao ser recebida pelo usuário, ela atua no mundo:

influencia decisões, molda crenças e orienta comportamentos. A distinção entre "gerar texto" (simulação) e "produzir um ato" (modelar a realidade) colapsa. Esse colapso é precisamente o risco: o texto alucinado já constitui uma ação, porque representação e ação se fundem na recepção humana. Talvez o mesmo possa ser dito sobre qualquer outro tipo de método de produção do conhecimento, mas com os modelos de linguagem generativa diante de uma maior especificidade das respostas e por conta de uma percebida maior participação do usuário na produção do conhecimento o problema se torna mais agudo.

É por isso que Klossowski afirma que a busca pela verdade não pode ser separada do delírio: o que é comunicado não é apenas "conhecimento", mas também uma força, um humor, uma coloração afetiva. O sujeito é a expressão das forças que o condicionam, o que pode ser pensado é produto das forças situacionais, o questionamento gira em torno do próprio domínio sobre os limites do pensamento e a produção de realidade. Os modelos generativos de linguagem são expressão dos algoritmos que o condicionam, levando os pesquisadores a questionarem modelos avaliativos. O que essa relação pretende demonstrar é que nem o homem nem a máquina produzem realidade livremente: ambos atualizam forças que os programam.

Para Klossowski (1997, p. xix), Nietzsche demonstra que a transmissão do conhecimento pressupõe uma inescapável "tonalidade da alma", um humor, a malha de afetos que acompanha o pensamento de cada um. Cada texto carrega essa "marca do autor". Mesmo quando uma ideia é racional, sua recepção pelo leitor é moldada pelo tom, ritmo, paixão e estilo.

Sistemas generativos de linguagem não têm uma "alma". Mas possuem um estilo, uma tonalidade, produzida pela arquitetura, pelos dados de treinamento e pelas prioridades de avaliação, mas também pela interação do usuário. Um modelo pode apresentar conhecimento com fluência, confiança, equilíbrio ou humildade — mas essas manifestações também não são neutras. Elas são marcas do processo de treinamento, refletindo vieses nos dados, no ajuste de parâmetros (internos e aqueles propostos por usuários) nos modelos de recompensa. Assim, toda resposta de IA transmite não apenas informação, mas também um modo de relação com a verdade (certeza, probabilidade, relação específica com o usuário diante do que foi solicitado). Esse é o simulacro de uma alma digital. A força não reside apenas na arquitetura do LLM, nem apenas no ato do usuário, mas em sua presuposição recíproca.

Em Nietzsche, segundo Klossowski, o delírio da certeza reside em perceber que a força para agir precede a representação do ato. A ação é primária, a representação é uma máscara secundária, e esse colapso torna o próprio pensamento "monstruoso". Há duas camadas que podem ser analisadas quanto à precedência de forças em modelos de linguagem generativa:

o ato humano (a pergunta) ou a comunicação com esses modelos, e o ato do sistema (a arquitetura).

A pergunta do usuário é já um ato que pressupõe força: molda, restringe e provoca o sistema. A alucinação não surge no vácuo. Ela é desencadeada pela pergunta. Nesse sentido, o ato humano é anterior à representação, pois obriga o modelo a gerar uma resposta, independentemente da existência de fundamentação. O LLM é estruturalmente compelido a gerar uma resposta com certeza, mesmo diante de lacunas na representação. Essa compulsão, responder, simular certeza, é a "força" interna ao modelo. Nesse sentido, a arquitetura precede qualquer representação específica, garantindo que o ato de responder aconteça, mesmo que seja alucinatório.

Portanto, não se trata de uma escolha entre humano ou sistema, mas de uma pressuposição recíproca — exatamente o ponto de Klossowski sobre delírio: o ato do usuário (fazer a pergunta) pressupõe o poder do sistema de responder. O ato do sistema (responder) pressupõe a força da pergunta humana que o ativa. O delírio surge porque cada um parece preceder o outro: sem o usuário, não há saída; sem a arquitetura do sistema, a pergunta não se materializa como resposta. Assim, tal como Klossowski descreve o colapso entre ato e simulacro, aqui observamos o colapso do ato humano (perguntar) e do ato do sistema (resposta alucinatória). Não existe um domínio puro de representação fora desse circuito.

A força não reside apenas na arquitetura do LLM, nem apenas no ato do usuário, mas na pressuposição recíproca entre ambos. A pergunta já contém a força que convoca a resposta. A arquitetura garante que uma resposta surgirá, com certeza, fundamentada ou não. A alucinação é, portanto, o delírio dessa reciprocidade: o ato de perguntar e o ato de responder colapsam em um círculo, onde simulação e realidade se confundem.

Referências bibliográficas

Bratton, B. H. (2015). The Stack: On Software and Sovereignty. MIT Press.

Colli, G., & Moniman, M. (1980). Nietzsche, F. Sämtliche Werke. Kritische Studienausgabe. Giorgio Colli and Mazzino Mominan. Berlin: Walter de Gruyter. En Klossowski, P. (1997). Nietzsche and the Vicious Circle. The Athlone Press.

Fisher, M. (1999). Flatline Constructs: Gothic Materialism and Cybernetic Theory Fiction. Tesis de doctorado en Filosofía, Universidad de Warwick.

Haraway, D. (2020). Manifiesto cíborg (Teorías del caos nº 1). Kaótica Libros.

Kalai, A. et al. (2025). Why Language Models Hallucinate. Disponible en: https://cdn.openai.com/pdf/d04913be-3f6f-4d2b-b283-ff432ef4aaa5/why-language-models-hallucinate.pdf. Acceso en: 30/09/2025.

Klossowski, P. (1997). Nietzsche and the Vicious Circle. The Athlone Press.

Land, N. (1992). The Thirst for Annihilation: Georges Bataille and Virulent Nihilism. Routledge Editions.

Land, N. (2021). Teleoplexia: Ensayos sobre Aceleracionismo y Horror. Holobionte Ediciones.

Mathiesen, T. (2013). Towards a Surveillant Society: The Rise of Surveillance Systems in Europe. Waterside Press.

Morozov, E. (2013). To Save Everything, Click Here: The Folly of Technological Solutionism. PublicAffairs.

Nietzsche, F. Ecce Homo. En Klossowski, P. (1997). Nietzsche and the Vicious Circle. The Athlone Press.

Parmar, M. (2025). "OpenAI is routing GPT-4o to safety models when it detects harmful activities". Bleeping Computer. Publicado el 29/05/2025. Disponible en: https://www.bleepingcomputer.com/news/artificial-intelligence/openai-is-routing-gpt-4o-to-safety-models-when-it-detects-harmful-activities/. Acceso en: 30/09/2025.

Poliks, M., & Trillo, R. (2025). Exocapitalism: Economies with Absolutely No Limits. BookPress EU.

Stiegler, B. (2010). For a New Critique of Political Economy. Polity Press.

Smith, D. (2005). "Klossowski's Reading of Nietzsche: Impulses, Phantasms, Simulacra, Stereotypes". Diacritics, 35(1), 8–21.

Tan, R., & Cabato, R. (2023). "Philippines: Scale AI creating 'race to the bottom' as outsourced workers face 'digital sweatshop' conditions incl. low wages & withheld payments". The Independent. Publicado el 30/08/2023. Disponible en: https://www.business-humanrights.org/en/latest-news/philippines-scale-ai-creating-race-to-the-bottom-as-outsourced-workers-face-poor-conditions-in-digital-sweatshops-incl-low-wages-withheld-payments/

Veyne, P. (2008). Foucault: el pensamiento y la persona. Texto y Grafia.

Virilio, P. (1995). Máquina de visión. Signo e Imagen.

Wright, I. (2025). "ChatGPT Energy Consumption Visualized". Business Energy UK. Publicado el 04/09/2025. Disponible en: https://www.businessenergyuk.com/knowledge-hub/chatgpt-energy-consumption-visualized/

Hacia una teoría del Estado de la inteligencia artificial: oclocracia algorítmica y tecnofeudalismo [*]

Por Emilio Ferrero García [**]

Resumen: La inteligencia artificial viene a mover el suelo sobre el que se cimentan los pilares de la Teoría del Estado clásica, irrumpiendo con la fuerza de un huracán la nueva revolución digital y los desafíos que para el derecho suscita. Así las cosas, a lo largo de las siguientes páginas abordaremos desde la ciencia constitucionalista el impacto del desarrollo y utilización algorítmica como punta de lanza contra las democracias constitucionales por parte del nuevo cesarismo y populismo computacional, cuestionando los fundamentos del sistema representativo a sustituir por nueva oclocracia digital. Un Gobierno de visión monista y decisionista desprovisto de contrapesos y limitaciones contramayoritarias, revisando viejas formas revestidas de una apariencia demagógica acampada fuera de las fronteras del Estado de derecho. Del mismo modo, se estudiará el modelo económico y político al que se vincula, desprovisto de las cortinas de humo que envuelven el nuevo tecnofeudalismo. Una estructura extraccionista y depredadora fundada en redes de dependencia y la privatización de la nube, como condición de los nuevos feudos digitales, en manos de una auténtica tecno-oligarquía, situados por encima y al margen del derecho de los ordenamientos jurídicos nacionales, alterando el complejo equilibrio y correlación de fuerzas del orden mundial surgido de la IIGM. Tal fenómeno provoca una nítida desconcentración, desplazamiento y progresivo vaciamiento del poder del

---

(*) Este capítulo se realiza en el marco del Proyecto de Investigación "Desinformación, odio y polarización: la afectación de derechos y libertades de personas vulnerables", concedido, dentro de la convocatoria Proyectos I+D+i 2020 del Ministerio de Ciencia e Innovación del Gobierno de España, con una cuantía de 24.200 euros. IP: Ángela Figueruelo Burrieza, (ref. PID2020-116603-RB-I00), así como en el marco de una estancia en la Facultad de Derecho de la Universidad de Coímbra (Portugal) como investigador invitado. Ha sido financiado con cargo a la convocatoria de concurrencia competitiva USAL 2023, cofinanciada por el Banco de Santander.

(**) Profesor de Derecho Constitucional de la Universidad de Salamanca. Director general del Observatorio Iberoamericano de Derechos LGTBIQ+. Coordinador académico del Máster en Identidad Sexual y Diversidad LGTBI+ y adjunto a la dirección del máster universitario en Estudios Interdisciplinares de Género en la USAL. Secretario adjunto de la Red GENET del Consejo Superior de Investigaciones Científicas (CSIC) y miembro del Instituto Iberoamericano de Derecho Constitucional, así como de los Centros de Investigación de Género y Para la Gobernanza Global. emilioferrero@usal.es.

Estado, agudizando la crisis de representación y autoridad en competencia con otros poderes privados no estatales. Finalmente, se plantearán una serie propuestas de abordaje jurídica de la IA y el establecimiento de un nuevo *check and balance* digital que revigorice la acción del Estado como garante de los derechos fundamentales ante el incierto siglo XXI.

Palabras clave: Oclocracia digital; Teoría del Estado; inteligencia artificial; tecnofeudalismo.

Abstract: Artificial intelligence comes to unsettle the very ground upon which the pillars of classical Theory of the State are built, the new digital revolution and the concomitant challenges to Law bursting forth with the force of a hurricane. Against this backdrop, the following pages address—from the vantage point of constitutional scholarship—the impact of algorithmic development and deployment as the spearhead of an emergent computational Caesarism and populism that assaults constitutional democracies, calling into question the foundations of the representative system and threatening to replace it with a new digital ochlocracy. A monistic, decisionist form of government, stripped of counter-majoritarian checks and limitations, reexamines old modalities now clothed in a demagogic guise and encamped beyond the frontiers of the Rule of Law. In like manner, this study examines the attendant political and economic model, stripped of the smokescreens that cloak the new techno feudalism: an extractivist, predatory structure founded on networks of dependency and the privatization of the cloud, which constitute the new digital fiefdoms in the hands of a genuine technooligarchy. These actors stand above and outside the law of national legal orders, thereby altering the complex equilibrium and correlation of forces that emerged from the Second World War. The phenomenon produces a clear disaggregation, displacement and progressive hollowing-out of state power, intensifying the crisis of representation and authority in competition with non-state private powers. Finally, a series of legal approaches to AI will be proposed, together with the establishment of a new digital system of checks and balances intended to reinvigorate the action of the State as guarantor of fundamental rights in the uncertain twenty-first century.

Keywords: Digital Mob rule; Theory of the State; artificial intelligence and techno-feudalism.

1. Introducción

A lo largo de las siguientes páginas abordaremos una cuestión ciertamente pluridimensional e intrincada, pero de enorme trascendencia y significación para la disciplina jurídica y su tratamiento dogmático de la inteligencia artificial. Una perspectiva nítidamente jurídica y constitucionalista, pero enhebrada con otras disciplinas como la Historia, la Filosofía

y la Ciencia Política, a fin de converger de forma holística en una aproximación científica en torno al estudio del algoritmo desde la teoría del Estado, a propósito del estudio de la oclocracia digital y el populismo computacional. Una amenaza existencial que supone a nuestro juicio uno de los más grandes desafíos en el horizonte jurídico-político de nuestras sociedades demoliberales, no solo frente las disyuntivas que nos presenta, sino ante la realineación de las relaciones económico-productivas y, sobre todo, reconfiguración del Poder como elemento objeto de estudio y limitación de la ciencia constitucional.

Creemos, no obstante, que esta nueva revolución tecnológica, divergente y por tanto innovadora, puede estudiarse, alejándonos del mundanal ruido y los árboles, que a veces no nos dejan ver el bosque, desde la mirada retrospectiva de halcón que otorga la historia y la teoría clásica, de modo que podamos observar, identificar y disertar con mayor precisión, profundidad y alcance a fin de extraer conclusiones sobre su naturaleza y horizonte de futuro.

Por ello, el presente trabajo orbita en torno al concepto de *oclocracia* y su transmutación en la sociedad digital, a fin de poder estudiar sus elementos presentes en la nueva IA y el sistema algorítmico con el consiguiente nacimiento de cesarismos y populismos computacionales, en tanto que actores impugnadores de la democracia representativa y constitucional. Una nueva realidad que, de otra parte, altera el reparto y limitación del poder, desplazado y desconcentrado en los nuevos señores del internet del tecnofeudalismo que se cierne.

Cuestionamientos, amenazas y desafíos no solo para nuestro modelo demoliberal social, sino existencialmente para el derecho mismo y el Estado como sujetos históricos centrales, analizando, pues, su usurpación de funciones y progresivas tentativas de sustitución, anulación o repliegue ante nuevas formas privadas de poder por medio de la titularidad y detentación del algoritmo como nuevos fundos digitales. Finalmente, abordaremos la proyección futura en clave jurídico-política, regulatoria e internacionalista como reivindicación doctrinal del Estado y la Constitución como escudo y motor histórico ante el replanteamiento y reformulación de nuevos pesos y contrapesos para la realidad algorítmica.

2. Conceptualización de la oclocracia digital desde la Teoría del Estado

Fue Polibio quien acuñó en sus *Historias* el término de oclocracia (*okhlokratia*) en su sucesión cíclica de sistemas políticos (*anaciclosis*), concebida como "total degeneración salvaje [de la democracia] hasta encontrar de nuevo un amo y monarca" (1981, VI, 9). El término, hoy en unísono desuso,

deriva del concepto que Platón llamó demagogia en su aristocrática *República* (2020, pp. 565-566), concebido como corrupción de la democracia (*dēmos-kratia*), igualmente presente en la *Política* de Aristóteles, donde la "muchedumbre" ejerce despóticamente el poder por encima y al margen de la Ley (2018, III, 5).

Ciertamente, se podrían evocar los evidentes recelos y escepticismos de los citados autores hacia la democracia, desconfiando de aquellas formas de Gobierno fácilmente tendentes a otorgar el poder al pueblo. No obstante, otros autores muy distinto signo, como sería el caso desde la teoría radical de Rousseau, en su *Contrato Social*, comparten una visión de la oclocracia como degeneración de la democracia, cuyo abuso del gobierno y disolución del Estado conduce a la anarquía (1812, p. 169). Por consiguiente, la interpretación de oclocracia que a lo largo de estas páginas daremos no pretende ser una crítica más o menos velada al sistema democrático, sino por el contrario, una reivindicación exaltada de su naturaleza frente a su oxímoron, el Gobierno iliberal y despótico de las multitudes (Ferreira, 2019, p. 144).

Así las cosas, la actual *oclocracia* algorítmica parte de la pérdida de lo que Cicerón acuñó *consensus iuris* (1991, p. 162), en que se cimenta el sistema político, en un contexto de encadenamiento masivo de diversas y concurrentes crisis cuasi existenciales para el viejo orden mundial surgido de la IIGM y fundado en los DDHH y el Estado social y democrático de derecho. Crisis de legitimidad y representación, económico-financiera, energética y medioambiental, de confianza y autoridad, institucional y social, de identidad y del sistema internacional multilateral. Una fragmentación social que lleva a la polarización y la confrontación de bloques (Sartori, 2005, p. 702), una vez eliminado el crédito y capacidad del Gobierno democrático para su mediación y moderación (Martin Lipset, 1993, p. 72).

Un choque frentista potenciado por unas redes sociales utilizadas como cámaras de eco para la amplificación y filtro algorítmico de aquellos mensajes dirigidos a una sobreestimulación social permanente (Gendler, 2018, p. 40-45), articulando una progresiva radicalización en torno al falso consenso de la *emocracia* y el Gobierno de las vísceras. He ahí la emergencia desestabilizadora de la inteligencia artificial en la cuarta revolución industrial en curso (Schawb, 2016, pp. 13-20) como elemento hegemónico —y contrahegemónico—.

3. EL ALGORITMO COMO PUNTA DE LANZA CONTRA LA DEMOCRACIA REPRESENTATIVA

Empero, el nuevo tecnopopulismo no solo aspira a cavar trincheras antagónicas dentro de lo social por medio de cadenas de equivalencia de demandas heterogéneas como praxis (Laclau, 2004, pp. 81 y 183). Propugna

desde posiciones nacionalistas extremistas la deshumanización de aquellos grupos que considera enemigos, condenados al "afuera" civil *agambeniano* y expulsados a los márgenes de la sociedad (Agamben, 2006, p. 32). No son fines ni medios nuevos, pero sí lo son las herramientas algorítmicas de las que se dotan, equivalentes en cuanto a potencialidad al tránsito del venablo a la bomba nuclear.

Su capacidad para la difusión automatizada de millones de mensajes a voluntad, incluyendo la microsegmentación psicográfica (Winner, 1991, p. 66), desinformación, discursos de odio, deshumanización de grupos o incitación a la violencia contra personas e instituciones, la manipulación de la opinión pública por medio de generación de *deepfakes* son solo algunas de las potenciales, pero contrastadas, amenazas para nuestro orden constitucional. Estas son convenientemente instrumentalizadas para la creación de un determinado clima histórico, político y moral que sirva de caldo de cultivo propicio para el triunfo de fuerzas extremistas del creciente populismo computacional (Zuboff, 2019, pp. 128 y ss.).

Por ello, esos afanes de inmediación real y voladura de los mecanismos intermediarios en la libre conformación de la voluntad popular son oportunamente facilitados por la aparente provisión de posibilidades por parte del desarrollo tecnológico que permite dotar una alternativa fuente de legitimidad destituyente y contrainstitucional pero alabada como participativa permanente. No en vano, y de la mano a dicho fenómeno, los algoritmos han tornado en aparato disciplinario de nuestro momento histórico (Beiguelman, 2021, p. 46).

Y ante esa desconfianza generalizada, consecuencia de la —compleja— incertidumbre tecnopolítica, no son pocas las respuestas simplistas y mesiánicas que ofrezcan opciones impugnatorias del *statu quo* en descomposición (Virio, 2016, pp. 61 y ss.), fijando como objetivo a batir la democracia representativa como tal. No se trata del precio a pagar de las nuevas invasiones bárbaras a cambio de la restauración del orden, sino la condición de posibilidad para quienes los contrapesos del constitucionalismo demoliberal representan cadenas demasiado pesadas en su visión monista y decisionista del poder, aun bajo la apariencia de una presunta democracia directa. Unos contrapesos liberales cuya vigencia la revolución tecnológica viene a poner a prueba decisivamente (Betzu *et al*, 2019, p. 256).

Así las cosas, si la democracia representativa se presenta como voluntad de los muchos por medio de unos pocos, la democracia directa carente de mecanismos institucionales de canalización y autolimitación se convierte en *oclocracia*, es decir, el gobierno de uno por medio de muchos. Porque siempre surge en esos momentos de crisis un tribuno que sirva de ariete para la desintegración social, cuya relación con el pueblo se pretende direc-

ta y sin intermediación institucional; un cesar cruzando cuantos rubicones delimiten la *res publica*.

No obstante, no ha de leerse con ello una crítica a la apertura o mayor participación de la ciudadanía en la toma de decisiones, todo lo contrario, al suponer esta no ya un mecanismo de calidad democrática sino su más profunda raíz y *conditio sine qua non* para su disfrute pleno. Sin embargo, la tentación algorítmica de una supuesta democracia directa no constituye más que un espejismo de empoderamiento popular, pero totalmente carente de mecanismo deliberativo alguno, lo que sabemos antecede a la tiranía de la mayoría que denunciaba Tocqueville (1911, pp. 361 y ss.), aun modernizada con la semblanza tecnológica y neofuturista que representa la brocha gorda de la inteligencia artificial.

4. TECNOPOPULISMO COMO NUEVA FORMA DE CESARISMO

La *oclocracia* necesariamente conduce (*cum-ducere)* [1] necesariamente al cesarismo (Cervelli, 1972, pp. 337-339) porque la muchedumbre (*turba*) —concepto diferente al de pueblo (*populus*) pero fruto de su destrucción—, cuando se le desprovee de su condición material de ciudadanía activa en su más plena acepción, es utilizada para iniciar el proceso destituyente que lleve al nuevo césar al poder sobre las cenizas de las instituciones derrocadas o conquistadas. Nos referimos, pues, a lo que, con buen criterio, desde la doctrina se ha señalado como "dictadura demagógica" como descomposición del Gobierno popular (Bobbio, 1986, p. 128) o "despotismo popular" (Treitchke, 1918, p. 190).

De ese Gobierno embrutecido, de esa omnipotencia nace el tirano que gobierna con el favor de aquellos que trata como esclavos (Mangoni, 1976, pp. 41-61). Una masa forjada a golpes, pero también a *tweets* y *posts*, aterrorizada, confundida y embrutecida, anulados sus mecanismos defensivos para resistir los cantos de sirena que aspiran a restaurar la seguridad a cambio de su libertad. Cada tiempo tiene sus césares, es la demagógica elocuencia violenta de Cleón de Atenas frente Diódoto (Tucídides, 2012, III, pp. 157-166), *duces*, centauros carismáticos, oradores, pistolos, plutócratas, etc., capaces de monopolizar el miedo y el odio en su toma del poder, incluso desde la sacralidad del tribuno (Neumann, 1973, pp. 333 y ss.). Una voladura de toda mediación institucional: de la tentadora *provocatio ad populum digital* de la pseudo democracia directa al embaucador *salve o Popolo d´eroi* del algorítmico (Ferrero, 2025, p. 190).

---

(1) A veces las palabras y su etimología son más elocuentes de lo que cabría imaginar, anticipando el verbo al *Duce* resultante de dicho proceso.

Un proyecto segregador y liberticida, fanático e intolerante en que expulse del cuerpo civil a aquellas personas y grupos en situación de vulneración estructural sometida a *apartheid*, criminalización o persecución; chivos expiatorios relegados a exogrupo, *homo sacer* u *hors l´hummanité* (Schmitt, 2002, pp. 61 y 73). Y, por último, una voluntad viciada y propensa a la evicción que los demagogos saben instrumentalizar (Padilla Gálvez, 2017, pp. 167-178). Y si esa capacidad en nuestro tiempo histórico se encuentra ejercida y amplificada exponencialmente por medios computacionales, entonces la titularidad, desarrollo, implantación, distribución y detentación de la inteligencia artificial irrumpe indiscutiblemente como nueva forma de poder. El *Cola di Rienzo algorítmico*.

5. Tecnofeudalismo y oligarquía algorítmica

Así las cosas, no nos encontramos *sensu stricto* frente una encrucijada de caminos, sino ante dos filos de un *gladium*, dos caras de una misma moneda, tan nociva o más que una espada: la oclocracia algorítmica, de una parte, y la oligarquía tecnológica de otra. Esa plutocracia digital viene a ocupar el rol central de quienes detentaron la propiedad de los medios de producción, en la consecuente transición del sistema capitalista clásico a una nueva forma, como antes se evolucionó del capitalismo mercantil, industrial, financiero y de consumo al capitalismo de la información y los datos o, incluso, de vigilancia (Denemark, 2024, p. 129).

Es entonces cuando entra en escena un nuevo actor y marco indeclinable en nuestra *dramatis personae*: el tecnofeudalismo. No han sido muy extendidas las deserciones doctrinales en torno al despliegue de un proceso de "refeudalización" de nuestras sociedades contemporáneas (Supiot, 2013, pp. 129-145) o al "neofeudalismo" eminentemente focalizado en las relaciones de dominio de las grandes corporaciones transnacionales como amenaza a los derechos fundamentales (Greenwood, 2017, pp. 163-221), recuperando la tesis habermasiana sobre la privatización de las formas de comunicación en la esfera pública contemporánea (*Refeudalisierung*), retrocediendo el espacio ganado por el Estado a los poderes privados (Habermas, 1991, p. 403).

Empero, este concepto pone la luz en la ruptura del contrato social en pro de lógicas de subordinación y exclusión, consecuencia de la evolución de las relaciones económicas, de la concentración de capital a un sistema de extraccionismo masivo de datos una estructuración de redes digitales de dependencia (Varoufakis, 2024, pp. 83-96).

La privatización de internet, equiparable en términos históricos a la privatización de la tierra comunal (*ager publicus*), véase ora la gradual y progresiva encomendación en la Roma pos-imperial, dando nacimiento al

feudalismo medieval canónico, ora las tierras comunales indígenas en la América pos-virreinal con la entronización jurídica de la burguesía como nueva clase dominante y la introducción jurídica mercantilista por medio del Código Civil Napoleónico en el siglo XIX. Y esa equivalencia radica en el dominio ejercido sobre ese "fundo digital", en que rige la jurisdicción y fuero de su señor: pago de nuevos diezmos en forma de datos personales e información que generamos y consumimos. ¿Pero a qué nos referimos cuando hablamos de feudos digitales? Al acceso a la nube, algoritmos, entornos virtuales, plataformas, redes sociales e infraestructuras en línea (Durand, 2020, pp. 123 y ss.).

Tal sistema, de palmarias relaciones asimétricas, requiere, no obstante, para su supervivencia de una capacidad para modelar y moldear el comportamiento humano por medio de algoritmos [(2)], injiriendo en la libre conformación de la voluntad individual y popular al alterar la capacidad de criterio de la ciudadanía para decidir autónoma e informadamente (Nieto Rojas, 2020, p. 326). Es decir, una dependencia garantizada por el algoritmo, fundado en un conocimiento con apariencia de omnisciencia total sobre sus usuarios y su realidad, acompañada de la creación de necesidades, con la nada colateral consecuencia de generación de adicciones, en una absorción configuradora en términos políticos y económicos de una rueda sin cesar alimentada por nosotros mismos. En definitiva, adormeciendo, pero, sobre todo, encolerizando, agitando y movilizando hordas de odio, haciendo honor a la clásica tesis, ocio digital garantizado mediante: *panem et circenses* (Juvenal, 1996, p. 130).

6. Los nuevos señores digitales y su amenaza al Estado de derecho

En vista de la exposición efectuada al sistema tecnofeudal y, puesto que basta un espejo para conocer sus feudatarios, solo nos restaría señalar sus oligarcas. No obstante, a estas alturas podemos identificarlos, tal vez con la mayor dosis de clarividencia que otros momentos históricos nos privan, presentando demasiadas máscaras, cortafuegos, velos y telones.

Sus nombres, pues, a nadie resultan ajenos: X, Tesla, Google, Amazon, Facebook y Meta, Apple y OpenAI; los otrora fulgurantes Plantagenet, Anjou, Lara, Welf, York, Trastámara, Hohenstaufen y Borgoña son los Musk, Bezos, Zuckelberg, Pichai, Altman y Andreessen del presente, si bien desprovistos de la epicidad romántica y caballeresca del sinfín de romances medievales, si bien no por falta de juglares y trovadores en las redes de su propiedad y jurisdicción, como lo fueron las plazas de *signoria* y castillos.

---

(2) A veces creemos que entrenamos a la IA cuando cada vez creo más que es la IA la que nos entrena a nosotros.

Señores de internet cuyos oligopolios ponen en peligro cierto nuestros derechos constitucionales (Rubio, 2018, p. 223).

Así las cosas, y siguiendo dicho paralelismo, resulta a nuestro juicio innegable ya la usurpación de roles estatales por parte de estos poderes privados ora en el marco doméstico, ora en el geopolítico. Empresas ya con mayor capacidad de influencia y presión que Estados, con una capitalización de mercado mayor al PIB de numerosos países [(3)]. Empero, quizá más descarnado sea el ejercicio de señoreaje, históricamente utilizado como atributo distintivo de soberanía (Bodin, 1577, pp. 154 y ss.), concretado en la acuñación de moneda por estos emporios empresariales, *v. g.*: *Libra/Diem* en Facebook o *Amazon Coins* en su homónimo fundo, o criptomonedas como Bitcoin, Tether o Ethereum (Rupeika-Apoga, 2025).

En consecuencia, asistimos a un desplazamiento paulatino, pero decidido, de la soberanía de los Estados a otros poderes privados, revirtiendo el camino que a lo largo de siglos constituyó el paso a la modernidad de la mano de la configuración del Estado (García Pelayo, 2004). No obstante, esta traslación a otros focos de poder viene de la mano del intervencionismo directo de estos señores digitales en política, no solo ejerciendo su presión entre bastidores sino, incluso, renunciando a añejos convencionalismos y actores partidistas interpuestos [(4)]. De esta forma, esta quinta columna tecno-oligárquica actúa como un virus, desmontando y boicoteando la acción estatal, recortando su objeto material, su presupuesto y su actuación.

Por ello, ver a Donald Trump en su toma de posesión en el Capitolio el pasado 2025 recuerda infaustamente a la seo de Frankfort el Meno o a la Capilla Palatina de Aquisgrán en tiempos pretéritos. Príncipes electores del Sacro Imperio Romano Germánico hoy son los tecno-señores feudales, eligiendo al nuevo emperador, aun sin Bula de Oro ni legitimidad alguna a sus espaldas. Si fuera una ficción nos hubiera parecido hace no tantos años demasiado burda.

7. Conclusiones, visión de futuro y manifiesto urgente en defensa del Estado

Visto lo anterior y ante una realidad tan compleja como acuciante, resulta necesario extraer ciertas conclusiones, de entre las que iniciaremos por evitar la resignificación de los términos y la disputa de sus significados. La oclocracia es una forma degenerada y excluyente de democracia fuera de sus fronteras materiales y situada por encima del Estado de derecho y las

(3) Véase Apple respecto al Reino Unido o Francia, otrora hegemones imperiales hoy en retirada.

(4) Emblemáticos son los ejemplos de Elon Musk como administrador —*de facto*— del Departamento de Eficiencia Gubernamental de EE. UU. o de millonarios como Mauricio Macri, Sebastián Piñeira o el propio Donald Trump, en Argentina, Chile y EE. UU. respectivamente.

garantías de las minorías. Por ello mismo, conviene tener presente que no es el carácter mayoritario la seña de identidad genuinamente característica de la democracia, sino, de hecho, su naturaleza contramayoritaria. Una caracterización no numérica (mayoritaria) sino cualitativa; un conjunto de principios, valores y derechos inalienables a preservar.

No cabe democracia sin Estado de derecho, igual que no cabe *libertas* sin *ius*, igual que no cabe democracia sin ley; ni voluntad popular sin supremacía constitucional, legitimidad última para poder hablar en pulcritud de gobierno democrático. No nos hallamos solo frente una controversia política, nos encontramos ante un problema jurídico de primer orden, que interpela a todos sus operadores, ante el combate declarado al principio fundamental atribuido a Adams de una *res publica de leyes y no de hombres*.

La libertad sin un ejercicio institucionalizado y dotado de límites y garantías no es sino el amparo del privilegio y la primacía de la ley del más fuerte. En ese sentido, la oclocracia digital viene a impugnar los clásicos contrapesos liberales de canalización institucional, apelando a un permanente —y retórico— recurso al pueblo como patente de corso en una visión todista y monista del poder *leviatánico* sin atención a limitación o control alguno. He ahí por qué la universalización sin contrapesos es la mera extensión de la barbarie y el despotismo.

Por todo ello, resulta imperioso fortalecer los mecanismos de control al poder, despertar al Estado de su largo letargo, poner pie en pared ante el avance imparable tecnofeudal, cesarista y oligárquico computacional de las grandes corporaciones digitales y restaurar el papel central del Estado como actor político y jurídico garante no solo de la paz civil, sino de los derechos fundamentales conquistados con sangre y consagrados en nuestras constituciones. En definitiva, recuperar una concepción del Estado como escudo del eslabón más vulnerado y como ingeniero social trasformador más cualificado para un incierto siglo XXI.

A tal fin, urge abordar la regulación real, efectiva, valiente y urgente de la IA, garantizar la transparencia en el funcionamiento del algoritmo y sus sesgos, a efectos de trazabilidad, explicabilidad y repetibilidad como condición de posibilidad de la debida tutela judicial efectiva para el respeto a los derechos fundamentales, hoy masivamente amenazados por el potencial uso de la IA, como control y poder limitador de esta nueva forma de poder.

Con ese objetivo, resulta ineludible cuidar y tejer nuevas alianzas regionales y trasatlánticas con visión de porvenir, particularmente en el orden hispanoamericano y europeo a fin de poder dar ajustada respuesta a los citados desafíos globales, para lo que se requiere de una fuerza en el mundo que ningún Estado-Nación por sí solo se encuentra en condiciones de ofrecer si no mediante la culminación del proceso de configuración federal de

unos Estados Unidos de Europa, que llegará, y una auténtica cooperación política hispanoamericana como alternativa garantista y horizonte esperanzador ante el pulso de los tiempos.

Finalmente, hemos de avanzar hacia una genuina ciudadanía digital, eficaz y canalizada institucionalmente ante una realidad ya digital pendiente de arbitrar. Apelar al espíritu del mejor constitucionalismo primigenio para (re)erigir como sujeto político y jurídico una nueva noción universalista de "Demos" digital: un revigorizado "*We the people* virtual" que traslade derechos y deberes, facultades y limitaciones jurídicas a la realidad material y cibernética.

Movilizar, en conclusión, al derecho en toda su plenitud y con todo su potencial para el cometido histórico de su supervivencia como norma ordenadora de la sociedad humana. Oponer a las garantías constitucionales el grito encolerizado de la muchedumbre clamando por salvar a Barrabás. Dotarnos, en definitiva, de una auténtica y demasiado tiempo postergada "Constitución del algoritmo" (Balaguer, 2022); de auténtico *check and balance* digital como nuevo paradigma. Asumir la prometeica tarea de someter a estos poderes tecno-oligárquicos al *Rule of Law* y al Estado constitucional para grabar en el frontispicio que haya de presentar en el futuro el estudio de nuestro actual momento jurídico la exitosa empresa de garantía y legado de un Gobierno de leyes y no de algoritmos.

Referencias bibliográficas

Agamben, G. (2006). *La comunidad que viene*. Pre-Textos.

Aristóteles (2018). *Política*. Editorial Austral.

Balaguer Callejón, F. (2022). *La constitución del algoritmo*. Fundación Manuel Giménez Abad de Estudios Parlamentarios y del Estado Autonómico.

Beiguelman, G. (2021). *Políticas da imagen: vigilancia e resistẽncẽa na dadosfera*. Ed Ubu.

Betzu, M., Coinu, G. & Demuro, D. (2019). Gobernanza de los macrodatos y democracia representativa. *Revista de Derecho Político*, 106, 253-264. https://doi.org/10.5944/rdp.106.2019.26156.

Bobbio, N. (1986). *El futuro de la democracia*. Fondo de Cultura Económica.

Bodin, J. (1577). *Les six libres de la République*. Jacques du Puys Ed.

Cervelli, I. (1972). Cesarismo e Cavourismo. *La cultura*, X, 337-391.

Cicerón, M. T. (1991). *Sobre la República*. Editorial Gredos.

Durand, C. (2020). *Tecnoféodalisme. Critique de l´economie numerique*. Découverte-Zones.

Ferreira, O. (2019). El poder de la multitud. Horizonte de la democracia. Ed. Eska.

Ferrero García, E. (2024). Del pacto social al proyecto de acuerdo de país frente a los discursos de odio contra grupos en situaciones de vulnerabilidad. En Figueruelo Burrieza, A. (Dir.), *Desinformación, odio y polarización II* (181-194). Ed. Aranzadi.

Ferrero García, E. (2025). Inteligencia artificial y sus riesgos para los derechos fundamentales desde la perspectiva de género: una visión constitucional. En VV.AA., *Humanismo digital y perspectiva de género* (181-194). Ed. Colex - Instituto de las Mujeres.

García Pelayo, M. (2004). *Federico II de Suabia y el nacimiento del Estado moderno*. Fundación Manuel García Pelayo.

Gendler, M. A. (2018). Internet, algoritmos y democracia: ¿Del sueño a la pesadilla? *Nueva Sociedad,* 294, 37-47.

Greenwood, D. J. H. (2017). Neofeudalism: The Surprising Foundations of Corporate Constitutional Rights. *University of Illinois Law Review*, 163 (2), 163-221. Disponible en: https://ssrn.com/abstract=2963887.

Habermas, J. (1991). *The Structural Transformation of the Public Sphere*. The MIT Press.

Juvenal (1996). *Sátiras*. Ed. Consejo Superior de Investigaciones Científicas.

Laclau, E. (2004). *La razón populista*. Fondo de Cultura Económica.

Mangoni, L. (1976). Cesarismo, bonapartismo, fascismo. *Studi storici*, 3, 41-61.

Neumann, F. (1973). *Lo stato democrático e lo stato autoritario*. Ed. II Mulino.

Nieto Rojas J. H. (2020). Desinformación, redes sociales y riesgos de seguridad. *Revista de Investigación CUGC*, 7, 295-349.

Padilla Gálvez, J. (2017). Democracy in Times of Ochlocracy. *Synthesis philosophica*, 32 (1), 167-178. https://doi.org/10.21464/sp32112.

Platón (2020). *La República*. Editorial Gredos.

Polibio (1981). *Historias* II. Editorial Gredos.

Rousseau, J. J. (1812). *El Contrato Social, ó principios del Derecho Político*. Ferrer de Orga.

Rubio Núñez, R. (2018). Los efectos de la posverdad en la democracia. *Revista de Derecho Político*, 103(2), 191–228. https://doi.org/10.5944/rdp.103.2018.23201.

Rupeika-Apoga, R. (2025). The Future of Money: Central Bank Digital. Currencies, Cryptocurrencies and Stablecoins. *Journal of Risk Financial Management,* 18(9), 469. https://doi.org/10.3390/jrfm18090469.

Sartori, G. (2005). *Parties and Party Systems: A framework for análisis*. Cambridge University Press.

Schawb, K. (2016) *The Fourth Industrial Revolution*. Editorial Debate.

Schmitt, C. (2002). *El concepto de lo político*. Buenos Aires: Ed. Struhart & Cia.

Supiot, A. (2013). The Public-Private Relation in the Context of Today's Refeudalization. *International Journal of Constitutional Law*, 11(1), 129-145. https://doi.org/10.1093/icon/mos050.

Tocqueville, A. (1911). *La Democracia en América* I. Ed. Daniel Jorro.

Treitchke, E. (1918), *Politica*, II. Ed. Gius. Laterza & Figli.

Tucídides (2012). *La Guerra del Peloponeso*. Editorial Gredos.

Varoufakis, Y. (2024). *Tecnofeudalismo. El sigiloso sucesor del capitalismo*. Ed. Deusto.

Virio, P. (2016). *La Administración del miedo*. Ed. Pasos Perdidos.

Winner, L. (1991). A Victory for Computer Populism, *Technology Review*, 94(4), 81-99. https://doi.org/10.1177/1050651996010001004.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs

Reflexões tecnopolíticas a respeito da inteligência artificial e o direito

Por Vinícius Pedro Teló (*)

Resumo: O artigo analisa criticamente a emergência das inteligências artificiais (IAs) como um projeto político e científico, situado no cruzamento das tecnologias de vigilâncias e as formas contemporâneas de governo da população. Partindo das contribuições principalmente de Michel Foucault sobre disciplina, biopolítica e governamentalidade, bem como das formulações de Gilles Deleuze sobre sociedade de controle, busca-se compreender de que forma as IAs se inserem nesse deslocamento epistemológico, indagando se esse movimento reconfigura os discursos sobre neutralidade tecnológica, segurança e risco. O método é genealógico-crítico, com técnica documental de revisão bibliográfica, que busca aproximações possíveis ao objeto e à hipótese levantada, sem a pretensão de resolver ou esgotar o tema. Conclui-se que as IAs não são uma tecnologia neutra, mas um dispositivo de poder que intensifica as lógicas de vigilância e controle social próprias da sociedade de informação, convertendo os indivíduos em "corpos-sinais" permanentemente rastreáveis. Ao mesmo tempo em que legitima-se pelo discurso de eficiência e segurança, as IAs opera como mecanismo de gestão tecnopolítica que amplia riscos, opera de maniera opaca, é alimentada por vieses e reforça desigualdades.

Palavras-chave: Inteligência Artificial; sociedade de controle; tecnopolítica; opacidade; sociedade da informação.

Reflexiones tecnopolíticas sobre la inteligencia artificial y el derecho

Resumen: El artículo analiza críticamente la emergencia de las inteligencias artificiales (IAs) como un proyecto político y científico, situado en la intersección de las tecnologías de vigilancia y las formas contemporáneas de gobierno de la población. Partiendo principalmente de las contribuciones de Michel Foucault sobre disciplina, biopolítica y gubernamentalidad, así como de las formulaciones de Gilles Deleuze sobre la sociedad de con-

(*) Vinícius Pedro Teló. Mestrando em Ciências Criminais na Pontifícia Universidade Católica do Rio Grande do Sul – PUC\RS. Bolsista PROPESQ PUCRS. E-mail: telo.vinicius@gmail.com. Currículo lattes: https://lattes.cnpq.br/9106109975691575.

trol, se busca comprender cómo las IAs se insertan en este desplazamiento epistemológico, cuestionando si este movimiento reconfigura los discursos sobre neutralidad tecnológica, seguridad y riesgo. El método es genealógico-crítico, con técnica documental de revisión bibliográfica, que pretende realizar aproximaciones posibles al objeto y a la hipótesis planteada, sin la pretensión de resolver o agotar el tema. Se concluye que las IAs no son una tecnología neutra, sino un dispositivo de poder que intensifica las lógicas de vigilancia y control social propias de la sociedad de la información, convirtiendo a los individuos en "cuerpos-señales" permanentemente rastreables. Al mismo tiempo que se legitima mediante discursos de eficiencia y seguridad, las IAs operan como un mecanismo de gestión tecnopolítica que amplía los riesgos de opacidad, sesgos y retroalimentación de desigualdades.

Palabras clave: Inteligencia artificial; sociedad del control; tecnopolítica; opacidad; sociedad de la información.

Technopolitical reflections on artificial intelligence and law

Abstract: The article critically analyzes the emergence of artificial intelligences (AIs) as a political and scientific project, situated at the intersection of surveillance technologies and contemporary forms of governing populations. Drawing primarily on Michel Foucault's contributions on discipline, biopolitics, and governmentality, as well as Gilles Deleuze's formulations on the society of control, the study seeks to understand how AIs are embedded in this epistemological shift, questioning whether this movement reconfigures discourses on technological neutrality, security, and risk. The method is genealogical-critical, employing documentary bibliographic review as a technique to explore possible approximations to the object and the proposed hypothesis, without the intention of resolving or exhausting the theme. The article concludes that AIs are not a neutral technology, but rather a dispositif of power that intensifies the logics of surveillance and social control characteristic of the information society, converting individuals into permanently traceable "signal-bodies". While legitimized through discourses of efficiency and security, AI operates as a technopolitical management mechanism that amplifies risks of opacity, bias, and the reinforcement of inequalities.

Key words: Artificial Intelligence; society of control; tecnopolitc; opacity; information society.

1. INTRODUÇÃO

A emergência das inteligências artificiais (IAs) como tema incontornável do presente não pode ser lida apenas como inovação tecnológica ou acidente da evolução científica. Sua inserção em múltiplas dimensões da vida social revela uma racionalidade mais profunda: a atualização de pro-

jetos políticos e científicos de governo da população. A IA, portanto, não é neutra, mas figura como peça de uma engrenagem que articula saberes, discursos e práticas de vigilância, transformando dados em matéria-prima de gestão social.

No campo jurídico, esse debate adquire contornos ainda mais sensíveis. A aplicação de algoritmos em tribunais, investigações criminais, reconhecimento facial e cruzamento de perfis genéticos mostra que o direito não é apenas espectador, mas território de experimentação e legitimação dessas tecnologias. Se, de um lado, os discursos de eficiência e segurança sustentam sua adoção, de outro, a opacidade, os vieses e a retroalimentação de desigualdades evidenciam riscos estruturais de um modelo que converte a sociedade em objeto calculável e governável.

É nesse cenário que se insere o presente trabalho, que busca compreender a emergência das IAs no contexto das tecnologias de vigilância e governo da população, e de que modo os discursos de neutralidade, eficiência e segurança legitimam tais práticas?

A investigação parte de uma abordagem genealógico-crítica, mobilizando especialmente as contribuições de Michel Foucault sobre biopoder e governamentalidade, as formulações de Gilles Deleuze sobre a sociedade de controle, bem como as análises recentes de Pablo Rodríguez e Byung-Chul Han sobre a sociedade da informação. A metodologia utilizada é qualitativa, com técnica documental de revisão bibliográfica, não para esgotar o tema, mas para produzir aproximações críticas que permitam situar as IAs como dispositivo de poder.

O artigo está organizado em dois capítulos, além das considerações finais. No primeiro, são exploradas os contextos históricos e as bases teóricas que permitem a aproximação entre inteligência artificial, Direito e vigilância. No segundo, são discutidos os discursos que buscam legitimar tais práticas de vigilâncias e os riscos sociais envolvidos.

2. A IA, o direito e o que a vigilância tem a ver com isso

A inserção das inteligências artificiais nas mais variadas atividades humanas pode parecer para alguns uma novidade ou mesmo um acidente fortuito da "evolução tecnológica". O desenvolvimento de uma inteligência artificial, de uma máquina pretensamente dotada de capacidade humana (ou sobre-humana) de compreensão, interpretação, comunicação, etc., está inserido em um projeto político e científico de governo da sociedade. Esse projeto emergiu da ferida aberta da modernidade e se propôs como a solução para a crise existencial e governamental resultante das duas grandes guerras e das crises humanitárias geradas pelo capitalismo. Esse solucionismo tecnológico, o qual foi pintado com um verniz discursivo poderoso

– dotar a máquina da capacidade humana, especificamente, do cérebro humano, órgão responsável por manter o corpo funcionando, desde as funções conscientes às automáticas e inconscientes –, é um projeto da modernidade (Comité Invisible, 2014).

A despeito dos discursos que circundam as IAs, não há nada de sobrenatural ou mesmo de natural (no sentido humano) em sua capacidade de processar informações. Os modelos atuais de IAs nada mais são do que algoritmos, ou seja, uma sequência de instruções derivadas do processo de programação que descrevem/instruem detalhadamente como deve ser feita determinada tarefa ou alcançado determinado objetivo por um programa, somada à capacidade de "aprendizagem", o que se refere à sua alimentação e retroalimentação com grandes bancos de dados, o que busca otimizar seus resultados (Pires e Silva, 2017).

E mesmo considerando o potencial das IAs de analisar grandes quantidades de dados e fornecer respostas à *prompts* (como são chamadas as perguntas do usuário que direcionam sua atividade), não há qualquer originalidade ou mesmo compreensão daquilo que fazem, uma vez que sua atividade pode ser resumida a uma operação matemática de probabilidade (Amaral e Salles, 2023) [1].

Para além de suas capacidades e limitações, importa perceber a exata encruzilhada em que as IAs passa a ser um projeto de governo e científico. Para isso, reputamos importante recorrer aos estudos de Michel Foucault sobre biopoder e biopolítica, bem como os estudos de Gilles Deleuze sobre sociedade de controle. Isso porque foi somente após o poder não ter como principal foco o indivíduo, mas também a população, com a necessidade de conhecê-la para melhor governa-la – e, assim, melhor extrair sua potencialidade, manejar e influenciar os fluxos econômicos, etc –, que as IAs surgem como um projeto político. O foco não será fazer a história do poder e das tecnologias de governo, mas sim estabelecer a base teórica necessária para avançarmos em nossa análise.

Inicialmente, devemos compreender o poder a partir de sua produtividade, e não a partir de um modelo jurídico, formal e proibitivo, como feito pela ciência política de maneira hegemônica até o século XX. Foucault propõe que não há "O" poder, mas que ele se exerce e é exercido em relações que se espalham por toda a sociedade, devendo ser compreendido como

(1) Segundo os autores: "E mesmo se essa questão tivesse sido superada [da originalidade], há um terceiro entrave que os pesquisadores da inteligência artificial não resolveram: as máquinas não são conscientes da natureza e do significado das tarefas que executam, embora efetivamente sejam hábeis para seguir aos comandos e agir de acordo com as instruções. Os sentidos de suas ações são dados pelos humanos que a gerenciam, de modo que as máquinas inteligentes dependem do intelecto humano para prenunciar as regras de manipulação e interpretação dos símbolos" (Amaral e Salles, 2023, p. 201).

uma tecnologia, operando de maneiras diversas, a partir de diferentes técnicas e procedimentos, a todo momento criado e aperfeiçoado, "no ateliê, no exército, em uma propriedade do tipo escravagista ou em uma propriedade onde há relações servis" (2012, p.172). O poder não está tão somente na interdição dos comportamentos proibidos, mas na produção: de corpos, comportamentos, de desejo, de subjetividades, etc. Segundo Foucault: "A função primitiva, essencial e permanente desses poderes locais e regionais é, na realidade, serem produtores de uma eficiência, de uma aptidão, produtores de um produto" (2012, p.173).

Enquanto produtor de um produto, o objeto do poder na sociedade disciplinar é o corpo dócil, mais facilmente governável e mais lucrativo. Isso era feito por meio dos grandes confinamentos: a escola, a fábrica, a caserna e o hospital (Foucault, 2012). Foucault identifica na arquitetura do panóptico de Jeremy Bentham a lógica, ou episteme, por trás da vigilância da modernidade, um novo princípio da construção. O panóptico, a prisão modelo, é construída em formato circular, com as celas dispostas ao interior do círculo e uma torre central que não permite que se veja seu interior. Assim, o prisioneiro não sabe quando está sendo vigiado. Pouco importa a presença ou não de um guarda em seu interior, uma vez que a própria estrutura causa a sensação de vigilância o que, por si só, produz parte do resultado almejado. A lógica de acumular informações sobre a população para melhor intervir na sociedade também já está aí presente (Rodríguez, 2019).

Surgem então duas técnicas políticas, ou tecnologias do poder, entre os séculos XVII e XVIII, de um lado a disciplina, ou anatomopolítica – cujo objeto é alcançar os átomos do corpo social, os indivíduos, "como vigiar alguém, como controlar a sua conduta, seu comportamento, suas atitudes, como intensificar a sua performance, multiplicar suas capacidades, como colocá-lo em seu lugar" (Foucault, 2012, p.177) –, e de outro, técnicas que visam a população, ou biopolítica – seres vivos, regidos por processos, por leis biológicas, com taxa de natalidade, taxa de mortalidade, curva de idade, morbidez, estado de saúde, potencialidades, etc (Idem, 2012).

Assim, para melhor intervir nesse corpo social, na população, cada vez mais passa a ser necessário conhecê-la. Essas técnicas, como se verá, são importantes para compreender a emergência das IAs, que surgem a partir da busca pela informação total sobre a sociedade, com o projeto de tornar possível ler, interpretar e sistematizar imensas quantidades de dados sobre a população.

Sobre esses processos de controle e acúmulo de informações, Jonathan Finn identifica alguns processos que intensificaram a busca por meios mais eficientes de interpretar grandes quantidades de informações. Falando especificamente sobre investigações criminais e identificação de pessoas, o autor aborda tecnologias policiais, como foi o caso, cronologicamente, da

fotografia, da datiloscopia e da descoberta do ácido desoxirribonucleico (DNA). Acontece que, antes do advento dos computadores e da *internet*, essas informações a respeito de indivíduos precisavam de um arquivo, um espaço físico de armazenamento, com métodos de organização que buscavam tornar possível acessar e navegar essas informações. O autor conta que, em 1877, a "galeria dos bandidos" de Berlim cresceu tanto que precisou ser subdividida em 10 grupos baseados no crime e exigiu três antecedentes para inclusão. À medida que o número de registros crescia, a utilidade do arquivo diminuía (2009).

Conhecer a população para melhor governá-la. Acumular cada vez mais informações sobre essa população, mas sem perder a utilidade dessas informações. Emergia cada vez mais a necessidade de interpretar essa grande quantidade de informações. Mas isso, por si só, não explica como chegamos ao momento em que uma infinidade de informações se torna possível a respeito dessa população, ainda que desde então já existissem problemas operacionais relacionados à interpretação e manuseio de informações impressas no papel.

Um processo importante para entendermos como chegamos até o momento atual foi descrito por Pablo Manolo Rodríguez, a partir de Foucault e Deleuze, como a crise dos interiores, das instituições de confinamento (prisão, fábrica, hospital, fábrica, escola e família) e também do interior subjetivo, que caracteriza o panorama biopolítico atual, que está localizado na dissolução dos limites entre interior e exterior. A crise dos interiores tem ligação com a baixa eficácia relativa da ligação entre vigilância e confinamento, característica das sociedades disciplinares. Na sociedade de controle nunca se termina nada: a empresa, a formação, o serviço são estados metaestáveis e coexistentes de uma modulação, como um deformador universal (2019).

Era por meio do confinamento e da vigilância decorrente que o Estado poderia conhecer e disciplinar os indivíduos e populações. Se escancarava a insuficiência dos interiores como tecnologia de governo.

Avançando um pouco, lembrando que nossa tarefa não é fazer a história das tecnologias de vigilância, mas sim compreender o contexto em que as IAs emergem como um projeto político, entre o final do século XX e o início do XXI, em mais ou menos três décadas, as câmeras de segurança se expandiram como pragas por todos os espaços: trabalho, espaços públicos, parques, praças, ruas e, finalmente, os céus, com os drones. Mas isso sequer esgota a potência de visualização desse fenômeno, pois as câmeras dos dispositivos móveis, telefones-celulares, fotografam a todo modo, em qualquer lugar, público ou privado, por livre e espontânea vontade dos usuários. Os dispositivos biométricos se instalaram a relativamente poucos anos, mas agora se estendem por grandes programas de reconhecimento facial, que

sequer estão limitados a sistemas de vigilância privados ou estatais (Rodríguez, 2019).

Surgem os cartões inteligentes (em Deleuze, *smart cards*), que carregam cada vez mais dados dos indivíduos, e já sem relação com uma finalidade específica – uma ação em determinado lugar. Empresas privadas, como *Google* e *Facebook*, acumulam esses dados pessoais (e até os vendem a agências estatais e privadas de segurança), demonstram a emergência de um novo espaço de vigilância, por meio da *Big Data*, que alude aos sistemas complexos de processamento e estabelecimento de padrões da enorme massa de dados que emergem não somente dos sistemas de vigilância anunciados como tais, mas também das redes sociais. Os sinais começaram a desligar-se dos corpos, a grande massa de dados se desterritorializa dos corpos e dos confinamentos e nem por isso se reterritorializa em outro espaço (Rodríguez, 2019).

Estamos diante de um processo de informatização da sociedade, que consiste na possibilidade de converter em dados homologáveis e em uma base tecnológica comum praticamente toda a vida comum de quase todos os indivíduos, desde o transpasse de fronteiras, a locomoção, a compra de produtos à expressão de opiniões (Rodríguez, 2019).

Isso permite identificar uma série de continuidades e rupturas com relação à episteme anterior à pós-modernidade, em razão da chamada sociedade da informação. Rodríguez aponta que em um primeiro momento, pode parecer apenas a complementação do processo moderno de burocratização da vigilância estatal do século XIX, mas há aí uma mudança também qualitativa no feito de se criar uma base de dados unificada e processável por qualquer atividade pública ou privada. Porém, a maior diferença se situa no fato de que a vigilância passou a ser aceita por todos, a partir do discurso de que tornaria a vida cotidiana mais prática, mais cômoda, mais segura, etc. O ápice da sociedade da informação é que a maior quantidade de tarefas seja realizada de maneira comum e seja transferida a um suporte digital (Rodríguez, 2019).

Além da difusão da vigilância para os espaços abertos, os dispositivos móveis que permitiram o mais firme passo em direção à sociedade de controle, pois eles permitem captar tudo com seus microfones e câmeras infinitesimais, sem necessidade de um ponto fixo de observação. Assim se fez a difusão da vigilância do confinamento para todos e cada um.

É o regime da informação, forma de governar por meio de dados, que produz sujeitos que participam de sua própria dominação. A vigilância por meio de dados é produzida a partir e feita pelos próprios indivíduos, enquanto performam nas redes abertas – celulares, computadores, videoga-

mes etc – permitindo a produção de saberes a respeito dos indivíduos e da sociedade pelas interações (Han, 2022).

Comparando em termos de subproduto do exercício do poder na sociedade disciplinar e na sociedade de controle, no primeiro produz-se o corpo-máquina, enquanto no segundo o corpo-sinal. "A vida cotidiana já não consegue não deixar rastros" [2] (Rodríguez, 2019, p.351).

Aí está toda a potencialidade das IAs para a sociedade de controle, enquanto projeto político e científico, torna possível a leitura, interpretação e manipulação do que antes era impossível. Claro, esses dados estão localizados em locais físicos, em imensos galpões espalhados pelo mundo (Harada, 2016), cuja função é armazenar dados quase infinitos – sites, contas de usuários, *chats* de conversas, *e-mails*, calendários, agendas, preferências, pesquisas de usuários, contas de jogos etc tudo deixa rastros no digital e fica armazenado. Mas não é necessário o impossível, que indivíduos vasculhem esse universo de informações, mas sim que um *prompt* inicie o trabalho do algoritmo que fará a varredura por essas informações. A realidade, virtualizada, se torna ainda mais vigiável.

Se a genealogia revela que as IAs emergem como continuidade/ruptura na sociedade de controle e no regime da informação, resta agora investigar quais os discursos que legitimam sua neutralidade, eficiência e segurança, ao mesmo tempo em que produzem opacidades e riscos.

3. Discursos, opacidade e riscos

Longe de ser uma tecnologia neutra ou sem riscos, as IAs representam apenas uma face da tecnopolítica criminal, forma de gestão da sociedade com a inserção de instrumentos tecnológicos que aprimoram o controle (Amaral e Dias, 2024). Questões como os interesses econômicos de empresas, a falibilidade dos seus resultados e a retroalimentação de preconceitos e vieses são blindadas a partir dos discursos tecno-científicos que buscam a legitimação das IAs e da vigilância, em especial a partir do dispositivo securitário.

Diversos estudos já apontaram inúmeros riscos inerentes às IAs: os vieses humanos e algorítmicos, alimentados a partir de seus desenvolvedores e dados utilizados no processo de *machine learning* (Gonzalez, Maldonado E. e Maldonado K, 2024); a opacidade inerente à construção do algoritmo, como ele aprende ou toma decisões, que leva pesquisadores do tema a os classificarem como caixas-pretas (Pasquale, 2015); a falta de precisão ou compromisso com critérios de verdade na formulação de respostas (Maleki,

(2) Tradução livre de: "[...] la vida cotidiana ya no podrá no dejar rastros".

Padmanabhan e Dutta, 2024); e o solucionismo tecnológico, como se a tecnologia, por si só ou mesmo com inferência humana, seria capaz de resolver problemas sociais complexos (Morozov, 2018).

No direito, as IAs já são aplicadas por tribunais, para interpretação de processos e auxílio na tomada de decisões, nas investigações criminas, por meio do Banco Nacional de Perfis Genéticos, no cruzamento de perfis genéticos, por meio das tecnologias biométricas de reconhecimento facial, e em diversas outras aplicações, cada uma com suas potencialidades e riscos, os quais, pela limitação de tamanho do trabalho não poderão ser enfrentados.

Han investigou esses discursos, que chamou de dataístas, para os quais o *Big Data*, as informações geradas por usuários e armazenadas por empresas privadas e públicas no ambiente digital, permitem "a possibilidade de observar a sociedade em toda sua complexidade, pelos milhares de conexões de trocas interpessoais" (Pentland, 2014, p.11, apud, Han, 2022, p. 68). O autor citado associa ainda os dados e sua análise por inteligências artificiais a um "olho divino, uma vista católica" (Idem), capaz de compreender a sociedade e solucionar problemas.

Essa racionalidade dataísta pressupõe que a partir da mineração dos dados a sociedade passa a ser um sistema social calculável, apto a ser lido e compreendido pelas inteligências artificiais, as quais são capazes de decisões inteligentes, mais racionais do que os humanos com sua capacidade limitada de processar informações (Han, 2022). Vemos a proposição de que as tecnologias de dados seriam infalíveis, neutras, de simples cálculo matemático, inclusive superiores à capacidade humana de compreender e processar informações.

Empregar essa gramática busca justamente blindar essas práticas de controle de qualquer crítica, inclusive serve ao propósito de afastar reflexões econômicas ou políticas, sempre presentes em qualquer projeto de gestão da vida. Quem poderia ser contra a informação ou à tecnologia? (Dias e Amaral, 2024).

Propomos que para a perspectiva dataísta também o discurso jurídico "não é outra coisa do que uma forma lenta e ineficiente de processamento de informação. As reivindicações de validade pelos participantes do discurso têm como fundamento um processamento insuficiente de informações" (Han, 2022, p.67).

Nasce a ideia de saber total sobre a sociedade, ideal dataísta. Pensa-se possível a mineração da realidade (*Reality mining*), a partir do qual serão equipados "sociômetros" às pessoas, que protocolam minuciosamente os comportamentos e a linguagem do corpo, gerando uma quantidade de dados analisáveis pela *Big Data* e pelas inteligências artificiais. "O Reality mi-

ning com sensores digitais torna toda a sociedade calculável e governável" (Han, 2022, p.76).

Longe da pretensão de desvelar as IAs como uma forma de controle autêntica e desvinculada de outras formas de controle social, a colocamos exatamente como um desdobramento de um certo modo de governo da vida. Uma face da tecnopolítica criminal.

Como apontado por Amaral e Dias, não há superação do poder disciplinar na sociedade de controle, mas sobreposição das práticas [3]. A título de exemplo, o cárcere, atualizado para as necessidades da sociedade de controle, gera cifras (no caso, perfis genéticos) sobre a população por ele alcançada, dados capazes de produzir previsibilidade de todos e qualquer indivíduo, grupo ou comportamento, o qual poderá ser "interpretado" pelas IAs.

E, se por um lado as IAs são protegidas discursivamente, por outro, a complexidade e falta de transparência (às vezes de compreensão dos próprios desenvolvedores) impede – por *design* e não por acidente – a compreensão pública e a contestação das suas aplicações na vida social. Os riscos já apontados – vieses, retroalimentação de preconceitos, falhas, solucionismo– não são meras falhas corrigíveis, mas consequências estruturais de um modelo que transforma a sociedade em objeto calculável e previsível.

Conclui-se, assim, que os contornos dessa aparente face da tecnopolítica criminal que, agora, acirra processos de vigilância a partir da digitalização da vida, acúmulo de dados e desenvolvimento de IAs capazes de navegar – de maneira mais ou menos controlável e verificável – essas nuvens de dados, emergem dentro de um projeto político e científico próprio das sociedades de controle.

Enfatizamos que o tema merece ainda maior aprofundamento, compreensão e diferentes aproximações, a partir de diferentes bases teóricas e ramos do conhecimento, mas desde já se escancarou a urgência de novas análises críticas, em razão do impacto direto dessas práticas na vida cotidiana e nos direitos fundamentais de todas as pessoas.

4. CONSIDERAÇÕES FINAIS

Este trabalho propôs que as inteligências artificiais são um projeto político e científico que emergiu nas sociedades de controle, aprimorando processos de informatização da vida e vigilância. Não são um artefato neutro, mas dispositivos de poder inscritos em uma racionalidade de governo pró-

(3) Amaral, Augusto Jobim Do; DIAS, Felipe Da Veiga. *Tecnopolítica Criminal*. 1. Ed. São Paulo: Tirant Lo Blanch, 2024.

pria das sociedades de controle. A genealogia buscou mostrar a transição do corpo-máquina ao corpo-sinal, em um regime de informação que converte a vida social em dados calculáveis. No plano discursivo, a promessa de neutralidade, eficiência e segurança opera como blindagem tecnopolítica, deslocando o debate de suas determinações políticas e econômicas para ajustes técnicos, ao mesmo tempo em que opacidades, vieses e retroalimentação de desigualdades se tornam efeitos estruturais.

Esse estado da arte das IAs, impõe uma crítica das condições de possibilidade das IAs como forma de governo, com ênfase na proteção de direitos fundamentais, na exigência de transparência e auditabilidade, na contestabilidade pública de decisões automatizadas e no controle democrático de infraestruturas e bases de dados.

Referências

Amaral, Augusto Jobim do; Dias, Felipe Da Veiga. *Tecnopolítica Criminal*. 1. Ed. São Paulo: Tirant Lo Blanch, 2024.

Comité Invisible. *À nos amis*. Paris: La Fabrique, 2014.

Deleuze, Gilles. *Conversações*, 1972/1990; tradução de Peter Pál Pelbart. São Paulo: Editora 34, 2017.

Finn, Jonathan Mathew. *Capturing the criminal image: From mug shot to surveillance society*. U of Minnesota Press, 2009.

Foucault, Michel (2012). *As malhas do poder*. In: MOTTA, Manoel Barros da (org.). Ditos e escritos VIII. Rio de Janeiro: Forense Universitária.

Gonzalez-Argote J, Maldonado E, Maldonado K. Algorithmic Bias and Data Justice: ethical challenges in Artificial Intelligence Systems. EthAIca. 2025; 4:159. https://doi.org/10.56294/ai2025159

Harada, E. Y. (2016/05/05). Conheça a estrutura monstruosa utilizada pela Google para seus servidores: Que tal fazer um tour pelos data centers da Gigante das Buscas. Tecmundo: https://www.tecmundo.com.br/google/104430-conheca-estrutura-monstruosa-utilizada-google-servidores.htm

Han, Byung-Chul. *Infocracia: digitalização e a crise da democracia*. Trad.: Gabriel S. Philipson. Petrópolis: Vozes, 2022.

Maleki, N., Padmanabhan, B. and Dutta, K., "AI Hallucinations: A Misnomer Worth Clarifying," 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, Singapore, 2024, pp. 133-138, doi: 10.1109/CAI59869.2024.00033.

Morozo, Evgeny. *Big Tech: a ascenção dos dados e a morte da política*; traduzido pr Claudio Marcondes – São Paulo: Ubu Editora, 2018.

Motta, Manoel Barros da (org.). *Michel Foucault: Arqueologia das Ciências e História dos Sistemas de Pensamento.* Ditos e Escritos. Vol. II. Rio de Janeiro: Forense, 2008.

PASQUALE, Frank. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Massachusetts: Harvard University Press, 2015.

Pires, Thatiane Cristina Fontão; Silva, Rafael Peteffi da. A responsabilidade civil pelos atos autônomos da inteligência artificial: notas iniciais sobre a resolução do Parlamento Europeu. Revista Brasileira de Políticas Públicas. Vol. 7, n. 3, 2017.

Rodríguez, Pablo Manolo. *Las palabras en las cosas: saber, poder y subjetivación entre algoritmos y biomoléculas*. Buenos Aires: Cactus, 2019.

Thompson, Willlian C., *The Myth of Infallibility*. In: Krimsky, Shelson; Gruber, Jeremy. *Genetic explanations: sense and nonsense*. 1 ed. – Cambridge, Massachusetts, and London, England: Harvard Unity Press, 2013.

Acesso à Justiça reloaded: decodificando a matrix para refundar a jurisdição

Por Angela Araujo da Silveira Espindola (*)

Resumo: O texto propõe uma reflexão sobre a relação entre acesso à justiça e cibercultura, articulando ambas as noções por meio da metáfora da Matrix e sob uma perspectiva de Direito e Literatura. A partir de uma atitude fenomenológico-existencial inspirada em Husserl, a autora busca compreender, mais do que medir, o fenômeno do acesso à justiça no contexto da transformação tecnológica contemporânea. Revisa criticamente as reformas judiciais inspiradas pelo Projeto de Florença e pelas políticas neoliberais das últimas décadas, apontando que a modernização do Poder Judiciário brasileiro privilegiou a eficiência e a metrificação em detrimento de um acesso verdadeiramente qualitativo à justiça. Nesse contexto, a cibercultura é entendida não apenas como um conjunto de tecnologias, mas como um processo social que pode tanto ampliar a inteligência coletiva quanto reforçar mecanismos de controle e desigualdade. O ensaio adverte sobre o risco de consolidar um tecnoneoliberalismo processual, no qual os algoritmos substituem a sensibilidade e a deliberação democrática. Decodificar a Matrix significa libertar o Direito da tirania das métricas e da linguagem binária, promovendo uma refundação da jurisdição a partir de perspectivas culturais diversas, humanas e latino-americanas.

Palavras-chave: Acesso à justiça; cibercultura; tecnoneoliberalismo; direito e literatura; matrix.

Acceso a la Justicia recargado: decodificando la matrix para refundar la jurisdicción

Resumen: El texto propone una reflexión sobre la relación entre acceso a la justicia y cibercultura, articulando ambas nociones mediante la metá-

---

(*) Doutora e Mestre em Direito pela Unisinos. Professora Associada do Departamento de Direito da UFSM. Professora do PPGD/UFSM. Líder dos Grupos de Pesquisa "Phronesis: Jurisdição e Humanidades" (UFSM/CNPq). Presidente da Rede Brasileira de Direito e Literatura (2023-2024 e 2025-2026).

fora de la *matrix* y desde una perspectiva de derecho y literatura. A partir de una actitud fenomenológico-existencial inspirada en Husserl, la autora busca comprender —más que medir— el fenómeno del acceso a la justicia en el contexto de la transformación tecnológica contemporánea. Revisa críticamente las reformas judiciales inspiradas por el Proyecto de Florencia y por las políticas neoliberales de las últimas décadas, señalando que la modernización del Poder Judicial brasileño ha privilegiado la eficiencia y la metrificación en detrimento de un acceso verdaderamente cualitativo a la justicia. En este marco, la cibercultura es entendida no solo como un conjunto de tecnologías, sino como un proceso social que puede tanto ampliar la inteligencia colectiva como reforzar mecanismos de control y desigualdad. El ensayo advierte sobre el riesgo de consolidar un tecno-neoliberalismo processual, donde los algoritmos sustituyen la sensibilidad y la deliberación democrática. Decodificar la matrix, significa liberar al derecho de la tiranía de las métricas y del lenguaje binario, promoviendo una refundación de la jurisdicción desde perspectivas culturales diversas, humanas y latinoamericanas.

Palabras clave: Acesso à justiça; cibercultura; tecnoneoliberalismo; direito e literatura; matrix

Access to Justice reloaded: decoding the matrix to rebuild jurisdiction

Abstract: The text proposes a reflection on the relationship between access to justice and cyberculture, articulating both notions through the metaphor of the Matrix and from a Law and Literature perspective. Drawing on a phenomenological-existential attitude inspired by Husserl, the author seeks to understand—rather than measure—the phenomenon of access to justice within the context of contemporary technological transformation. It critically reviews the judicial reforms inspired by the Florence Project and by the neoliberal policies of recent decades, pointing out that the modernization of the Brazilian Judiciary has prioritized efficiency and quantification to the detriment of truly qualitative access to justice. In this framework, cyberculture is understood not merely as a set of technologies but as a social process that can both expand collective intelligence and reinforce mechanisms of control and inequality. The essay warns of the risk of consolidating a techno-neoliberal proceduralism, in which algorithms replace sensitivity and democratic deliberation. Decoding the Matrix, the author suggests, means freeing Law from the tyranny of metrics and binary language, promoting a re-foundation of jurisdiction from diverse, human, and Latin American cultural perspectives.

Keywords: access to justice; cyberculture; techno-neoliberalism; law and literature; matrix

1. Considerações iniciais: bem-vindo(a) ao deserto do real

O mote deste artigo em tom de ensaio são duas palavras chaves "acesso à justiça" e "cibercultura". Ambas são aqui tomadas como pontos de partida para a fixação do nosso campo semântico. Estas duas palavras-chaves, no nosso imaginário jurídico, vão nos reportar até duas obras referenciais: "Acesso à Justiça", de Mauro Cappelletti e Bryan Garth (1988) e "Cibercultura", de Pierre Lévy (2009). E estas palavras-chaves serão amalgamadas na metáfora da Matrix, a partir de uma perspectiva metodológica de "Direito e Literatura" (KARAM, 2017; TRINDADE; BERNSTS, 2017; TRINDADE; GUBERT, 2008).

Antes de mais nada, é importante situarmos brevemente o "movimento do acesso à justiça" ocidental a partir das reflexões trazidas pelo *Projeto de Florença* e pelas recomendações do Banco Mundial dispostas no *Documento Técnico 319: o setor judiciário na América Latina e no Caribe, elementos para reforma* (DAKOLIAS, 1996) e a partir dele. Ao fazer isso, estabelecemos como marco inicial as duas últimas décadas do século XX, no mundo Ocidental e como direção a terceira década do século XXI, já que estamos na segunda década deste século.

Feito isso, vamos problematizar o cenário das ondas reformistas do processo civil brasileiro e o advento do Código de Processo Civil, de 2015, que comemora seu quinquênio de vigência neste ano de 2021.

Essa contextualização considerará (e dará destaque) à virada tecnológica no processo (em sua dupla dimensão, qual seja: (a) a dimensão decisória e (b) a dimensão da gestão do processo, buscando compreender como a emergência de uma cibercultura impacta os caminhos não só da humanidade, da aprendizagem, da modernidade, mas também do "movimento do acesso à justiça".

Adotaremos aqui o tom de um ensaio acadêmico e não de um artigo científico propriamente dito. O objetivo central deste ensaio, centrado na preocupação com o acesso à justiça, é apontar alguns dos desafios trazidos pelo conjunto de tendências denominado de cibercultura no âmbito da Jurisdição. Nossa pergunta norteadora (e experimental, haja vista o tom ensaístico escolhido) será: *diante dos desafios da cibercultura, como pensar o acesso à justiça, quantitativo e qualitativo, para as próximas duas décadas?*

Observe que estamos assumindo que a cibercultura demanda que sejam retomadas e revisadas as compreensões tradicionais de acesso à justiça, assim como também estamos assumindo que as reformas do final do século XX e início do século XXI concentraram-se muito mais em pontos de vista quantitativo (métricas), não raro em detrimento de pontos de vista

qualitativo (PEDRON, 2008), e que isso se operou não só *pelo* judiciário, mas também *no* judiciário e *para além* do poder judiciário [1]. Registre-se aqui que, por acesso à justiça qualitativo não se está partindo de uma perspectiva binária em que se analisa a dualidade de um mesmo fenômeno, investigando o acesso à justiça quantitativo *vs* qualitativo, mas antes partindo da perspectiva de consolidação do Estado Democrático de Direito e da efetivação de direitos, que passa pela também pela teoria da decisão. (PEDRON, 2013; TEIXEIRA, 2011)

Nosso esforço é para que possamos não só "medir" o acesso à justiça, mas sobretudo "compreendê-lo", "senti-lo". Para tanto, precisamos ir além das *métricas*, pensando em *harmonia, melodia e ritmo* e, quem sabe, em uma *poli-métrica*. Por fim, vale adiantar que os desafios ao acesso à justiça não são necessariamente desafios de um determinado tempo (a modernidade dos séculos XX ou XXI) ou espaço (ciberespaço)!

Talvez seja útil advertir que vamos adotar – em coerência com nossa condição de pesquisadora - uma atitude (metodológica) fenomenológica-existencial, pressupondo a unidade espaço-tempo-coexistência e, portanto, considerando que "toda consciência é consciência de algo", apoiando-nos na perspectiva de Husserl (2012, 2020). Pretendemos, portanto, tecer uma análise compreensiva e não explicativa dos fenômenos (LAPORTE; VOLPE, 2000, p. 52), tomando a realidade como uma construção do pensamento e, como tal, produzida pela linguagem. A metáfora da Matrix (bastante compatível com esta atitude metodológica) irá nos guiar neste ensaio.

2. MATRIX: A METÁFORA DO SISTEMA JURÍDICO E A DECODIFICAÇÃO DO MOVIMENTO DO ACESSO À JUSTIÇA

Matrix é uma trilogia norte-americana de ficção científica. "The Matrix", o primeiro filme, surgiu em 1999. As sequencias "The Matrix Reloaded" e "The Matrix: Revolutions" surgem logo depois, em 2003, com intervalo de seis meses entre uma e outra. Recentemente, quase 20 anos depois, estamos na expectativa de assistirmos a "The Matrix 4", que ainda está em sessão de teste e – segundo especula-se – receberá o nome "The Matrix: Resurrection". Ao tempo da conclusão deste ensaio (meados de 2021), a previsão é de que "Matrix 4" chegue aos cinemas em dezembro de 2021.

Seguindo a metáfora cinematográfica, a narrativa do filme se passa em um futuro distópico, representando uma realidade simulada criada por má-

(1) Essa preocupação com uma dimensão quantitativa em detrimento de uma perspectiva qualitativa é um sintoma de que a Crise da Jurisdição e até mesma a Crise do Direito, vem sendo enfrentadas como um problema de função e não como um problema também de estrutura (paradigmático). Nesse sentido, ver Espindola (2008) e Pedron (2008).

quinas sencientes para subjulgar a humanidade enquanto se valem de seus corpos (e mentes) como fonte de energia (pilhas ou baterias).

Ao longo da narrativa, observamos que a humanidade dentro da Matrix não é exatamente (ou exclusivamente) uma bateria, uma fonte de energia, mas unidades computacionais. Sendo assim, sair da Matrix não é exatamente abandonar ou libertar-se da Matrix, mas sim "tomar consciência", "dar-se conta" de que estamos na Matrix, para tanto é preciso decodificá-la, desvelá-la, para então reinicializá-la (refundá-la) sob novo paradigma. A título ilustrativo, para construir nosso argumento a partir da metáfora proposta, vale registrar a cena do filme em que há a seguinte afirmativa: "libere sua mente, para pular de um prédio para outro". A assertiva sugere que mesmo "dentro" da Matrix, podemos "estar fora" da Matrix, apropriando-se das habilidades humanas (da *artesania do processo* – expressão que cunhamos para a observância das garantias constitucionais ao interpretar e aplicar o Direito) e atuando como *outsiders* mesmo estando dentro da Matrix (decodificada).

Na tentativa de decodificar a Matrix e, portanto, decodificar as amarras para a efetivação do Acesso à Justiça, identificando a força do paradigma dominante, oportuno observarmos a obra "Acesso à Justiça", de Mauro Cappelletti e Bryan Garth (1988).

Não é novidade que o Projeto Florença foi importantíssimo para os estudos do direito processual e para as melhorias na prestação jurisdicional em muitos países ocidentais. Foram estudos empíricos realizados por Mauro Cappelletti, Bryan Garth e Nicoló Trocker que construíram, em nosso imaginário jurídico, a perspectiva das três ondas renovatórias, os três entraves para o acesso à justiça, nomeando-os para que fosse possível criar estratégias de enfrentamento.

Estas três ondas renovatórias, quais sejam, a preocupação (a) com os obstáculos econômicos (assistência judiciária aos pobres); (b) com os obstáculos organizacionais (representação dos interesses difusos) e (c) com os obstáculos burocráticos e formalísticos (novo enfoque de acesso à justiça, judicial e extrajudicial, medidas repressivas, reparatórias e preventivas) guiaram, num tenso contexto de implantação (ou tentativa de implantação) do paradigma do Estado Democrático de Direito, as reformas constitucionais e processuais nos anos 80 e 90 no mundo ocidental.

O movimento do acesso à justiça impulsionou a força normativa da Constituição de 1988 e outras legislações infraconstitucionais a partir de então, gerando as ondas reformistas do final dos anos 90 e início dos anos 2000. Podemos dar destaque aqui as preocupações com a proteção de grupos vulnerabilizados ou invisibilizados, as preocupações com as questões consumeristas e ambientais (de um modo geral preocupações com direitos

difusos e coletivos e até com direitos individuais homogêneos), mas também a ampliação das defensorias públicas, a criação dos juizados especiais e a preocupação com a oralidade, com o sincretismo processual e com cognição materialmente sumária no processo.

A Constituição de 1988 abrigou o movimento de acesso à justiça incorporando a preocupação com as ondas renovatórias, inspirando as ondas reformistas no Brasil. Era necessário, não só pensar em reformas legislativas e políticas públicas, mas também na prestação e na gestão jurisdicional à luz do paradigma constitucional que se forjava e se desejava fortalecer para estruturar um Estado Democrático de Direito. Era preciso implementar um programa de reforma do poder judiciário que desse conta da emergência deste novo paradigma - decodificando a Matrix - que reverberasse na atuação e na gestão jurisdicional e, quiçá, no imaginário social (ESPINDOLA, 2016).

3. O MOVIMENTO DO ACESSO À JUSTIÇA E AS DUAS ÚLTIMAS DÉCADAS DO SÉCULO XX

Qualquer problematização em torno do tema "acesso à justiça" passa pela conexão entre a "teoria do processo", a "teoria do estado" e a "teoria do direito". Consequentemente, temos que considerar relevante compreender a teoria do processo não pela perspectiva tradicional dos cortes históricos propostos por Zamora e Castillo (1974), mas especialmente pelo modo como os paradigmas jurídico-estatais ecoaram no sistema processual (ESPINDOLA, 2008; NUNES, 2009; NUNES; BAHIA; PEDRON, 2020).

Reconstruir o sistema processual, identificando a existência de sistemas processuais pré-liberais (século XIII a XVIII), antes da estruturação das bases de um liberalismo processual (século XIX e XX) e do surgimento da socialização do processo (primeira metade do século XX e pós-guerra) é identificar (dar-se conta) os bastidores do Projeto de Florença de Acesso à Justiça. É compreender que as crises do Estado Social (especialmente, no caso do Brasil, onde vivemos as crises de um Estado Social que nunca foi efetivamente implementado) geraram a necessidade de buscar novas estratégias e alternativas, justificando novos estudos no direito, e especialmente novos estudos sobre a jurisdição e o processo [(2)] que considerassem a importância de examinar os impactos do paradigma jurídico-estatal dominante.

A abertura democrática e as transições das roupagens do Estado Moderno, ocorreram no mundo ocidental como um desbloqueador da litigio-

(2) Não será possível nesse espaço realizar essa reconstrução do sistema processual, recomenda-se, contudo, que a leitura seja complementada com Nunes (2009) e Espindola (2005 e 2008b), bem como Bolzan de Morais.

sidade, gerando uma verdadeira inflação legislativa e uma revisão do papel do intérprete. No Brasil – que à época padecia das crises de um Estado social interrompido – as premissas da socialização do processo, as práticas sociais e processuais de um modelo publicista e social foram deturpadas ou mal compreendidas resultando em um trato privatístico das esferas pública (NUNES, 2009, p. 155), corrompendo o modelo idealizado da socialização do processo. E isso se deu, pois as ondas renovatórias dos anos 80 e as ondas reformistas dos anos 90 estouraram em areias que já haviam sido banhadas por ondas neoliberais em *terra brasilis*.

Às vésperas do século XXI, tivemos as alterações trazidas pela Lei 8.952/94, modificando o CPC/1973, em especial o art. 273, regulamentando a antecipação de tutela e também os arts. 461 e 461-A, que instituíram a tutela específica nos casos de obrigação de fazer, não-fazer e entregar. Somese também a Lei n. 9.079/95 e a Lei n. 9.139/95 que instituíram o procedimento monitório e as reformas no recurso de Agravo, respectivamente. Os "grandes inimigos", apontados à época, eram os "males corrosivos do tempo no processo" (DINAMARCO, 2001, p. 140). As mini-reformas dos anos 1994 e 1995, ganharam novo fôlego, intensificando-se, a partir de 1999, aproximando o processo do avanço das novas tecnologias de comunicação: a Lei n. 9.800/99 trouxe a possibilidade inovadora de peticionamento remoto (à distância), mediante o uso do Fax. Temos aqui um marco importante para a virada tecnológica no processo. O crescimento do ciberespaço ganhava velocidade e o conjunto de técnicas (materiais e intelectuais), de práticas, de atitudes, de modos de pensamento e de valores que se desenvolviam no ciberespaço (a cibercultura) avançava igualmente (LEVY, 2018, p. 13).

O Documento Técnico n. 319, do Banco Mundial, redigido nos anos 90 (acima mencionado), deu o tom das prioridades e da direção das ondas reformistas na América Latina do século XXI (DAKOLIAS, 1996). No Brasil, no âmbito do processo civil, este período foi frenético em reformas (e de grande atividade no mercado editorial, com inúmeras produções de livros e artigos sobre as novas e as novíssimas reformas da época). Era o contexto das mini-reformas, como denominou Dinamarco (2001, p. 32), cujo objetivo era eliminar (ou mitigar) empecilhos ao acesso à justiça.

Sublinhe-se que o Banco Mundial estava na arena direcionando as reformas do judiciário nos países da América Latina e no Caribe, com as recomendações para um programa global de reforma do judiciário a ser adaptado as situações específicas de cada país, visando maior confiança no mercado e no setor privado, com o Estado atuando como um importante facilitador e regulador das atividades. O objetivo era estimular "reformas para aprimorar a qualidade e a eficiência da Justiça, fomentando um ambiente propício ao comércio, ao financiamento e investimentos" (DAKOLIAS, 1996). E tudo isso chegou ao Brasil quando estávamos vivendo um contexto marcado pelos avanços de políticas neoliberais onde perspectivas econô-

micas que traduziam a implementação das reformas reduzindo o processo a uma perspectiva meramente instrumental, funcionalizada. O objetivo das recomendações do Banco Mundial era priorizar um processo célere e o desenvolvimento do setor privado (DAKOLIAS, 1996, p. 10). Estas recomendações foram escutadas no Brasil ao som do modelo neoliberal implantado nos anos 90.

Concluída a última década do século XX - a década das "mini-reformas" - avançávamos para a primeira década do século XXI - a década da "reforma da reforma", para nos apropriarmos aqui de outra expressão de Dinamarco (2003). Uma nova conjuntura tentava organizando-se e os três poderes reuniram-se para direcionar as reformas do processo, da jurisdição e do Poder Judiciário, buscando reler, traduzir e adaptar as ondas renovatórias.

4. A CIBERCULTURA, O ACESSO À JUSTIÇA E AS DUAS PRIMEIRAS DÉCADAS DO SÉCULO XXI

Iniciado o novo século, a reforma do judiciário ainda estava incompleta e os desafios ao acesso à justiça ainda eram muitos, somando-se ao conjunto de tendências da cibercultura. Estudos e discussões visando construir alternativas políticas, administrativas e jurisdicionais pululavam na América Latina como um todo, mas também na Europa continental. No Brasil, as preocupações eram sobretudo com o crescimento econômico (FARIA, 2010).

Em 2004, após 13 anos de tramitação, aprovou-se a Reforma do Poder Judiciário no Brasil com a publicação da Emenda Constitucional n. 45, em 17 de novembro de 2004. Na época, tramitavam outras 17 propostas de emendas constitucionais (PECs) (ANGRA, 2005).

O problema da efetividade da prestação jurisdicional e as crises do modo estatal de d*izer o direito* – jurisdição – refletiam questões infraestruturais (relacionadas a escassez de recursos, inadaptações de caráter tecnológico), questões formativas dos atores da prestação jurisdicional diante do aumento da complexidade social, bem como questões relacionadas ao aumento dos bens de direito e da multiplicação dos sujeitos de direito assim como dos pluralidade dos *status* dos sujeitos de direito nas relações jurídicas, mas sobretudo estas crises refletiam questões de cunho paradigmático estrutural, dificilmente abaláveis por reformas meramente legislativas ou funcionais.

A ampliação do acesso à justiça e a melhoria na qualidade dos serviços prestados era o grande mote. É nesse cenário que foram assinados o I e o II Pactos de Estado, reunindo os três Poderes de Estado.

O primeiro deles chamou-se "Pacto em favor de um judiciário mais rápido e republicano", firmado à época pelo Presidente da República Luís Inácio Lula da Silva, pelo Presidente do STF, Nelson Jobim, pelo Presidente do Senado, José Sarney e pelo Presidente da Câmara, João Paulo Cunha. Esse pacto foi lançado logo após a reforma do judiciário, em dezembro de 2004, recomendando, entre outras medidas, a aprovação de um conjunto de projetos para alterar a legislação civil, trabalhista e penal. A reforma do judiciário já havia instituído a Súmula Vinculante nas decisões do STF e criado o Conselho Nacional da Justiça (CNJ), era agora preciso avançar nas reformas atinentes à legislação processual (ANDRADE, 2006; SADEK, 2010). Pensar funcionalmente (e não estruturalmente) a crise do judiciário era o foco.

O I Pacto (2004) foi mais específico que o II Pacto (2009) e listava 11 objetivos, a saber: (1) implementação da reforma constitucional do Judiciário; (2) reforma do sistema recursal e dos procedimentos; (3) defensoria pública e acesso à justiça; (4) juizados especiais e justiça itinerante; (5) execução fiscal; (6) precatórios; (7) graves violações contra os direitos humanos; (8) informatização; (9) produção de dados e indicadores estatísticos; (10) coerência entre a atuação administrativa e as orientações jurisprudenciais já pacificadas; (11) incentivo à aplicação das penas alternativas. Este acordo entre os Poderes impulsionou a aprovação de mais de 20 projetos de lei, especialmente no âmbito do Código de Processo Civil. Novas minirreformas do direito processual orientadas por estes objetivos tiveram início.

Cinco anos depois, era hora de assinar um novo pacto: *II Pacto Republicano de Estado por um sistema de justiça mais acessível, ágil e efetivo* [3], elegendo quais as materiais prioritárias para a nova década, a saber: (1) proteção dos direitos humanos e fundamentais; (2) agilidade e efetividade da prestação jurisdicional; (3) acesso universal à justiça.

Os Pactos fizeram-se acompanhar de um sistema de planejamento, controle, métricas e preocupações eficientistas, calcadas na lógica do produtivismo: surgiam o Justiça em Números de 2004 (ano-base 2003) e as Metas do CNJ (2009). Os Relatórios "Justiça em Números", que surgem em 2004, são a principal e mais confiável base de dados quantitativa das estatísticas oficiais do Poder Judiciário. Estes relatórios colocaram holofotes em informações sobre a estrutura e a litigiosidade dos tribunais até então desconhecidos, fornecendo indicadores para subsidiar a gestão judiciária nacional [4].

As metas do CNJ começaram a ser traçadas no ano de 2009 e resultam anualmente de compromissos firmados entre os presidentes dos tribu-

(3) Disponível em http://www.planalto.gov.br/ccivil_03/outros/iipacto.htm

(4) Sobre os relatórios consultar https://www.cnj.jus.br/pesquisas-judiciarias/justica-em-numeros/

nais [5]. Estas metas observam e guiam-se pelos Relatórios "Justiça em Números". Nessa trilha, a Meta 2 mereceu bastante destaque, vez que elegeu como foco os estoques de processos causadores do congestionamento nos tribunais.

Como bem observaram Sadek e Arantes, todo o contexto das reformas do judiciário, em prol do acesso à justiça, foram marcados por dois fenômenos: a importância assumida por essa questão no debate público e, ao mesmo tempo, a dificuldade de construir acordos suficientes para a implementação de mudanças (SADEK, 2010). De fato, a agenda política dos anos 90 modificou o perfil do Estado e sua relação com a economia e com a sociedade e, fundamentalmente, isso acabou por ecoar na prestação jurisdicional, não só sob o ponto de vista dos custos da estrutura jurisdicional e do desempenho do sistema judiciário, mas sobretudo em função do papel político do Poder Judiciário na democracia brasileira, enquanto controle de abusos dos demais poderes do Estado. Gradualmente o diagnóstico de Nicola Picardi ao sinalizar a vocação do nosso tempo para a jurisdição e para a doutrina jurídica foi se confirmando. (PICARDI, 2008, p. 2). De fato, hoje podemos falar no agigantamento do Judiciário. Registre-se, que Picardi (2008), refere que o século XXI seria um século vocacionado para o Judiciário, mas também pela doutrina jurídica.

Com a atenção voltada para o Poder Judiciário, apostou-se que o crescimento da economia dependeria de um "ambiente saudável para os negócios" por meio de tribunais capazes de proporcionar estabilidade e segurança aos agentes. A preocupação na primeira década do século XXI era sobretudo com a construção de uma ordem legal clara, precisa e confiável, e com a conversão das instituições judiciais em cortes eminentemente técnicas, altamente profissionalizadas, capazes de tomar decisões coerentes, previsíveis e não retroativas, afastar o risco regulatório e assegurar fluidez para os negócios (FARIA, 2010, p. 88).

A partir da segunda década do século XXI a preocupação passa a ser medir, metrificar a atuação e o desempenho dos tribunais.

As estratégias de "metrificação" adotadas no Brasil, refletem de certa forma, estratégias também adotadas na União Europeia. Em 2013 – primeira década do século XXI – foi lançado o "Painel de Avaliação da Justiça na EU", no intuito de acompanhar as reformas do judiciário. No entanto, se cotejarmos os números do "Painel de Avaliação da Justiça na EU de 2020" [6] facilmente vamos observar o contraste entre os sistemas judiciários brasileiro e europeus, revelando a gravidade do problema do acesso à justiça no Brasil.

---

(5) Sobre as Metas do CNJ, consultar https://www.cnj.jus.br/gestao-e-planejamento/metas/sobre-as-metas/

(6) Disponível em https://ec.europa.eu/info/files/2020-eu-justice-scoreboard-factsheet_en

Se compararmos a Itália – o mais litigante dos países europeus, podemos observar que as respostas jurisdicionais na Itália são mais céleres e que a taxa de congestionamento e o estoque de processos são melhores que os números do judiciário brasileiro (WOLKART, 2019, p. 89).

Como bem-sinalizado pelas pesquisas de Faria, a tendência a partir da segunda década do século XXI, no Brasil, seria o aumento no ritmo de regressão tanto dos direitos sociais quanto dos direitos humanos consagrados ou tutelados pelo direito positivo, implicando no rebaixamento qualitativo da própria cidadania, diante do enxugamento do Estado-nação e da retração da esfera pública, havendo, ainda, a sobreposição do papel do consumidor ao papel do trabalhador (FARIA, 2010, pp. 104–105).

Iniciativas de acesso à justiça certamente não se resumem a uma preocupação numérica ou ao atingimento de métricas, os "cálculos" precisam dar conta de uma preocupação com a efetividade da prestação jurisdicional, mas não só, vez que por acesso à justiça não se pode entender apenas o acesso aos tribunais, tampouco a uma perspectiva doméstica, nacional, de acesso à justiça. A preocupação exclusiva com o eficientismo pode levar ao paradoxo kafkiano, no qual a busca pelo acesso à justiça obstaculiza o próprio acesso à justiça (WOLKART, 2019, p. 63) e o remédio se transforma em veneno, matando o paciente. A cibercultura, a virada tecnológica, neste ponto, surgiria como aliada ou como vilã?

Ultrapassada a segunda década do século XXI é preciso questionar as premissas anteriores e reavaliar o momento histórico, visualizado, por óbvio, o impacto da cibercultura para os avanços ou realinhamentos na rota do acesso à justiça.

Para Pierre Levy, ao falar sobre a cibercultura, a metáfora bélica do "impacto" da tecnologia é inadequada (LÉVY, 2009, p. 21), pois a tecnologia não seria um ator autônomo separado da sociedade e da cultura, assim como estas não seriam entidades passivas percutidas por um agente exterior. Aquilo que definimos como "novas tecnologias" na verdade consistem na atividade multiforme de grupos humanos, um devir coletivo complexo que se cristaliza em volta de objetos materiais, de programas de computadores e de dispositivos de comunicação ou do uso de tecnologia na gestão do judiciário ou no apoio ao processo decisório (LÉVY, 2009, p. 28). Quando nos referimos às "novas tecnologias" ou à "virada tecnológica" estamos nos referindo ao processo social em toda sua opacidade, estamos nos referindo a atividade dos outros. E, claro, se os "impactos" são negativos é natural que alguém seja responsabilizado, incriminado; e se os "impactos" são positivos é natural que alguém seja deificando.

Embora desejável, não vamos conseguir aqui dar conta de tudo que envolve a compreensão da cibercultura em Levy, mas é importante trazer o re-

gistro de que é no ciberespaço que, segundo ele, se desenvolverá uma inteligência coletiva, com a mesma probabilidade que se desenvolvam bobagens coletivas, assim como dominação, exploração, isolamento, dependência e sobrecarga cognitiva (LEVY, 2018).

Na nossa leitura, a cibercultura e a virada tecnológica no processo não trazem em si avanços ou retrocessos para o acesso à justiça, embora ela possa – e a probabilidade parece-nos bastante alta - ocultar problemas que já vinham sendo identificados assim como ocultar problemas estávamos desvelando ou que ainda desconhecemos. Ora, o movimento do acesso à justiça deságua, no Brasil, no contexto de um neoliberalismo processual, onde as métricas, o produtivismo e o eficientismo eram os itens prioritários. A cibercultura e a virada tecnológica no processo, desacompanhadas de uma refundação da jurisdição – uma (re)volta (d)as suas estruturas paradigmáticas – poderá fortalecer o neoliberalismo processual, que assume a roupagem de um tecnoneoliberalismo processual, expressão que temos utilizado justamente para representar como um determinado paradigma dominante pode ocultar-se e fortalecer-se no ciberespaço, hackeando a cibercultura, nos mantendo na Matrix (multiplicando o agente Smith). Há grande chances de que, continuando o debate, isso nos leve a problematizar o ensino do direito, assim como a educação para os direitos humanos. (ESPINDOLA; SEEGER, 2018)

5. Considerações finais: Reinicializar ou Destruir a Matrix: quem controla quem programa? Quem é o Arquiteto?

Na tentativa de manter coerência com a metáfora proposta, é bom lembrar que em Matrix: Reloaded, quando Neo encontra o criador da Matrix, o Arquiteto, este lhe conta que ele é uma parte intencional do sistema, criado para evitar uma queda fatal e que deveria escolher entre voltar para a fonte e reiniciar a Matrix ou derrubar a Matrix, desconectando (e matando) a todos. O dilema da escolha neste caso representa a binariedade, a linearidade, a linguagem de programação com padrões binários. Se o Arquiteto é o criador da Matrix (o sistema jurídico) e se Neo (o intérprete) é "irrevogavelmente humano"; então embora o processo tenha "alterado sua consciência", como anunciado no diálogo com o Arquiteto, não é a linguagem binária que "programa" (ou comanda) o intérprete, ainda que o auxilie na redução das complexidades. O Arquiteto cria a Matrix, mas é o intérprete que programa o Arquiteto e sua linguagem binária.

Se de fato o nosso tempo é vocacionado para a jurisdição (PICARDI, 2008), o tema do acesso à justiça estará em pauta ao longo de todo o século XXI, precisando ser sempre atualizado. Se o arbítrio do intérprete (ativismo judicial) era (e ainda é) uma preocupação legítima na atuação jurisdicio-

nal, temos que estar atentos ao agigantamento da tecnologia assim como às arbitrariedades do algoritmo. Na verdade, o problema não está na tecnologia ou em seu agigantamento ou em identificar seu impacto. De fato, como diz Toynbee, a tecnologia em si não é neutra, vez que carrega em si formas particulares de conhecimento e práticas que se impõem aos usuários, que se sentem obrigados a submeter-se a elas (TOYNBEE, 1968). De fato, subestimar a tecnologia como manifestação meramente instrumental conduz a uma abordagem binária, dualista, linear, simplista (HUI, 2020).

Nesse sentido, a proposta de Yuk Hui de combate o monotecnologismo (HUI, 2020, p. 200) como também de combate ao solucionismo tecnológico (BERARDI, 2019; MOROZOV, 2018) e os alertas sobre os algoritmos de destruição em massa, os ADMs (O'NEIL, 2020). Decodificar a Matrix, para que possamos Refundar a Jurisdição, escapando de um tecnoliberalismo processual atualizando as compreensões sobre acesso à justiça, implica necessariamente em afastar essa sincronização e confrontarmos o conceito de tecnologia em si, afastando-nos de uma perspectiva fundada numa força exclusivamente produtiva e do aumento da mais-valia. Sobretudo para nós, com demandas latino-americanas de acesso à justiça, temos essa urgência. No âmbito da jurisdição (e consequentemente do acesso à justiça) a pergunta continua sendo quem controla quem controla, embora possa ter sua linguagem atualizada para quem programa o programador.

Contextualizar e problematizar o acesso à justiça nos liberta da tirania das métricas, da linguagem binária (simplista) e da ameaça de um tecnoneoliberalismo processual, os quais provocam o agigantamento do acesso à justiça quantitativo e o silenciamento do acesso à justiça qualitativo, nos afastando das promessas do Estado Democrático de Direito e nos sincronizam com um discurso hegemônico e monotecnológico. Nosso debate, nossos desafios e nosso fazer são científicos, assim como as respostas do Direito também devem ser científicas. No entanto, juristas que somos precisamos estar atentos a fenômenos do nosso território e nosso tempo e, para tanto, devemos acessar realidades latino-americanas, sua cultura, sua literatura, sua arte, sua poesia e todas as manifestações humanas. Isso sim, irá impactar a tecnologia, diversificando-a e customizando-a às demandas de um acesso à justiça do nosso tempo. Cientes de que o debate é extenso e profundo e que, como pesquisadoras do direito, devemos adotar a postura de betas perpétuos [7], este ensaio é um convite para que possamos pensar e dialogar sobre o tema do acesso à justiça no contexto da cibercultura, decodificando a Matrix para que possamos, superando a linguagem algorítmica binária, criar mundos e refundar a Jurisdição.

---

(7) O tema da educação jurídica parece-nos, portanto, incontornável e que merece ser combinado com a formação de uma cultura literária do direito no Brasil (KARAM TRINDADE, 2020; PAIXÃO, 2020).

Referências bibliográficas

Alcalá-Zamora y Castillo, N. *Estudios de teoría general e histórica del proceso*. México: UNAM, 1974.

Andrade, F. M. de. Reforma do poder judiciário: aspectos gerais, o sistema de controle de constitucionalidade das leis e a regulamentação da súmula vinculante. *Revista de Informação Legislativa*, v. 43, n. 171, p. 177–197, 2006.

Angra, W. de M. *Comentários à reforma do poder judiciário*. Rio de Janeiro, RJ: Forense, 2005.

Berardi, F. *Depois do futuro*. São Paulo, SP: Ubu Editora, 2019.

Cappelletti, M.; Garth, B. *Acesso à Justiça*. Porto Alegre, RS: Safe, 1988.

Dinamarco, C. Rangel. *A reforma do código de processo civil*. 4. ed. São Paulo, SP: Malheiros, 2001.

Dinamarco, C. Rangel. *A reforma da reforma*. 6. ed. rev. São Paulo, SP: Malheiros, 2003.

Espindola, A. A. da S. *Superação do racionalismo no processo civil enquanto condição de possibilidade para a construção das tutelas preventivas: um problema de estrutura ou função? (Ou: por que é preciso navegar em direção à ilha desconhecida e construir o direito processual civil)*. Tese de doutorado. São Leopoldo, RS: Universidade do Vale do Rio dos Sinos, abr. 2008. Acesso em: 7 set. 2021.

Espindola, A. A. da S. Entre a insustentabilidade e a futilidade: a jurisdição, o direito e o imaginário social sobre o juiz. *Anamorphosis. Revista Internacional de Direito e Literatura*, v. 2, n. 2, p. 293–320, 31 jan. 2016.

Espindola, A. A. da S.; Seeger, L. D. S. O ensino jurídico no Brasil e o senso comum teórico dos juristas: um olhar a partir de Warat. *Revista de Direito da Faculdade Guanambi*, v. 5, n. 02, p. 92–120, 1 dez. 2018.

Faria, J. E. *Direito e conjuntura*. São Paulo, SP: Saraiva, 2010.

Hui, Y. *Tecnodiversidade*. São Paulo, SP: Ubu Editora, 2020.

Husserl, E. *Investigações lógicas: fenomenologia e teoria do conhecimento*. Rio de Janeiro, RJ: Forense Universitária, 2012.

Husserl, E. *A ideia da fenomenologia: cinco lições*. São Paulo, SP: Editora Vozes, 2020.

Karam, H. O direito na contramão da literatura: a criação no paradigma contemporâneo. *Revista Eletrônica do Curso de Direito da UFSM*, v. 12, n. 3, p. 1022–1043, 2017.

Karam Trindade, A. El aporte de José Calvo González a la cultura literaria del derecho en Brasil. *LawArt*, v. 1, n. 1, 30 jan. 2020.

Laporte, A. M.; Volpe, N. *Existencialismo: uma reflexão antropológica e política a partir de Heidegger e Sartre*. [s.l.]: Editora Juruá, 2000.

Lévy, P. *Cibercultura*. São Paulo, SP: Editora 34, 2009.

Morozov, E. *Big Tech: a ascensão dos dados e a morte da política*. São Paulo, SP: Ubu Editora, 2018.

Nunes, D. J. C. *Processo jurisdicional democrático*. Curitiba, PR: Juruá, 2009.

Nunes, D. J. C.; Bahia, A.; Pedron, F. Q. *Teoria geral do processo*. Salvador, BA: JusPodivm, 2020.

O'Neil, C. *Algoritmos de destruição em massa*. Santo André, SP: Editora Rua do Sabão, 2020.

Paixão, C. O lugar da literatura na educação jurídica: três urgências contemporâneas. *LawArt*, v. 1, n. 1, 30 jan. 2020.

Pedron, F. B. Q. *O poder judiciário e(m) crise*. Belo Horizonte, MG: [s.n.].

Pedron, F. Q. Acesso à justiça qualitativo no Estado Democrático de Direito. *Revista Jus Navigandi*, 2013.

Picardi, N. *Jurisdição e processo*. Rio de Janeiro: Forense, 2008.

Sadek, M. T. *Reforma do judiciário*. Rio de Janeiro, RJ: Centro Edelstein de Pesquisas Sociais, 2010.

Teixeira, L. F. *Acesso à justiça qualitativo*. Dissertação de Mestrado. Pouso Alegre, MG: [s.n.].

Toynbee, A. J. *O desafio do nosso tempo*. Rio de Janeiro, RJ: Zahar, 1968.

Trindade, A. K.; Bernsts, L. G. O estudo do direito e literatura no Brasil: surgimento, evolução e expansão. *Anamorphosis. Revista Internacional de Direito e Literatura*, v. 3, n. 1, p. 225–257, 2017.

Trindade, A. K.; Gubert, R. M. Direito e literatura: aproximações e perspectivas para se repensar o direito. In: *Direito & literatura: reflexões teóricas*. Porto Alegre: Livraria do Advogado, 2008. p. 11–66.

Wolkart, E. N. *Análise econômica do processo civil*. São Paulo, SP: Thomson Reuters Brasil, 2019.