

DERECHOS HUMANOS ANTE EL RETO DE LA INTELIGENCIA ARTIFICIAL



Directores: Gabriela C. Cobo del Rosal Pérez. Pablo Javier Fernández de Casadevante Mayordomo

INSTITUTO DE ESTUDIOS JURÍDICOS INTERNACIONALES, CONDE DE ARANDA

Dykinson, S.L.

Directores

GABRIELA C. COBO DEL ROSAL PÉREZ

PABLO FERNÁNDEZ DE CASADEVANTE MAYORDOMO

DERECHOS HUMANOS ANTE EL RETO DE
LA INTELIGENCIA ARTIFICIAL

Edición a cargo de Gabriela C. Cobo del Rosal Pérez

Dykinson, S.L.

Todos los derechos reservados. Ni la totalidad ni parte de este libro, incluido el diseño de la cubierta, puede reproducirse o transmitirse por ningún procedimiento electrónico o mecánico. Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra (www.conlicencia.com; 91 702 19 70 / 93 272 04 47)

Editorial DYKINSON, S.L. Meléndez Valdés, 61 - 28015 Madrid
Teléfono (+34) 91 544 28 46 - (+34) 91 544 28 69
e-mail: info@dykinson.com / www.dykinson.es / www.dykinson.com

Este libro ha sido sometido a evaluación por parte de nuestro Consejo Editorial.

Para mayor información, véase Consejo Editorial:
www.dykinson.com/quienes_somos

© Los autores
Madrid, 2025

Portada: Ana de la Cerda Aparicio. Fenetre73, @ana.de.la.cerda.pintura

ISBN: 979-13-7047-099-9

A los que trabajan por la Paz

ÍNDICE

CAPÍTULO 1. DERECHOS HUMANOS ANTE EL RETO DE LA INTELIGENCIA ARTIFICIAL, Gabriela C. Cobo del Rosal Pérez..... pp. 15-25

INTRODUCCIÓN

1. LA INTELIGENCIA ARTIFICIAL COMO “CIRCUNSTANCIA”
 2. DERECHOS HUMANOS ANTE EL RETO DE LA INTELIGENCIA ARTIFICIAL: UNA APROXIMACIÓN MULTIDISCIPLINAR
 3. REFLEXIÓN FINAL A MODO DE CONCLUSIÓN
- BIBLIOGRAFÍA

CAPÍTULO 2. COOPERACION INTERNACIONAL INSTITUCIONALIZADA FRENTE A LOS RIESGOS PARA LA DEMOCRACIA, LA SEGURIDAD Y LOS DERECHOS HUMANOS DERIVADOS DEL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL, M^a Ángeles Cano Linares, ... pp. 27-74

INTRODUCCIÓN

1. PRINCIPALES RIESGOS DERIVADOS DEL USO/MAL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL
 - 1.1 El uso malicioso de los sistemas IA. Ciberdelitos y Ciberseguridad
 - 1.2 Su incidencia en el estado de derecho: derechos humanos, democracia, brecha digital y gobernanza.
 - 1.3 Las neuro tecnologías: Revolución en ciernes
 2. PROPUESTAS EN EL AMBITO UNIVERSAL: PACTO DIGITAL MUNDIAL Y ÉTICA EN LA INTELIGENCIA ARTIFICIAL
 - 2.1 Naciones Unidas: Pacto Digital Mundial
 - 2.2. La labor de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura y su enfoque ético.
 3. RIESGOS ESPECIFICOS EN CASO DE CONFLICTO ARMADO
 - 3.1 Amenazas híbridas, conflictos cibernéticos
 - 3.2. Inteligencia Artificial y Derecho Internacional Humanitario
 - 3.3 La preocupación del Consejo de Seguridad de Naciones Unidas
 4. LA DEFENSA DE LA DEMOCRACIA Y EL ESTADO DE DERECHO: CONVENIO MARCO DEL CONSEJO DE EUROPA
 5. CONCLUSIONES
- APÉNDICE BIBLIOGRÁFICO

CAPÍTULO 3. POSICIÓN DE CASILLA “d8”: SUPERVISIÓN HUMANA, Ana Isabel Carreras Presencio pp.75-90

INTRODUCCIÓN

1. EL SESGO
 - 1.1.- El sesgo en la IA.
 - 1.2.- El sesgo de género en la IA.

2. POSICIÓN DE CASILLA “d8”
 - 2.1.- La interseccionalidad
 - 2.2.-Supervisión humana. Buenas prácticas
3. EDUCACIÓN EN IA.
4. CONCLUSIONES
- BIBLIOGRAFÍA

CAPÍTULO 4. SESGOS ALGORÍTMICOS Y DERECHOS HUMANOS: DESAFÍOS ÉTICOS Y REGULATORIOS EN LA ERA DE LA INTELIGENCIA ARTIFICIAL,

Angélica María Castro Acosta..... pp. 91-142

INTRODUCCIÓN

1. MARCO TEÓRICO

- 1.1. Conceptualización de los Sistemas de IA
- 1.2. Entendiendo los Algoritmos

2.. TAXONOMÍA DE LOS SESGOS ALGORÍTMICOS

- 2.1. Tipos de Sesgos
 - 2.1.1. Sesgo de Representación (*Representation Bias*)
 - 2.1.2. Sesgo de Medida (*Measurement Bias*)
 - 2.1.3. Sesgo de etiquetado (*Label Bias*)
 - 2.1.4. Sesgo por objetivo (*Objective Bias*)
 - 2.1.5. Sesgo Emergente en Sistemas Complejos (*Emergent Bias*) (Comportamiento no previsto)

3. DERECHOS HUMANOS EN LA ERA DIGITAL

- 3.1. Marco normativo europeo y español
- 3.2. Impacto sobre derechos humanos y principios afectados
 - 3.2.1. Derecho a la igualdad y no discriminación
 - 3.2.2. Derecho a la privacidad y protección de datos personales
 - 3.2.3. Debido proceso y derecho a una decisión motivada
- 3.3. Previniendo los Sesgos Algorítmicos

4. CONCLUSIONES

APÉNDICE BIBLIOGRÁFICO

CAPÍTULO 5. ABOGACÍA E INTELIGENCIA ARTIFICIAL: ALGUNAS NOTAS SOBRE LAS OPORTUNIDADES Y DESAFÍOS PARA LA PROFESIÓN,

Irene Blanca Cerrillo Torquemada..... pp. 143-168

INTRODUCCIÓN

1. UTILIZACIÓN POR LOS ABOGADOS: *LEGAL PROMPTING*

2. RECOMENDACIONES OFICIALES PARA EL ÁMBITO DE LA ABOGACÍA

3. CONCLUSIONES

APÉNDICE BIBLIOGRÁFICO Y FUENTES

CAPÍTULO 6. IDENTIFICACIÓN BIOMÉTRICA REMOTA EN EL ÁMBITO POLICIAL, ¿CÓMO AFECTA A LA PROTECCIÓN DE DATOS DE CARÁCTER PERSONAL?, *Cristina Domingo Jaramillo*.....pp. 169-195

INTRODUCCIÓN

1. DATOS BIOMÉTRICOS COMO DATOS PERSONALES ESPECIALMENTE SENSIBLES

1.1. Distinción de los datos biométricos en función de la intrusividad en su recolección

1.2. Disponibilidad o visibilidad del rostro en el espacio público

2. IMPACTO DE LA BIOMETRÍA EN LA PROTECCIÓN DE DATOS PERSONALES: ESPECIALMENTE EL RECONOCIMIENTO FACIAL

2.1. La protección de datos personales en los distintos momentos en los que opera el sistema

2.2. Fuentes de las que se extrae la información: ¿Qué ocurre con las imágenes publicadas en redes sociales?

3. CONCLUSIONES

APÉNDICE BIBLIOGRÁFICO

CAPÍTULO 7. ENTRE LA LEGALIDAD Y EL ALGORITMO: PRINCIPIOS PROCESALES E INTELIGENCIA ARTIFICIAL EN LA EJECUCIÓN DE MEDIDAS DEL PROCESO PENAL DE MENORES, *Elena Pineros Polo*.....pp. 197-224

INTRODUCCIÓN: EJECUCIÓN DE MEDIDAS EN CLAVE GARANTISTA Y DIGITAL

1. CATÁLOGO DE MEDIDAS Y CRITERIOS DE IMPOSICIÓN CON EFECTO EN LA EJECUCIÓN

2. ARQUITECTURA DE LA EJECUCIÓN: ROLES, INSTRUMENTOS Y PROCEDIMIENTOS

3. PRINCIPIOS PROCESALES RECTORES EN LA EJECUCIÓN DE MEDIDAS

4. INTELIGENCIA ARTIFICIAL Y EJECUCIÓN DE MEDIDAS: COMPATIBILIDAD Y SALVAGUARDAS

4.1. La Inteligencia artificial como elemento de apoyo en la fase de ejecución.

4.2. Riesgos procesales y criterios de compatibilidad.

5. CONCLUSIONES

BIBLIOGRAFÍA

CAPÍTULO 8. JUSTICIA EN PLAZO, SIN DILACIONES, E INTELIGENCIA ARTIFICIAL: ALGUNAS PROPUESTAS, *Aurelio de Prada García*..... pp. 225-248

INTRODUCCIÓN

1. *DE NOBIS IPSIS NON SILEMUS*
2. *THE FIRST THING WE DO, LET'S KILL ALL THE LAWYERS*
3. ALGUNAS PROPUESTAS
4. ¿PLAZOS RAZONABLES Y DILACIONES INDEBIDAS?
5. A MODO DE CONCLUSIÓN

BIBLIOGRAFÍA UTILIZADA

CAPÍTULO 9. GEOPOLÍTICA, GOBERNANZA E INTELIGENCIA ARTIFICIAL, *Fernando Gil González*.....pp. 249-264

INTRODUCCIÓN: LA INTELIGENCIA ARTIFICIAL Y LA TRANSMISIÓN DEL CONOCIMIENTO

1. LA CONVERGENCIA DE BIG DATA: APRENDIZAJE AUTOMÁTICO
2. LA CUESTIÓN DE LA SOBERANÍA DE LOS GOBIERNOS
3. EL DUOPOLIO CHINO-ESTADOUNIDENSE: ESTADOS UNIDOS Y CHINA
4. ¿GEOPOLÍTICA DE BLOQUES? LA SITUACIÓN DIGITAL DE LA UNIÓN EUROPEA
5. EPÍLOGO: ¿HACIA DÓNDE NOS DIRIGIMOS?
6. CONCLUSIONES

BIBLIOGRAFÍA

CAPÍTULO 10. INTELIGENCIA ARTIFICIAL, ROBOTIZACIÓN DE LA JUSTICIA: ESPECIAL REFERENCIA A COMO SE VEN AFECTADOS LOS DERECHOS HUMANOS Y LAS GARANTÍAS PROCESALES, *María Concepción Rayón Ballesteros*..... pp. 265-287

INTRODUCCIÓN

1. CONCEPTOS FUNDAMENTALES: IA JUDICIAL Y ROBOTIZACIÓN DE LA JUSTICIA

- 1.1. Inteligencia Artificial Judicial: definición y tipología
- 1.2. Diferencias conceptuales: IA vs. Robotización

2. MARCO NORMATIVO INTERNACIONAL Y EUROPEO

- 2.1. La Carta de Derechos Fundamentales de la Unión Europea y el Convenio Europeo de Derechos Humanos
- 2.2. Reglamento General de Protección de Datos (RGPD)
- 2.3. Reglamento (UE) 2024/1689 sobre Inteligencia Artificial (Ley de IA)
- 2.4. Recomendaciones del Consejo de Europa y la UNESCO

- 2.5. Derecho comparado: experiencias internacionales
- 2.6. Necesidad de una regulación procesal específica
- 3. PRINCIPALES DERECHOS HUMANOS AFECTADOS
 - 3.1. Dignidad humana
 - 3.2. Derecho a un juicio justo, con todas las garantías y al derecho de defensa
 - 3.3. Igualdad y no discriminación
 - 3.4. Derecho a la privacidad y protección de datos personales
 - 3.5. Libertad de información
 - 3.6. Tutela judicial efectiva y acceso a los tribunales
- 4. RIESGOS JURÍDICOS Y ÉTICOS
 - 4.1. Sesgos algorítmicos
 - 4.2. Opacidad y el problema de la “caja negra”
 - 4.3. Pérdida de supervisión humana y automatización de la autoridad
 - 4.4. Ciberseguridad
 - 4.5. Deshumanización de la justicia
- 5. PROPUESTAS Y BUENAS PRÁCTICAS
 - 5.1. Principio “*Human in the Loop*”: supervisión humana significativa
 - 5.2. Auditorías algorítmicas obligatorias y certificación
 - 5.3. Explicabilidad y transparencia algorítmica
 - 5.4. Regulación procesal
 - 5.5. Alfabetización para operadores jurídicos
 - 5.6. Principios de diseño ético: equidad, inclusión y accesibilidad
 - 5.7. Estrategia de ciberseguridad integral
 - 5.8. Gobernanza algorítmica y supervisión institucional con creación de entornos de prueba o *sandbox*
- 6. CONCLUSIONES
- BIBLIOGRAFÍA

CAPÍTULO 11. DERECHOS FUNDAMENTALES DEL ART. 20. 1, a) y d) CE: IMPACTO DE LA INTELIGENCIA ARTIFICIAL, *Pedro Ródenas Cortés*.....pp. 289-318

BREVE PROEMIO

- 1. CONCEPTO Y CARACTERÍSTICAS DE LA INTELIGENCIA ARTIFICIAL
 - 1.1. Antecedentes y fundamentos
 - 1.2. Concepto, características y tipología
- 2. LIBERTAD DE EXPRESIÓN Y DE INFORMACIÓN
 - 2.1. Titularidad y contenido del derecho de libertad de expresión e información
 - 2.2. Límites y conflictos con otros derechos
 - 2.3. Requisitos de la libertad de información: veracidad y diligencia
 - 2.4. Garantías constitucionales

- 2.5. Referencia a la dimensión internacional de la libertad de expresión
- 3. IMPACTO DE LA INTELIGENCIA ARTIFICIAL
 - 3.1. Funcionalidad de la inteligencia artificial confrontada a la libertad de expresión e información
 - 3.2. Nuevas formas de participación y debate público condicionadas por la inteligencia artificial
 - 3.3. Amenazas emergentes de la inteligencia artificial sobre la libertad de expresión e información
 - 3.4. El dominio de la IA en el comportamiento e información recibida por jóvenes
- 4. LA RESPONSABILIDAD CIVIL DERIVADA DEL MAL USO DE LA INTELIGENCIA ARTIFICIAL
- 5. CONCLUSIONES
- BIBLIOGRAFÍA
- REPERTORIO JURISPRUDENCIAL

CAPÍTULO 12. IMPACTO DE LA INTELIGENCIA ARTIFICIAL EN EL ÁMBITO DE LA PROPIEDAD INTELECTUAL, TENSIONES Y DESAFÍOS, *Alejandro Rosillo Fairén*.....pp. 319-346

- 1. ORÍGENES DE LA IA. APROXIMACIÓN A SU DEFINICIÓN. ACERCAMIENTO A SU DIMENSIÓN ÉTICA
- 2. MARCO NORMATIVO BÁSICO. HACIA UNOS PRINCIPIOS UNIVERSALES DE LA IA. LÍMITES ÉTICOS
- 3. AUTORÍA DE OBRAS REALIZADAS CON IA.
- 4. RIESGOS EN EL USO DE LA IA POR ENTES PRIVADOS Y PÚBLICOS
 - 4.1 ANÁLISIS PRÁCTICO DE CONFLICTOS CON LA IA.
 - 4.1.1 LA IA DENTRO DEL PROCESO JUDICIAL: HERRAMIENTA NO SUSTITUTIVA DE LA LABOR DE LAS PARTES
- 5. CONCLUSIONES
- BIBLIOGRAFÍA

CAPÍTULO 13. CUESTIONES DE PROPIEDAD INTELECTUAL E INDUSTRIAL VINCULADAS A LA INTELIGENCIA ARTIFICIAL. ESPECIAL MENCIÓN AL CAMPO DE LA SALUD, *Juan Emilio Suñé Cano*.....pp. 347-388
375

- INTRODUCCIÓN
- 1. APROXIMACIÓN AL CONCEPTO DE PROPIEDAD INTELECTUAL E INDUSTRIAL
- 2. LA INTELIGENCIA ARTIFICIAL, CONCEPTOS Y ESPECIFICACIONES DE LA INTELIGENCIA ARTIFICIAL GENERATIVA
 - 2.1. Tipos de inteligencia artificial
 - 2.2. El entrenamiento y alimentación de la Inteligencia Artificial

2.3. Redes Generativas Antagónicas

3. INTELIGENCIA ARTIFICIAL GENERATIVA Y EL RÉGIMEN JURÍDICO DE LA PROPIEDAD INTELECTUAL (DERECHO DE AUTOR)

- 3.1. El principio antropocéntrico de la autoría: núcleo dogmático del sistema
- 3.2. La naturaleza jurídica de los prompts expresión, idea o instrucción técnica
- 3.3. Obras generadas por IA: originalidad, reproducción y riesgos jurídicos
- 3.4. Autoría y titularidad de derechos en obras generadas o asistidas por IA
- 3.5. La Respuesta de las legislaciones de Propiedad Intelectual al Desarrollo de la IA

4. ALGUNOS ASPECTOS DE PROPIEDAD INDUSTRIAL E INTELIGENCIA ARTIFICIAL

5. INTELIGENCIA ARTIFICIAL EN EL ÁMBITO DE LA SALUD, LA BALANZA DE RIESGOS Y BENEFICIOS

- 5.1. Oportunidades Clave en el Sistema Sanitario
- 5.2. Aceleración de la Investigación y el Desarrollo de Vacunas
- 5.3. Riesgos y Desafíos Críticos en la Implementación de la IA
- 5.4. IA para la Aceptación Pública y la Comunicación
- 5.5. IA en el ámbito radiológico

6. CONCLUSIONES

APÉNDICE BIBLIOGRÁFICO



CAPÍTULO 1.
DERECHOS HUMANOS ANTE EL RETO
DE LA INTELIGENCIA ARTIFICIAL¹

GABRIELA C. COBO DEL ROSAL PÉREZ²

Profesora Titular de Historia del Derecho
Universidad Rey Juan Carlos

Sumario: INTRODUCCIÓN 1. LA INTELIGENCIA ARTIFICIAL COMO “CIRCUNSTANCIA” EN OCCIDENTE 2. LOS DERECHOS HUMANOS ANTE EL RETO DE LA I.A.: UNA APROXIMACIÓN MULTIDISCIPLINAR 3. REFLEXIÓN FINAL 4. A MODO DE CONCLUSIÓN BIBLIOGRAFÍA

“Vivir es sentirse fatalmente forzado a ejercitar la libertad”³

INTRODUCCIÓN

Se acude con frecuencia a la célebre afirmación cuya autoría sobra mencionar “yo soy yo y mi circunstancia” para excusarse de alguna actitud o fracaso ante una

¹ La presente obra se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos. Es el resultado conjunto de los trabajos que como Grupo vieron la luz sus primeros resultados en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, desarrollados mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

² Profesora Titular de Historia del Derecho y de las Instituciones Españolas en la Universidad Rey Juan Carlos. Doctora en Derecho. Cuenta con numerosas estancias de formación académica, docentes y de investigación en prestigiosas universidades como la Universidad Karl-Franzens Graz; la Universidad Degli Studi di Milano; la Universidad de Harvard o la Universidad San Marcos. Ha escrito numerosos trabajos al publicados en editoriales y revistas científicas de reconocido prestigio. orcid.org/0000-0001-8515-4743. gabriela.cobodelrosal@urjc.es

³ Ortega y Gasset, J., *La rebelión de las masas*, Prólogo de Julián Marías, (1929), ePUBByrm, 2021, p. 48.

situación. Tal frecuente cita normalmente omite su final, tal vez intencionadamente o por desconocimiento, pues la cita completa en realidad termina de diciendo, “y si no

la salvo a ella no me salvo”. En el sentido que aquí nos ocupa cabe decir que la circunstancia que plantea el reto de la Inteligencia Artificial (I.A.) es una de las que más nos debería unir como sociedad en la actitud que frente a ella quisiéramos tener. Ello resulta evidente si nos fijamos, por ejemplo, en la transversalidad de esta herramienta o en su vertiginosa y, en muchos casos, silente velocidad, no solamente en su capacidad de respuesta sino también en el desarrollo y evolución de la misma, especialmente la generativa⁴. La aplicación de la I.A. evidentemente afecta al conjunto global de toda la humanidad, sin embargo, estas breves páginas quedan circunscritas a lo que se entiende por forma de vida occidental, la cual por otra parte afecta en mayor o menor medida a todos los continentes. La velocidad del desarrollo de la I.A. nos recuerda que es urgente una puesta en común y una respuesta determinada en torno a la dirección que en Occidente queremos tomar como sociedad, ello a pesar de que la Unión Europea, -reflejo institucional de lo que entendemos por cultura occidental-, camina más lentamente que China o Estados Unidos especialmente en lo que concierne a su desarrollo⁵. Probablemente nos enfrentamos con un nuevo paradigma en materia de protección de los derechos fundamentales⁶. A tal exacto el pensamiento orteguiano plasmado en su casi centenaria obra, *La rebelión de las masas*, contrario a resultar obsoleto parece ser una lectura apropiada como hilo conductor a la hora de ordenar algunas ideas en torno a esta cuestión, pues en sus reflexiones Ortega y Gasset se adelantó a muchos males que el uso incorrecto de I.A. nos puede acarrear. De su magnífica obra se colige que parte de la solución descansa en la educación entendida en un sentido amplio, educación en valores y diálogo entre las ciencias y, sin duda, el diálogo entre la ciencia y la técnica.

Ratzinger en una de sus publicaciones sobre el problema de identidad que atraviesa Europa se preguntaba si Occidente formaba parte de la “civilización de la técnica”, la civilización “del comercio” o incluso si asistíamos al “final de las culturas europeas”⁷. Han pasado más de veinte años desde que el Papa Benedicto XVI formulase aquellas cuestiones y todo parece apuntar que la I.A. alcanza a Occidente sin haber resuelto, al menos del todo, la crisis de identidad de Europa y por tanto de la

⁴ Recoge en síntesis una evolución de los principales ámbitos en los que se ha ido incorporando y una relación bibliográfica de interés. De los Ríos Ruíz, A., “Derechos Humanos e Inteligencia Artificial”, *Revista de Derecho, Empresa y Sociedad*, N°25, 2025, pp. 179-187.

⁵ Vid. Gascón Marcén, A., “Derechos humanos e Inteligencia Artificial”, VV.AA., *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*, Centro de Estudios Políticos Constitucionales, Madrid: 2020, pp. 335-350, p. 348.

⁶ Vid., Sarrión Esteve, J., “El Derecho constitucional en la era de la Inteligencia Artificial, los robots y los drones”, VV.AA., *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*, Centro de Estudios Políticos Constitucionales, Madrid: 2020, pp. 321-333, p. 328.

⁷ Ratzinger, J., «Europa, sus fundamentos espirituales ayer, hoy y mañana», *Seminarios sobre los Ministerios de la Iglesia*, vol. 51, 177, (2005), pp. 416-429, p. 423. Disponible en <https://www.seminariosdigital.es/index.php/RevistaSeminarios/article/view/706> También en Europa. Raíces, identidad y misión, Traducción Pablo Largo, Madrid 2004, pp. 9-33

civilización que vive a la hora occidental⁸. En este sentido parece que lo dicho al respecto en el pasado siglo por Ortega tiene vigencia todavía:

“Los principios en que se apoya el mundo civilizado —el que hay que sostener— no existen para el hombre medio actual. No le interesan los valores fundamentales de la cultura, no se hace solidario de ellos, no está dispuesto a ponerse en su servicio”⁹.

Julían Marías, discípulo principal del referenciado filósofo, al tratar de allanar al lector el camino para comprender la *La rebelión de las masas* aclaraba que su maestro entendía que:

“[...] el gran título de honor de nuestro tiempo, en los países occidentales, es el acceso de las masas a la vida histórica, al goce de las creaciones de la civilización, a las posibilidades humanas que más de dos milenios de esfuerzo han hecho posibles. Lo grave, lo inquietante, lo patológico no es eso, sino que eso se comprometa y ponga en peligro al no tener en cuenta que en el hombre todo es inseguro y problemático. No se puede vivir humanamente más que exigiéndose, manteniendo una tensión creadora, un estado de permanente alerta, un impulso hacia lo alto”¹⁰.

Y es que la realidad social a la que nos referimos no puede entenderse aislada de su Historia o de las corrientes de pensamiento imperantes en la sociedad. La fortaleza y la fragilidad de las democracias de Occidente descansan en la educación. La democracia quiebra y se resiente solamente con una voz más alta que interrumpa un diálogo de forma abrupta. En suma, la fortaleza del sistema que plantea la democracia descansa en su fragilidad. Es comparable a un cristal de bohemia cuya belleza en parte reside en su delicadeza la cual exige un cuidado y silencio reverentes. Occidente vive en democracia, es decir, en un sistema cuya fragilidad al igual que el cristal de bohemia reclama del silencio, de la lentitud imprescindible que acompasa el diálogo y del celoso y atento cuidado de las instituciones que acogen este sistema. De todo ello agudamente alertaba Ortega y Gasset al decir que:

“La civilización no está ahí, no se sostiene a sí misma. Es artificio y requiere un artista o artesano. Si usted quiere aprovecharse de las ventajas de la civilización, pero no se preocupa usted de sostener la civilización [...] se ha fastidiado usted. En un dos por tres se queda usted sin civilización. ¡Un descuido, y cuando mira usted en derredor, todo se ha volatilizado! Como si hubiese recogido unos tapices que tapaban la pura naturaleza, reaparece repristinada la selva primitiva. La selva siempre es primitiva. Y viceversa: todo lo primitivo es selva”¹¹.

⁸ De lo cual hemos tratado de ocuparnos modestamente en “Consideraciones en torno a las tres colinas que conformaron la identidad de Europa”, *Vergentis. Revista de Investigación de la Cátedra Internacional Con-junta Inocencio III*, N° 16, enero-junio 2023, pp. 19-41.

⁹ Ortega y Gasset, J., *La rebelión...*, p. 84.

¹⁰ Ortega y Gasset, J., *La rebelión...* p. 10.

¹¹ *Ibid.*, p.

La irrupción de la I.A. puede modificar en grave medida muchos de los pilares en los que se sustenta nuestra forma de vida conocida como “occidental”. La presente reflexión aquí compartida trata de ofrecer elementos útiles desde la ciencia para frenar ahora con más urgencia que antes, la “rebelión de las masas” preconizada por Ortega el pasado siglo, que ahora encuentra un vehículo increíblemente más eficaz para aumentar su capacidad destructiva. Una rebelión que, como aclaraba Julián Marías, no es una rebelión social contra un tirano, sino una rebelión del hombre “contra su propia condición y función” contra “una inautenticidad, una falsedad, una enfermedad social”¹². Pues para el referido filósofo:

“El hombre-masa es el hombre cuya vida carece de proyectos y va a la deriva. Por eso no construye nada, aunque sus posibilidades, sus poderes, sean enormes”¹³.

Este hombre masa para Ortega hace un siglo era el que tomaba las decisiones de su tiempo. ¿Es acaso el mismo que toma las decisiones del actual? De ser así, Ortega, le vaticinaba un corto recorrido:

“Si ese tipo humano sigue dueño de Europa y es, definitivamente, quien decide, bastarán treinta años para que nuestro continente retroceda a la barbarie. Las técnicas jurídicas y materiales se volatilizarán con la misma facilidad con que se han perdido tantas veces secretos de fabricación. La vida toda se contraerá. La actual abundancia de posibilidades se convertirá en efectiva mengua, escasez, impotencia angustiosa; en verdadera decadencia. Porque la rebelión de las masas es una misma cosa con lo que Rathenau llamaba «la invasión vertical de los bárbaros»

Importa, pues, mucho conocer a fondo a este hombre-masa, que es pura potencia del mayor bien y del mayor mal”¹⁴.

En este sentido y sin querer recurrir al alarmismo, hacemos nuestras las palabras del referido Ortega ya centenarias, quien sin negar la realidad del progreso entendía que ningún progreso es seguro y que ninguna evolución estaba exenta de la amenaza de la involución y del retroceso, por entender que en la Historia todo es posible, tanto el “progreso triunfal e indefinido” como la “periódica regresión”. Agudamente afirmaba que:

“La civilización, cuanto más avanza, se hace más compleja y más difícil. Los problemas que hoy plantea son archiintrincados. Cada vez es menor el número de personas cuya mente está a la altura de los problemas. La posguerra nos ofrece un ejemplo bien claro de ello. La reconstrucción de Europa —se va viendo— es un asunto demasiado algebraico, y el europeo vulgar se revela inferior a tan sutil empresa. No es que falten medios para la solución. Faltan cabezas. Más exactamente: hay algunas cabezas, muy

¹² Ibid., p. 10.

¹³ Ibid., p. 50

¹⁴ Ibid., p. 53.

pocas, pero el cuerpo vulgar de la Europa central no quiere ponérselas sobre los hombros”¹⁵.

1. REFLEXIONES EN TORNO A LOS DERECHOS HUMANOS ANTE EL RETO DE LA INTELIGENCIA ARTIFICIAL

Antes de que estallase la Segunda Guerra Mundial Ortega y Gasset alertaba acerca de los comunismos y de los fascismos, con su visionaria forma de observar el mundo. Afirmaba ya entonces que eran movimientos típicos del hombre-masa¹⁶. También aseguraba que las naciones europeas eran insuficientes para alcanzar sus metas y resolver sus problemas ya que éstos traspasaban las fronteras y solamente unidas podríamos resolverlos. Por citar otro ejemplo de la vigencia de sus predicciones, aseguraba que lo que acontecía en Europa respecto al acceso de las masas al poder, ya aconteció en Roma en tiempos de Ortega, esto es, aproximadamente cien años atrás. Y es que la época de la decadencia romana fue la época de los edificios colosales, la época en la que se terminaron por subvertir los órdenes y las minorías gobernantes fueron sustituidas por las masas que se colocaron en su lugar¹⁷. Sobradas son pues las muestras de los aciertos del citado autor en sus visionarias predicciones. Reconsiderar sus reflexiones puede ser útil para ponderar con mayor tino los retos a que se enfrenta Occidente con la I.A. Actualmente, se suma al distanciamiento respecto de su Historia y de los factores que cohesionaron su forma de vida, la desconfianza en sus dirigentes y en su propia inteligencia. Es por ello imprescindible que al menos desde la Universidad se trate de recuperar la confianza en esta última: la razón, en este sentido pueden recuperarse las palabras de Julián Marías en su *Aquí y ahora*:

“Los hombres de Occidente volverán a confiar en la inteligencia, que ha sido su gran fuerza milenaria, en cuanto la vean funcionar, es decir, aplicarse a la faena de dar razón a las cosas. Necesitan ver trabajar, con fruición y sin gestos, a los que tienen como misión buscar la verdad; y se sentirán asociados a su esfuerzo y llenos de ánimo aún en medio de mayores dificultades, siempre que tengan conciencia de que se está luchando con ellas y de que en esa lucha se emplea a fondo el instrumento con que es dado al hombre arrancar a la realidad su secreto: la razón”¹⁸.

Y es que si nos mantenemos desconfiamos de la razón tenemos más propensión a hacer un peor uso de la I.A. Alertaba Ortega respecto a la ciencia y a la técnica aclaraba ya entonces que es en la ciencia Europea en la que se ha asentado Occidente aclarando que:

¹⁵ Ibid., p. 85

¹⁶ Ibid., p. 87

¹⁷ Ibid., p. 23.

¹⁸ Marías, J., “La autoridad intelectual”, *Aquí y ahora*, Buenos Aires, 1954, p. 1948.

“No toda técnica es científica. El que fabricó las hachas de sílex, en el período che-lense, carecía de ciencia y, sin embargo, creó una técnica ... sólo la técnica moderna de Europa tiene una raíz científica, y de esa raíz le viene su carácter específico, la posibilidad de un ilimitado progreso. Las demás técnicas —mesopotámica, nilota,

griega, romana, oriental— se estiran hasta un punto de desarrollo que no pueden sobrepasar, y apenas lo tocan comienzan a retroceder en lamentable involución”¹⁹.

Y sobre ello advertía además que:

“[...] resulta que el hombre de ciencia actual es el prototipo del hombre-masa. Y no por casualidad, ni por defecto unipersonal de cada hombre de ciencia, sino porque la ciencia misma —raíz de la civilización— lo convierte automáticamente en hombre-masa; es decir, hace de él un primitivo, un bárbaro moderno”²⁰.

La nueva realidad en la que se sitúan los derechos fundamentales ante la I.A. obliga a poner al día todos los mecanismos jurídicos que ha informado su conformación original y su preservación a fin de que los avances propuestos por el nuevo paradigma tengan puntos de referencia. La irrupción de la I.A., especialmente de la I.A. generativa en la esfera de la protección de los derechos fundamentales, puede afectar a aspectos consustanciales a la naturaleza de algunos de ellos como los que se desprenden de la propiedad intelectual en relación con los derechos de autor. A la luz del grado de afectación de la presencia de la I.A. en la dimensión social y, más específicamente, en la protección de los derechos fundamentales, se propone este Grupo de Investigación el análisis crítico de la normativa de aplicación en materia de I.A. y de las instituciones jurídicas implicadas desde una perspectiva multidisciplinar. Ello lleva implícito el estudio de los derechos fundamentales afectados en su conceptualización, desarrollo y delimitación. El análisis crítico de la normativa vigente referida a la I.A., sin alejarlo de la real interpretación de la jurisprudencia en la medida de lo posible a la hora de delinear las fronteras de los derechos fundamentales afectados por el nuevo paradigma creado por la I.A. y las nuevas tecnologías y la revisión de la doctrina, cuando menos permite atisbar la previsibilidad de las líneas delimitadoras de los derechos fundamentales en su protección, analizando su expansión, contracción o posible colisión, contribuyendo de este modo a la seguridad al sistema jurídico.

¹⁹ Ibid., p. 101

²⁰ Ibid., p. 101.

2. LOS DERECHOS HUMANOS ANTE EL RETO DE LA I.A.: UNA APROXIMACIÓN MULTIDISCIPLINAR

Los siguientes trabajos aquí recogidos ofrecen una aproximación multidisciplinar al tema central: la protección de los derechos fundamentales ante el reto de la I.A. Abordar una investigación sobre la temática planteada y desde la perspectiva multidisciplinar propuesta, precisa un previo y profundo conocimiento del Derecho vigente y aplicado que a su vez reclama una visión más amplia del conocimiento científico que precisa el diálogo permanente entre las distintas ramas del saber. Habida cuenta de ello la mayor parte de los investigadores provienen de disciplinas jurídicas de Derecho “positivo”, sin embargo, no exclusivamente. Ello por entender que es preciso para el jurista del Derecho positivo conocer la realidad vigente a la luz de la Filosofía. Pues algunas precisiones conceptuales y planteamientos de base deben nutrirse necesariamente desde las ramas del Derecho “no positivo” para poder adelantarnos a las consecuencias no deseadas que la norma puede traer, especialmente cuando se considera ésta al margen de su Historia o al margen de la corriente filosófica que de fondo le inspira. El uso de estas ramas del saber jurídico, resulta preciso cuando no imprescindible para describir la realidad social que acoge el gran cambio que supone el reto de la I.A. y las nuevas tecnologías²¹. Es por todo ello que se reúnen aquí áreas como la Historia del Derecho; Filosofía del Derecho; Derecho Civil; Derecho Internacional Público o la Criminología.

La presente obra colectiva queda compuesta por trece trabajos que se presentan ordenados alfabéticamente desde el índice general. En dichos trabajos los autores se responsabilizan individualmente de las consideraciones y criterios que en ellas recogen. Desde distintas ramas jurídicas y ramas afines como la Criminología centran su atención en el reto de la I.A. a fin de prevenir, alertar o, en su caso, ofrecer soluciones a las posibles dificultades ante las que el Derecho se encuentra.

M^a Ángeles Cano Linares es Profesora Titular de Derecho Internacional Público en la Universidad Rey Juan Carlos. Contribuye a la presente obra colectiva con el Capítulo Segundo titulado “Cooperación internacional institucionalizada frente a los riesgos para la democracia, la seguridad y los derechos humanos derivados del uso de sistemas de inteligencia artificial”. Aborda el tema de la vertiginosa irrupción de los sistemas de I.A., previniendo acerca de las nuevas e importantes amenazas que su uso inadecuado conlleva y alerta acerca de los graves dilemas éticos que plantea su uso. Propone como rumbo a seguir una inexorable respuesta determinada y “sin fisuras”, para lo cual reclama como imprescindible una estrecha cooperación internacional

²¹ Ofrece una visión panorámica de los derechos humanos que pueden verse afectados por la Gascón Marcén, A., “Derechos humanos e Inteligencia Artificial”, VV.AA., *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*, Centro de Estudios Políticos Constitucionales, Madrid: 2020, pp. 336 y ss.I.A. así como de la reacción de la comunidad internacional

orientada al mantenimiento de la paz, la democracia, así como el estado de derecho y fundamentalmente el respeto de los derechos humanos.

Angélica María Castro Acosta es Profesora de Derecho Constitucional en la Universidad Europea de Madrid. Participa con el Capítulo Tercero titulado “Sesgos algorítmicos y derechos humanos: desafíos éticos y regulatorios en la era de la inteligencia artificial”. Propone una aclaración de determinados conceptos referidos a los sesgos algorítmicos tras analizar su impacto en la protección de derechos humanos. Ofrece un estudio de algunos casos, así como de los marcos regulatorios internacionales, -EU AI Act y las directrices de UNESCO-, con la intención anunciada por la autora de concluir con una propuesta de protocolo capaz de integrar mecanismos que garanticen el desarrollo de la I.A. de un modo transparente, auditable y ético.

Ana Isabel Carreras Presencio es Profesora Permanente Laboral en el Área de Derecho Internacional Público de la Universidad Rey Juan Carlos. Participa con el Capítulo Cuarto titulado “Posición de casilla “d8”: supervisión humana”. Alerta con ejemplos acerca de los riesgos de los sesgos en la I.A. y de cómo éstos pueden atentar contra el derecho a la igualdad y a la no discriminación. Plantea la necesidad de una educación en I.A. así como un control humano directo en la respuesta que proporcione la I.A., incidiendo en la importancia de esta vigilancia humana recordando en todo momento que no es más que una herramienta tecnológica.

Irene Blanca Cerrillo Torquemada es abogada en el Área de Estrategia, Seguimiento y Transformación de los Servicios Jurídicos de Banco Bilbao Vizcaya Argentaria SA. Participa con el Capítulo Quinto titulado “Abogacía e Inteligencia Artificial: algunas notas sobre las oportunidades y desafíos para la profesión”. En su trabajo de marcado carácter pragmático aborda cómo el vigente Derecho en su aplicación directa es receptor de los modelos de lenguaje grandes (LLM, siglas en inglés). Dichos sistemas, capaces de generar textos de directa utilidad para el tráfico jurídico y de ofrecer respuestas concretas sobre asuntos jurídicos precisos tales como análisis de normas o de resoluciones, ofrecen una serie de ventajas e inconvenientes sobre los que la autora comparte sus reflexiones en torno a las consecuencias que el uso o el abuso de tales herramientas puede acarrear a los ciudadanos en su acceso a la justicia. Alerta en sus páginas y desde la práctica jurídica acerca del potencial “cambio de paradigma” en la aplicación del Derecho. Considera que si bien el abogado que logre un mejor dominio de la I.A. gozará de una mayor ventaja diferencial no parece muy próxima la posibilidad real de delegar en la I.A. “la tarea del asesoramiento o defensa jurídica”, pues entiende que la verdadera diferencia, “seguirá radicando en el profesional humano, con una mayor base de experiencias vitales y mejor capacidad de raciocinio jurídico”.

Cristina Domingo Jaramillo es Investigadora Postdoctoral en el Departamento de Derecho Penal de la Universidad de Granada. Participa con el Capítulo Sexto titulado “Identificación biométrica remota en el ámbito policial, ¿Cómo afecta a la protección de datos de carácter personal?”. Estudia el Reglamento Europeo de Inteligencia

Artificial en su regulación en torno a la regulación que propone en el ámbito de los sistemas de identificación biométrica remota en espacios públicos. Dicho asunto afecta principalmente a los derechos y libertades vinculados a la protección de datos de carácter personal. En su estudio la autora estudia algunos aspectos especialmente controvertidos respecto a los ataques que precisamente a este derecho puede traer el uso de los sistemas de identificación biométrica remota, y alerta que el impacto puede llegar a rayar lo excesivo por lo que considera que tal vez no debería haberse incluido en el citado Reglamento.

Elena Pineros Polo es Profesora Permanente Laboral en el Área de Derecho Procesal de la Universidad Rey Juan Carlos. Participa con el Capítulo Séptimo titulado “Entre la legalidad y el algoritmo: principios procesales e Inteligencia Artificial en la ejecución de medidas del proceso penal de menores”. La autora estudia la ejecución de medidas en el proceso penal de menores fase procesal que termina por evidenciar el propósito reeducativo del sistema. El trabajo analiza la Ley Orgánica 5/2000 y su desarrollo reglamentario especialmente centrándose en las medidas que propone esta norma. Ofrece un estudio en torno a la distribución de las diferentes competencias así como de los principales principios procesales que informan la ejecución. Destina una parte final al estudio del posible uso de sistemas de I.A. como medidas apoyo alertando acerca de los riesgos asociados a la protección de datos, sesgos, posible automatismo decisorio, ofreciendo soluciones compatibles garantistas en las que con el control humano prevalece en la decisión judicial.

Aurelio de Prada García es Profesor Titular de Filosofía del Derecho en la Universidad Rey Juan Carlos. Participa con el Capítulo Octavo titulado “Justicia en plazo, sin dilaciones, e Inteligencia Artificial: algunas propuestas”. En las páginas que componen su capítulo completa el análisis realizado en otros trabajos previos del autor que parten todos de un caso real en el que presuntamente se habría violado el Derecho a la tutela efectiva. De su análisis en este trabajo concluye, utilizando las palabras del autor, que: “si bien sería de gran ayuda, la implementación de la IA en la administración de justicia, tardanza incluida, no sería sin embargo suficiente para corregir todos los defectos que se siguen de ese análisis. Habría que combinar el uso de la IA con un replanteamiento general de la consideración de los juristas y del propio derecho, incluyendo su consideración peyorativa y, por tanto, un replanteamiento de la enseñanza jurídica, así como del acceso a la carrera judicial para el que sería requisito sine qua non el haber sido parte en un juicio”.

Fernando Gil González es Profesor Ayudante Doctor de Historia del Derecho en la Universidad de Alcalá de Henares. Participa con el Capítulo Noveno titulado “Geopolítica, gobernanza e Inteligencia Artificial”. Estudia la capacidad de la I.A. a la hora de deformar la realidad política, incidiendo en la importancia del estudio a fondo de la geopolítica para una mejor gobernanza. Con perspectiva histórica se aproxima a las estructuras de gobernanza planetaria. Destina una parte final a estudiar valoraciones de futuro en determinadas prácticas y procesos de gobernanza pública en los que la I.A. tenga parte.

María Concepción Rayón Ballesteros es Profesora de Derecho Procesal (acreditada para Prof. Titular) Facultad de Derecho de la Universidad Complutense de Madrid. Participa con el Capítulo Décimo titulado “Inteligencia Artificial, robotización de la justicia: especial referencia a como se ven afectados los derechos humanos y las garantías procesales”. En su trabajo estudia el impacto de la toma de decisiones automatizadas y de los sistemas algorítmicos que apoyan al juzgador. Analiza los conceptos de I.A. aplicada a la justicia y de “justicia robotizada” y cómo estas tecnologías pueden afectar a los derechos fundamentales. Así mismo también alerta acerca de los posibles riesgos de opacidad algorítmica, sesgos, falta de control humano en la función jurisdiccional, los cuales estudia también. Destina una parte final a la propuesta de soluciones normativas y prácticas para garantizar que la incorporación de la I.A. a la justicia se realice de un modo garantista con el Estado de Derecho.

Pedro Ródenas Cortés es Profesor de Derecho Civil de la Universidad de Extremadura. Participa con el Capítulo Undécimo titulado “Derechos fundamentales del art. 20. 1, a) y d) ce: impacto de la inteligencia artificial” centrando su estudio en el potencial alcance de la I.A. al ejercicio del derecho fundamental de la libertad de expresión y de información contenido en el artículo 20 de la Constitución Española. En su trabajo evidencia que la I.A. es capaz de operar como un nuevo intermediario informativo con una legítima intervención siempre y cuando garantice la supervisión humana y la protección de los derechos fundamentales. También se ocupa del impacto de la I.A. en la formación de la opinión pública, con especial énfasis en las generaciones más jóvenes, quienes recurren a ella como herramienta informativa y soporte afectivo. Concluye subrayando que el vínculo entre la IA y la libertad de expresión precisa una regulación más robusta en armonía con el con la protección de los derechos fundamentales y el compromiso conjunto del Estado y la ciudadanía.

Alejandro Rosillo Fairén es Profesor Titular de Derecho Civil de la Universidad Europea. Participa con el Capítulo Decimosegundo titulado “Impacto de la inteligencia artificial en el ámbito de la propiedad intelectual, tensiones y desafíos”. En su trabajo realiza un estudio crítico que gira en torno a la evolución de la I.A. y su impacto en la propiedad intelectual. En sus páginas con visión crítica alerta acerca de los riesgos de su aplicación indebida como la destrucción masiva y la falta de ética en la programación y uso de la misma. Entre otras medidas propone establecer controles en su desarrollo capaces de fomentar un uso responsable sin caer en la censura.

Juan Emilio Suñé Cano es Profesor Asociado de Derecho Internacional Público en la Universidad Complutense de Madrid. Participa con el Capítulo Decimotercero titulado “Cuestiones de Propiedad Intelectual e Industrial vinculadas a la Inteligencia Artificial. Especial mención al campo de la salud”. En su estudio alerta sobre los nuevos retos jurídicos que plantea la I.A. especialmente la IA generativa en el ámbito de la Propiedad Intelectual principalmente por haber sido éste un campo radicado en la autoría humana que ahora se enfrenta a tecnologías capaces de generar contenidos de forma autónoma. Dichos contenidos de íntegra creación tecnológica al carecer de autoría legal quedan en el dominio público, de tal modo que como analiza el autor, nociones como originalidad y autoría entran en jaque así como el uso de datos protegidos en su entrenamiento. Trata también el ámbito de la Propiedad Industrial y el

sector sanitario donde si bien constata los beneficios incuestionables de la correcta aplicación de la I.A. alerta con espíritu crítico acerca de los posibles riesgos de sesgo, privacidad y responsabilidad.

REFLEXIÓN FINAL A MODO DE CONCLUSIÓN

En la base de los problemas frente a los riesgos de una mala praxis en el uso de la I.A. se encuentran al menos tres elementos que interactúan entre sí: la falta de conexión, conocimiento y comprensión con las raíces históricas que cohesionaron y conformaron la cultura conocida como occidental; el triunfo de las masas en la dirección de los Estados y la pérdida en Occidente de la esperanza y de la confianza en la inteligencia para dar respuesta a las preguntas existenciales más trascendentes para el hombre.

En suma, Occidente necesita recuperar la confianza en la inteligencia humana: la razón. Ello es especialmente importante para desarrollar un uso adecuado de la I.A. Para ello es imprescindible que la razón en Occidente renueve su compromiso con la búsqueda de la verdad y se enfrente con ilusión a las preguntas verdaderamente importantes que afectan al sentido de la existencia humana sin perder por el camino la esperanza.

BIBLIOGRAFÍA

De los Ríos Ruíz, A., “Derechos Humanos e Inteligencia Artificial”, *Revista de Derecho, Empresa y Sociedad*, N°25, 2025, pp. 179-187.

Gascón Marcén, A., “Derechos humanos e Inteligencia Artificial”, VV.AA., *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*, Centro de Estudios Políticos Constitucionales: Madrid, 2020, pp. 335-350.

Marías, J., *Aquí y ahora*, Colección Austral: Buenos Aires, 1954.

Ortega y Gasset, José, *La rebelión de las masas*, 1929, Prólogo de Julián Marías, ePUByrm, 2021.

Sarrión Esteve, J., “El Derecho constitucional en la era de la Inteligencia Artificial, los robots y los drones”, VV.AA., *Setenta años de Constitución Italiana y cuarenta años de Constitución Española*, Centro de Estudios Políticos Constitucionales, Madrid: 2020, pp. 321-333.



CAPÍTULO 2

COOPERACION INTERNACIONAL INSTITUCIONALIZADA INSTITUCIONALIZADA FRENTE A LOS RIESGOS PARA LA DEMO- CRACIA, LA SEGURIDAD Y LOS DERECHOS HUMANOS DERIVA- DOS DEL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL²²

M^a ÁNGELES CANO LINARES

Profesora Derecho Internacional público y RR.II.²³

Universidad Rey Juan Carlos

angeles.cano.linares@urjc.es

Resumen

El rápido auge de los sistemas de inteligencia artificial (IA) ha generado nuevas oportunidades potencialmente a nivel global en prácticamente todos los ámbitos. Junto a las ventajas derivadas de desarrollos de sistemas de IA, estos rápidos cambios plantean importantes nuevos riesgos y amenazas y profundos dilemas éticos, haciendo surgir la necesidad imperiosa de responder a ellos con una respuesta sin fisuras, lo que requiere de una estrecha cooperación internacional para preservar la paz, la democracia, el estado de derecho y el respeto de los derechos humanos.

Palabras clave

Sistemas de inteligencia artificial; derechos humanos, neuro tecnologías, amenazas híbridas, armas autónomas, ciberataques, brecha digital, ODS, neuro tecnologías, biometría.

²² El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ca8b86cec8e8/content>

²³ Miembro del Grupo de investigación de alto rendimiento en Libertad, Seguridad y Ciudadanía en el Orden Internacional y del Grupo de investigación de alto rendimiento en Inmigración y Gestión de la Diversidad Cultural, ambos de la Universidad Rey Juan Carlos. Este trabajo ha sido realizado en el marco del “Vacíos normativos y desarrollo progresivo de la Agenda 2030 y del principio de sostenibilidad. Especial relevancia para España”(PID 2022-13833390B-I00). ORCID: 0000-0002-1440-5880. Scopus ID: 57199672283.

Abstract

The rapid rise of artificial intelligence (AI) systems has created new opportunities with global potential in every field. Alongside the advantages derived from developments in AI systems, these rapid changes pose significant new risks and threats and profound ethical dilemmas, giving rise to an urgent need to respond to them with a seamless response, which requires close international cooperation to preserve peace, democracy, the rule of law and respect for human rights.

Key words

Artificial intelligence systems; human rights; neurotechnology; hybrid threats; autonomous weapons; cyberattacks, digital divide, SDG, biometrics.

Sumario: INTRODUCCIÓN 1- PRINCIPALES RIESGOS DERIVADOS DEL USO/MAL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL 1.1 El uso malicioso de los sistemas IA. Ciberdelitos y Ciberseguridad 1.2 Su incidencia en el estado de derecho: derechos humanos, democracia, brecha digital y gobernanza. 1.3 Las neuro tecnologías: Revolución en ciernes 2. PROPUESTAS EN EL AMBITO UNIVERSAL: PACTO DIGITAL MUNDIAL Y ÉTICA EN LA INTELIGENCIA ARTIFICIAL 2.1 Naciones Unidas: Pacto Digital Mundial 2.2. La labor de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura y su enfoque ético 3. RIESGOS ESPECIFICOS EN CASO DE CONFLICTO ARMADO 3.1 Amenazas híbridas, conflictos cibernéticos 3.2. Inteligencia Artificial y Derecho Internacional Humanitario 3.3 La preocupación del Consejo de Seguridad de Naciones Unidas 4. LA DEFENSA DE LA DEMOCRACIA Y EL ESTADO DE DERECHO: CONVENIO MARCO DEL CONSEJO DE EUROPA 5. CONCLUSIONES APÉNDICE BIBLIOGRÁFICO

INTRODUCCIÓN

El rápido auge de los sistemas de inteligencia artificial (IA) ha generado nuevas oportunidades potencialmente a nivel global en prácticamente todos los ámbitos. Desde la automatización de tareas rutinarias, y la optimización de procesos productivos a los robots inteligentes para tareas repetitivas, de las plataformas adaptativas que personalizan el aprendizaje y las herramientas de traducción y accesibilidad al monitoreo de cultivos con sensores y drones y la predicción de plagas y optimización de riego; desde la detección temprana de enfermedades al desarrollo de medicamentos mediante simulaciones, de los vehículos autónomos a los sistemas inteligentes de tráfico, los ejemplos se multiplican.

Sin embargo, junto a las ventajas derivadas de la automatización de tareas, la reducción de costos, el aumento de la eficiencia en procesos repetitivos, la potencial mejora en la toma de decisiones, el análisis de grandes volúmenes de datos, la personalización con experiencias adaptadas en educación, marketing y salud y la disponibilidad 24/7, estos rápidos cambios derivados de desarrollos de sistemas de IA también

plantean profundos dilemas éticos y no menos importantes nuevos riesgos que surgen del potencial que tienen los sistemas basados en IA para reproducir prejuicios, contribuir a la degradación del clima y amenazar los derechos humanos, entre otros.

Estos riesgos asociados a la IA se suman a las desigualdades ya existentes, aumentándolas aún más en perjuicio de todos los históricamente marginados. Sin embargo, también los más afortunados, aquellos que viven en sociedades desarrolladas en las que se puede disfrutar de las ventajas derivadas de la IA, se enfrentan a nuevos riesgos y desafíos que afectan tanto al ámbito laboral como al personal y al más íntimo del ser humano. Así, la sustitución de tareas humanas por automatización comporta la pérdida de empleos, los algoritmos entrenados con datos sesgados pueden generar decisiones injustas ampliando los sesgos y la discriminación, el uso indebido de datos personales y la vulnerabilidad frente a ciberataques afecta a nuestra privacidad y seguridad. Además, se ha incrementado nuestra dependencia tecnológica, con la reducción de habilidades humanas y excesiva confianza en los sistemas por no a mencionar el impacto ético y legal y los problemas generados por la falta de regulación clara sobre responsabilidad y transparencia. Un ejemplo de especial relevancia es la generalización del uso de la biometría en esta era de la inteligencia artificial con, por ejemplo, la posibilidad de identificación biométrica a distancia, en la que se incluye, el reconocimiento facial, la categorización biométrica o el reconocimiento de emociones.

Y aún más alarmante es el desarrollo de las neuro tecnologías cuyo inmenso potencial en términos de beneficios en el ámbito de la salud es equiparable a los riesgos que plantean. Aunque atrae menos la atención mediática que la inteligencia artificial, la neuro tecnología, asociada a la IA, se está desarrollando con la misma rapidez²⁴.

La neuro tecnología incluye herramientas que pueden interactuar directamente con el sistema nervioso para medirlo, modularlo o estimularlo. Ofrece beneficios prometedores, particularmente en medicina: la estimulación cerebral profunda alivia los síntomas de trastornos como la depresión y la enfermedad de Parkinson y las interfaces entre el cerebro y el ordenador permiten a las personas con discapacidades controlar las prótesis o comunicarse a través del pensamiento.

Sin embargo, mientras que su uso médico está estrictamente regulado, la neuro tecnología en muchos casos sigue careciendo de regulación en otras áreas. Muchos consumidores utilizan esta tecnología sin saberlo, a través de dispositivos comunes como diademas o auriculares conectados, que utilizan datos neuronales para monitorear la frecuencia cardíaca, el estrés o el sueño. Estos datos altamente sensibles pueden revelar pensamientos, emociones y reacciones, y pueden compartirse sin consentimiento. Los

²⁴ Un informe de la UNESCO revela un aumento del 700 % en la inversión en empresas de neuro tecnología entre 2014 y 2021. UNESCO: Unveiling the Neurotechnology Landscape Scientific Advancements Innovations and Major Trends, 2023. [Unveiling the neurotechnology landscape: scientific advancements innovations and major trends - UNESCO Digital Library](#)

datos cerebrales son considerados ultrasensibles y requieren de protección especial en términos similares a los datos biomédicos. Pueden ser usados con fines comerciales o políticos, manipulando preferencias sin consentimiento. De ahí que se hayan articulado los denominados neuro derechos diseñados para proteger la actividad cerebral, los datos del cerebro y la autonomía mental, ante los avances de las neuro tecnologías vinculadas a los sistemas de IA. Derechos como la libertad cognitiva para decidir sobre mejoras o evitar manipulaciones mentales, privacidad mental para proteger los pensamientos y datos cerebrales, integridad mental con el fin de prevenir manipulaciones no autorizadas o continuidad psicológica destinada a proteger la identidad personal y la historia interior²⁵ son algunos de estos denominados neuro derechos, esto es derechos humanos en un nuevo ámbito, el propio de una sociedad neuro tecnologizada.

Pero los riesgos y desafíos no afectan exclusivamente a las personas en tanto que sujetos individuales, sino que se proyectan igualmente en su convivencia con otros, en la organización social y política con la proliferación de herramientas generativas que multiplican las *fake news*, reforzando burbujas informativas y dificultando la distinción entre información real y creada; El uso de datos personales para perfilar y manipular a la ciudadanía, a través de *deepfakes* o *bots* personalizados; Las decisiones automatizadas en políticas públicas o justicia que pueden estar basadas en sesgos ocultos y carecer de mecanismos explicativos, lo que reduce la rendición de cuentas. Todo ello comporta generar una confianza posiblemente excesiva en modelos predictivos y generar una visión tecnocrática en la que la IA se considere la única alternativa válida frente al juicio humano. Además, el uso de IA puede crear contenido electoral falso, falsas voces o noticias manipuladas lo que puede subvertir los procesos democráticos. Los *chatbots* y herramientas que adaptan mensajes psicográficos individualizados dificultan tanto la regulación como el control de campañas engañosas²⁶.

Igualmente se debe llamar la atención sobre la concentración de poder, con el dominio de grandes empresas tecnológicas y la creciente influencia de la IA en el sector público susceptibles de poner en riesgo la pluralidad de voces y el equilibrio democrático, así como la fragilidad que genera el vacío normativo. La ausencia de una regulación unificada fomenta una fragmentación normativa, donde los intereses regionales o privados determinan el rumbo, poniendo en riesgo la convivencia política interna

²⁵ Ana Isabel Carreras Presencio “Inteligencia artificial en el siglo XXI. Necesidad de gobernanza y ética” en *Nuevas realidades: un enfoque desde el derecho*, dir. M.^a Ángeles Cano Linares, Ana Isabel Carreras Presencio, 2025, ISBN 979-13-7006-164-7, pp. 14-34. Bernadete Dias, André Lins, Thais Fernandes “Neurorights: Is the creation of new human rights effective in protecting human dignity” International Bar Association 2025.

²⁶ Juan Emilio Suñé Cano “La Unión Europea frente a la mentira digital” en *Nuevas realidades: un enfoque desde el derecho*, dir. M.^a Ángeles Cano Linares, Ana Isabel Carreras Presencio, 2025, ISBN 979-13-7006-164-7, pp. 52-81.

nacional e internacional, configurando un nivel cada vez más decisivo en el tablero geoestratégico.

En una resolución de la Asamblea Parlamentaria del Consejo de Europa se había afirmado, ya en 2017, que “la convergencia entre la nanotecnología, la biotecnología, la tecnología de la información y las ciencias cognitivas, así como la velocidad con que se introducen en el mercado las aplicaciones de las nuevas tecnologías, tienen consecuencias no sólo para los derechos humanos y la manera de ejercerlos, sino también para el concepto fundamental de lo que caracteriza al ser humano” y que como consecuencia de la omnipresencia de esas tecnologías y sus aplicaciones se están desdibujando “las fronteras entre el hombre y la máquina, entre las actividades en línea y fuera de línea, entre el mundo físico y el virtual, entre lo natural y lo artificial, y entre la realidad y la virtualidad. La humanidad está aumentando sus capacidades potenciándolas con la ayuda de máquinas, robots y *software*. Hoy es posible crear interfaces funcionales cerebro-ordenador. Se ha producido un cambio del ser humano “tratado” al ser humano “reparado”, y lo que ahora se vislumbra en el horizonte es el ser humano “aumentado”.

Por su parte, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) comienza su Recomendación sobre la Ética de la Inteligencia Artificial, aprobada en 2021²⁷, reconociendo que “las repercusiones positivas y negativas profundas y dinámicas de la inteligencia artificial en las sociedades, el medio ambiente, los ecosistemas y las vidas humanas, en particular en la mente humana, debido en parte a las nuevas formas en que su utilización influye en el pensamiento, las interacciones y la adopción de decisiones de los seres humanos y afecta a la educación, las ciencias sociales y humanas, las ciencias exactas y naturales, la cultura y la comunicación y la información”

Y añade que “las tecnologías de la IA pueden ser de gran utilidad para la humanidad y que todos los países pueden beneficiarse de ellas, pero que también suscitan preocupaciones éticas fundamentales, por ejemplo, en relación con los sesgos que pueden incorporar y exacerbar, lo que puede llegar a provocar discriminación, desigualdad, brechas digitales y exclusión y suponer una amenaza para la diversidad cultural, social y biológica, así como generar divisiones sociales o económicas; la necesidad de transparencia e inteligibilidad del funcionamiento de los algoritmos y los datos con los que han sido entrenados; y su posible impacto en, entre otros, la dignidad humana, los derechos humanos y las libertades fundamentales, la igualdad de género, la democracia, los procesos sociales, económicos, políticos y culturales, las prácticas científicas y de ingeniería, el bienestar animal y el medio ambiente y los ecosistemas”.

²⁷ En la Conferencia General de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), reunida en París del 9 al 24 de noviembre de 2021, en su 41ª reunión.

En este trabajo se examinarán los principales riesgos derivados del uso de sistemas de IA, con incidencia particular en el caso de los conflictos armados y el derecho internacional humanitario, así como en las respuestas que desde el ámbito internacional, universal y regional europeo se han adoptado con la finalidad de prevenirlos.

1. PRINCIPALES RIESGOS DERIVADOS DEL USO/MAL USO DE SISTEMAS DE INTELIGENCIA ARTIFICIAL PRINCIPALES RIESGOS DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL (IA)

Si bien aún se afirma que los sistemas de IA transformarán el mundo en un futuro, y probablemente lo hagan aún más, lo cierto es que ya lo están haciendo y generando con ello una serie de riesgos que han de ser visibilizados para poder tomar conciencia individual y colectivamente de ellos. Solo con esa concienciación será posible, sino evitar, al menos paliar sus consecuencias.

1.1. Ciberdelitos y ciberseguridad

Una primera y clara clasificación binaria de estos riesgos distingue entre los daños y riesgos generados por el simple uso de los sistemas de IA de los riesgos y daños derivados del mal uso o del uso con fines criminales de los diversos sistemas de IA. En este segundo caso, que se examina en este apartado, el problema no es la tecnología ni el fin para el que fue creada sino el uso que se haga de ella

Se entiende por mal uso de la inteligencia artificial cualquier utilización de los sistemas IA con fines maliciosos, dañinos o ilícitos. Es decir, usar la IA para causar algún tipo de perjuicio, difundir información falsa con alguna intencionalidad de manipulación o engaño, para cometer algún delito o, simplemente, utilizarla con resultados perjudiciales para una o varias personas o entidades, aun cuando, en primer lugar, el objetivo no fuera causar el daño producido.

Con carácter general, el mal uso o uso malicioso de los sistemas de IA está relacionado con la manipulación y el engaño con fines tan diversos como el plagio, la pornografía y el de ánimo de lucro ilícito. Por ello, entre los usos de la inteligencia artificial con fines maliciosos más destacados y habituales encontramos el plagio y el fraude de la IA generativa, los *deepfakes* y la desinformación, la creación de contenido pornográfico y la suplantación de identidad.

Respecto al primero, el plagio y el fraude de la IA generativa, se debe indicar que la IA generativa, que resulta muy útil para realizar, e incluso automatizar, diferentes tareas, también puede ayudarnos a enfocar y preparar temas para la creación de contenidos o para la redacción de trabajos e informes o ayudarnos a corregir textos. Sin embargo, la puesta a disposición del público general de este tipo de IA ha creado toda una

fuelle de conflictos con los derechos de autor y la creación de contenidos originales. Así, las IAs generativas de texto se emplean por parte de estudiantes de todo el mundo para hacer trabajos, por empresas para crear contenidos para sus páginas web, especialmente orientados al marketing digital y al marketing de contenidos mientras que las IAs de generación de imágenes se usan para obtener imágenes sin necesidad de contratar a ilustradores y fotógrafos o recurrir a un banco de imágenes, que, además, empiezan a tener también contenido generado por los propios sistemas IA.

Uno de los daños ya causados por estos sistemas es el derivado de su entrenamiento con contenidos sin atender ni respetar los derechos de autor. El riesgo de pérdidas de empleo en ciertos creadores de contenidos es claro, así como el menoscabo a la creatividad y originalidad humanas si la alimentación y entrenamiento de los sistemas de los sistemas se realiza con contenidos creados por otras IAs.

En el marco de las IA generativas, particularmente de imágenes, uno de los malos usos más extendidos, al alcance del público en general gracias a ellas, es la creación de *deepfakes*²⁸ empleados para llevar a cabo acciones maliciosas o para extender bulos y desinformación con fines diversos.

Así, cada vez es más difícil distinguir entre una imagen real y una imagen generada por IA, o unas declaraciones de cualquier persona lo que facilita el engaño y la manipulación de todos. Sumamos con ello una nueva incertidumbre a nuestra vida cotidiana, incluso más radical que el principio de indeterminación de la física cuántica.

Además, uno de los usos maliciosos, cada vez más extendido, es el de la creación de imágenes, videos o audios usando técnicas de *deep learning*, para generar contenido falso que parece completamente real con fines pornográfico. Con este mismo fin, existen igualmente aplicaciones basadas en IA que ofrecen “imágenes desnudas” basadas en fotos reales subidas en alguna aplicación, siendo su versión más grave la creación de contenidos pornográficos utilizando las imágenes de menores de edad.

Siguiendo en el ámbito delictivo, la IA generativa facilita igualmente la suplantación de identidad con la apertura de perfiles falsos en redes sociales o haciéndose pasar por otra persona, lo que constituyen otros malos usos de la IA destinados generalmente a llevar a cabo ataques de *phishing* o *spear phishing* o intentos de chantajes. Y estos usos pueden desencadenar efectos muy perjudiciales y manipulación no solo a título sino afectando a sociedades enteras. Y ello porque a nivel colectivo y como consecuencias del mal uso de la inteligencia artificial, se incrementan las posibilidades de generar desinformación y, por lo tanto, generar mayor desconfianza entre la ciudadanía y la manipulación de esta. Quizás el ejemplo más extremo conocido hasta la fecha haya sido el

²⁸ Imágenes, videos o audios manipulados o generados mediante inteligencia artificial, especialmente usando técnicas de *deep learning*, para crear contenido falso que parece completamente real.

de la propagación, en 2017 en Myanmar, de mensajes de odio y desinformación contra la población rohingyá, creados por medio de los algoritmos de Facebook.

Por ello, cabe afirmar que si con las nuevas tecnologías de la información y comunicación, surgieron de nuevos tipos de delito, la denominada ciberdelincuencia, con la IA su incidencia se ha incrementado haciendo más necesarias que nunca medidas de ciberseguridad²⁹, pues la realidad es que a medida que la interconexión y el uso de IA crecen, también lo hacen los vectores de ataque, exigiendo mayores iniciativas de seguridad.

Teniendo en cuenta que nos encontramos tanto en un contexto de globalización de base económica como en uno de interconectividad digital, con comunicaciones instantánea y acceso inmediato a la información resulta completamente ilusorio, al igual que ocurre en otros ámbitos, pensar en respuestas nacionales para hacer frente a los retos y riesgos que se plantean. La única opción posible es una respuesta global con la creación de un marco regulatorio tendente a mitigar los múltiples riesgos que plantean el rápido desarrollo de los sistemas de IA, sin por ello menoscabar sus igualmente claros beneficios.

En este campo, como en muchos otros, el Consejo de Europa fue pionero con la adopción³⁰ del ya en vigor Convenio de Budapest sobre Ciberdelincuencia que tiene como fin armonizar la legislación nacional sobre ciberdelitos, facilitar la investigación y el enjuiciamiento de delitos informáticos y promover la cooperación internacional eficaz y rápida. Se trata del primer tratado internacional que afronta los ciberdelitos³¹.

A nivel universal, la Asamblea General de Naciones Unidas adoptó el 24 de diciembre de 2024³² la Convención de Hanoi contra la ciberdelincuencia. con fines similares a la Convención del Consejo de Europa, con la importante ventaja de su alcance potencialmente universal. En la misma se incluyen delitos como *ransomware*, fraude financiero, pornografía no consensuada al tiempo que reconoce explícitamente los riesgos para derechos humanos y privacidad.

²⁹ Entendida como el conjunto de prácticas, tecnologías y procesos para proteger redes, sistemas y datos contra accesos no autorizados o maliciosos.

³⁰ Adoptado por el Consejo de Europa el 23 de noviembre de 2001 en Budapest y en vigor desde el 1 de julio de 2004.

³¹ Con la tipificación de delitos como acceso no autorizado, fraudes informáticos, difusión de pornografía infantil y violaciones a la seguridad informática, estableciendo procedimientos de investigación y cooperación transfronteriza y asistencia legal mutua, incluida una red de contacto. Se complementa con 2 Protocolos adicionales, el primero dedicado a la xenofobia y racismo y el Segundo protocolo adicional (ETS No. 224) que refuerza cooperación transfronteriza, acceso a datos de usuarios y urgencias, con salvaguardas de derechos fundamentales.

³² Asamblea General de la ONU mediante resolución 79/243. Abierto a la firma el 25 de octubre de 2025 en Hanoi (Vietnam), y también en Tokio hasta el 31 de diciembre de 2026.

1.2. SU INCIDENCIA EN EL ESTADO DE DERECHO: DERECHOS HUMANOS Y DEMOCRACIA.

Los principales riesgos identificados como consecuencia de la proliferación de sistemas de IA y su uso, malicioso o no, se pueden agrupar bien en función del o de los derechos humanos especialmente afectados o, bien, función del sector al que se aplican los sistemas IA como una vía para aumentar la eficiencia y reducir el sesgo humano ya sea, por ejemplo, la administración pública, el ámbito sanitario o el judicial³³.

En el caso de los derechos los más afectados son el derecho a la privacidad, intimidad, el principio de no discriminación, el derecho a la libertad de expresión, el derecho al trabajo, a la educación, a la propia identidad y al desarrollo de todos, incluidos los más vulnerables.

Uno de los desarrollos que genera más preocupación son las herramientas biométricas (reconocimiento facial, sistemas de identificación remota, categorización y “reconocimiento de emociones”) y su impacto en derechos fundamentales como la privacidad, la protección de datos, la no discriminación o la libertad de expresión y reunión.

Algunos de los daños potenciales trascienden este ámbito individual y pueden comportar riesgos y amenazas a nivel de organización política y social, tanto a nivel nacional como internacional. Así, por ejemplo, la desinformación política puede comportar campañas de manipulación que pueden implicar una injerencia en los asuntos internos de otros Estados en contra del artículo 2 de la Carta de Naciones Unidas, mucho más difícil de probar que las clásicas actuaciones. Otra manifestación a nivel internacional es el incremento de la desigualdad entre Estados desarrollados y Estados en vías de desarrollo, en contra del Objetivo de Desarrollo Sostenible (ODS) 10 dando lugar a la que se conoce como brecha digital.

Se examinan a continuación algunos de estos riesgos en relación con los principales derechos afectados.

En primer lugar, prácticas y desarrollos, en particular los derivados de mecanismos de vigilancia masiva³⁴, que afectan a la privacidad y a la intimidad. Es el caso de

³³ Sin embargo, investigaciones recientes revelan que estos sistemas pueden reproducir y amplificar desigualdades estructurales, vulnerando principios fundamentales de derechos humanos. Cf. Vielka M. Párraga. “From efficiency to exclusion rethinking AI governance in public institutions”.

³⁴ Por ejemplo, el artículo "Respect for privacy and personal data protection in the context of mass surveillance of communication" de Esther Salamanca Aguado analiza el impacto de programas de vigilancia masiva de comunicaciones (revelados por Edward Snowden) en derechos humanos, centrándose en el art. 8 del Convenio Europeo de Derechos Humanos.

los sistemas de IA aplicados al reconocimiento facial y monitoreo en tiempo real que pueden vulnerar el derecho a la intimidad y a la libertad de movimiento y sobre los que se han alzado numerosas voces advirtiendo sobre la necesidad de prohibirlas.

Y ello porque algunos de estos sistemas de IA pueden proceder a la identificación biométrica remota, con comparación de datos biométricos a distancia con bases de datos sin conocimiento previo de la presencia del individuo mientras que otros son sistemas de reconocimiento de emociones, con inferencia de emociones o intenciones existiendo igualmente sistemas de categorización biométrica, con asignación a categorías como sexo, edad, origen étnico, orientación sexual o política basada en datos de esta naturaleza³⁵.

Por lo tanto, las tecnologías biométricas procesan características físicas³⁶ y fisiológicas o conductuales³⁷ y, son clasificadas, por un lado, en unimodales y multimodales y, por otro, en las denominadas “duras” con alta fiabilidad para identificación) frente a las “suaves”, demográficas y/o emocionales para mejorar rendimiento.

Este uso de la IA, combinada con datos biométricos se aplica ya en sectores como control de fronteras, finanzas, educación y servicios públicos. Y como todos los sistemas de IA, la biometría aparece como extremadamente útil para la autenticación y seguridad, incluso la adopción de biometría en el voto puede robustecer la integridad de las elecciones, pero sin salvaguardias estrictas, como limitaciones a la identificación biométrica en tiempo real en espacios públicos, exigencias de evaluación previa de impacto y, en algunos casos, prohibiciones estas tecnologías pueden suscitar temores de vigilancia, discriminación de ciertas comunidades y posible manipulación si no existen garantías claras de transparencia, control ciudadano y protección de datos personales.

Una de las consecuencias de la vigilancia masiva es la potencial erosión de las libertades fundamentales, desalentando por ejemplo protestas de carácter público lo que puede impactar en la libertad de movimiento y en la de expresión y, en última instancia, afectar al funcionamiento del sistema democrático. Además los sistemas IA pueden incurrir en sesgos discriminatorios en relación con las minorías por no mencionar que la falta de consentimiento y transparencia viola la autonomía individual. Por su parte, la categorización de las emociones abre un campo escasamente contemplado, que puede afectar no solo a la privacidad sino a la intimidad y a la autonomía.

³⁵ Fuster Gonzales, Gloria, Nadola Peeters, Michalina. “A new study on biometrics & AI written by, titled 'Person identification, human rights and ethical principles: Rethinking biometrics in the era of artificial intelligence’”. *Scientific Foresight Unit (STOA)STOA*. European Parliamentary Research Service, diciembre 2021.

³⁶ Como huellas, iris, rostro, voz, ADN

³⁷ Como firma, marcha, interacción con teclado/ratón, entre otras.

En todo caso, las regulaciones sobre inteligencia artificial están redefiniendo el uso de tecnologías biométricas con resultados dispares en distintas regiones del mundo. Así, en Estados Unidos, prima la seguridad nacional, el control fronterizo y la identificación digital del votante mientras que la Unión Europea (UE) aparece como líder normativo con la aprobación del Reglamento de IA³⁸ que introduce una clasificación por niveles de riesgo y sitúa muchas aplicaciones biométricas en el nivel de mayor riesgo³⁹.

Por otro lado, y, como ya se ha apuntado, el principio de no discriminación se ve igualmente afectado por el desarrollo de sistemas IA, con los sesgos que suelen implicar al ser entrenados con datos sesgados, como los que se dan en todas las sociedades, y pueden reproducir y amplificar desigualdades raciales, de género o socioeconómicas, afectando el derecho a la igualdad y la no discriminación. Por ello, a las causas tradicionales de no discriminación debería añadirse la de no discriminación algorítmica. Su impacto es, como siempre ocurre, mayor en los grupos más vulnerables. Así, por ejemplo, niños y adolescentes son especialmente susceptibles a la manipulación algorítmica, lo que exige educación digital y normas éticas específicas al tiempo que la automatización impulsada por IA amenaza principalmente a empleos de mujeres, jóvenes o personas con baja formación digital, lo que puede aumentar la exclusión económica. Además, si los más vulnerables no cuentan con acceso a las herramientas de IA, muchas de las cuales no son gratuitas, quedan aún más rezagados.

Surge así la denominada brecha digital como riesgo en la era de la IA, entendida como la diferencia entre quienes tienen acceso y utilizan las nuevas tecnologías de la información y del conocimiento, incluida la conectividad a Internet, los dispositivos con conexión a Internet y las habilidades digitales, y quienes no lo tienen⁴⁰.

En realidad, en la brecha digital convergen varias brechas en una, con una brecha tecnológica, una brecha de contenidos, lingüística y cultural⁴¹ una brecha de género y una brecha comercial.

³⁸ 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° 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

³⁹ Esto se debe a que el tratamiento de datos biométricos (en especial el reconocimiento facial en espacios públicos y sistemas de identificación remota) puede amenazar gravemente la privacidad, favorecer la vigilancia masiva y generar discriminaciones si no se controla adecuadamente. AI-Regulations-and-Biometric-Technology-in-2025--BioID-Blog

⁴⁰ De acuerdo con la definición formulada en Tal y como se define en la publicación “Addressing the Digital Divide” de 2021 del Programa de las Naciones Unidas para los Asentamientos Humanos.

⁴¹ Los modelos de lenguaje (funcionan muy bien en inglés, pero tienen un rendimiento pobre en idiomas con pocos datos, lo que genera exclusión cultural y lingüística. Esto limita el acceso a información,

La brecha digital amplifica los riesgos asociados a la IA al dejar fuera a personas, comunidades y países que no cuentan con infraestructura, datos, habilidades ni apoyo institucional. Sin una intervención global coordinada, la IA puede reforzar y profundizar desigualdades existentes en lugar de cumplir su potencial de servicio público y progreso compartido.

Así, la Organización de las Naciones Unidas (ONU) advierte que, sin políticas inclusivas, la IA podría revertir décadas de avances, ampliando brechas económicas y de capacidades entre países ricos y pobres⁴². Mientras que un informe del Programa de las Naciones Unidas para el desarrollo (PNUD) describe este fenómeno como una “nueva gran divergencia”, donde las naciones con mejor infraestructura y gobernanza acumularán los beneficios, mientras que las demás se quedarán aún más atrás⁴³.

En relación con la libertad de expresión, de pensamiento y la manipulación informativa, ya se ha señalado que herramientas como *deepfakes* y *bots* generativos pueden difundir desinformación, socavando la credibilidad institucional e incluso los procesos democráticos poniendo de manifiesto una estrecha relación entre democracia y desinformación. La IA generativa, las redes sociales y los algoritmos de recomendación multiplican el alcance y la eficacia de la desinformación y de las operaciones de manipulación informativa extranjera (FIMI), explotándose estas herramientas, como se examinará en un apartado posterior, como instrumento de las denominadas guerras híbridas y cognitivas.

En todo, surge con fuerza la necesidad de consolidar el derecho a la autodeterminación informativa y su contenido normativo, entendido como la capacidad de una persona para determinar, en principio, cuándo, cómo y en qué medida sus datos personales pueden ser divulgados, consultados, tratados y utilizados por agentes públicos o privados. En el contexto de la vigilancia digital, donde la recopilación de datos es masiva, constante, omnipresente y, a menudo, de naturaleza opaca, este derecho funciona como una garantía fundamental de la libertad personal y la dignidad humana, diseñada para garantizar que las personas mantengan un control significativo sobre sus datos personales.

En todo caso parece igualmente incuestionable que en el marco de la gobernanza de la IA las directrices éticas, por sí solas, no bastan frente a sistemas capaces de

educación, oportunidades laborales y protección contra desinformación para hablantes de idiomas minoritarios.

⁴² <https://news.un.org/es/story/2023/03/1519152>.

⁴³ Cinco grandes potencias económico-tecnológicas y unas 100 empresas concentran el 40 % de la inversión privada en IA, excluyendo a 118 países del Sur Global de los debates de gobernanza Programa de las Naciones Unidas para el desarrollo (PNUD). Estrategia digital. 2022- 2025. https://digitalstrategy.undp.org/documents/Digital-Strategy-2022-2025-Full-Document_ES_Interactive.pdf

decisiones opacas a gran escala, y que la regulación debe incluir mecanismos ágiles para designar nuevos sistemas de alto riesgo, participación de sociedad civil en la evaluación de riesgos, y prohibiciones claras de ciertos usos de alto impacto insiste en la transición de marcos puramente éticos, absolutamente necesarios pero insuficientes a regulaciones vinculantes de carácter universal de forma que se desarrolle lo que se ha denominado una “IA confiable”. dimensión geopolítica: el enfoque europeo de “IA centrada en el ser humano” y fuerte protección de derechos se presenta como posible ventaja competitiva y base para liderar estándares internacionales frente a modelos más autoritarios.

Se identifican requisitos básicos para una “IA fiable”: intervención y supervisión humanas, solidez y seguridad técnica, privacidad y gestión de datos, transparencia y explicabilidad, diversidad/no discriminación, bienestar social y ambiental, y rendición de cuentas. confiable y centrada en el humano, clave para el crecimiento económico y derechos fundamentales en la UE⁴⁴.

Y para finalizar solo apuntar dos cuestiones más. La de la sostenibilidad de la IA debido al alto consumo energético y de recursos que implica, presionando los suministros globales, y la cuestión de su fiabilidad, no solo por sus sesgos sino por sus no tan infrecuentes alucinaciones.

1.2. LAS NEURO TECNOLOGÍAS: REVOLUCIÓN EN CIERNES

Aunque atrae menos la atención mediática que la inteligencia artificial, las neuro tecnologías se están desarrollando con la misma rapidez. Estrechamente relacionadas con los sistemas de IA, las neuro tecnologías incluyen herramientas que pueden interactuar directamente con el sistema nervioso para medirlo, modularlo o estimularlo. Y plantean, al igual que hemos venido analizando con los sistemas IA, un carácter dual. Por un lado, ofrece beneficios prometedores, particularmente en medicina⁴⁵ mientras que, por otro, plantea nuevos retos y desafíos éticos.

La neuro tecnología incluye dispositivos y procedimientos para monitorear, investigar o manipular la estructura y función neural, desde imágenes cerebrales hasta interfaces cerebro-computadora, herramientas han evolucionado para diagnosticar y tratar enfermedades. Sin embargo, su expansión más allá de la medicina, en *neuromarketing*, *neurogaming* o mejora cognitiva, genera dilemas inéditos. Los datos neurales se convierten en bienes valiosos para fines comerciales y combinados con la denominada “huella digital”, la que dejan las búsquedas en línea y la geolocalización), amplifican riesgos de discriminación y vigilancia.

⁴⁴ David Kemp, “What if generative AI is reaching its limits?”, *European Parliamentary Research Service*, noviembre 2025.

⁴⁵ La estimulación cerebral profunda alivia los síntomas de trastornos como la depresión y la enfermedad de Parkinson y las interfaces entre el cerebro y el ordenador permiten a las personas con discapacidades controlar las prótesis o comunicarse a través del pensamiento.

Así, mientras su uso médico está estrictamente regulado, la neuro tecnología carece, en muchos casos, de regulación en otras áreas. Muchos consumidores utilizamos esta tecnología sin saberlo, a través de dispositivos comunes como diademas o auriculares conectados, que utilizan datos neuronales para monitorear la frecuencia cardíaca, el estrés o el sueño. Estos datos altamente sensibles pueden revelar pensamientos, emociones y reacciones, y pueden compartirse sin nuestro consentimiento.

Por ello, los Estados miembros de la UNESCO adoptaron el pasado 12 de noviembre de 2025, una Recomendación sobre la ética en la neuro tecnologías, en línea con la que adoptaron en 2021 sobre la ética en la IA , estableciendo salvaguardias esenciales para asegurar que la neuro tecnología contribuya a mejorar la vida de quienes más la necesitan sin poner en peligro los derechos humanos, que se examina en un apartado posterior y está basada en el informe del Comité Internacional de Bioética de la UNESCO sobre *Cuestiones Éticas de la Neuro tecnología*, adoptado en diciembre de 2021.

Ese informe clasifica las neuro tecnologías en varias categorías. En primer lugar, Las neuroimágenes no invasivas⁴⁶, que no obstante plantean preocupaciones éticas en interpretación y hallazgos incidentales. En segundo, Los Neuro dispositivos⁴⁷ que restauran funciones sensoriales o motoras, aunque con ciertos riesgos⁴⁸. Las interfaces cerebro-computadora (BCI) traducen señales cerebrales en comandos para prótesis o realidad virtual, mientras que la IA en neurociencias analiza patrones masivos de datos neurales lo que desafía algunos de los principios bioéticos claves establecidos en la Declaración Universal de Bioética y Derechos Humanos de 2005.

La integridad cerebral/mental y la dignidad humana se pueden ver amenazadas por manipulaciones invasivas que alteran la identidad personal y la continuidad psicológica. La autonomía puede verse comprometida si las tecnologías influyen en decisiones sin consentimiento pleno, especialmente en personas vulnerables como niños o pacientes con problemas de salud mental.

Por ello, los neuro derechos buscan ser un marco normativo adaptado a los retos que plantean las neuro tecnologías, como la lectura o modificación del cerebro, reforzando la protección de la privacidad mental, la autonomía cognitiva y la identidad personal en una nueva era tecnológica⁴⁹.

⁴⁶ Como EEG (alta resolución temporal), fMRI (anatomía funcional), PET o fNIRS, miden actividad eléctrica, metabólica o sanguínea cerebral.

⁴⁷ Como la estimulación cerebral profunda (DBS) para Parkinson o tDCS (estimulación transcraneal).

⁴⁸ Complicaciones quirúrgicas, efectos psicológicos (hipersexualidad, ludopatía).

⁴⁹ Dias, Bernadete André Lins, Thais Fernandes. “Neurorights, Is the creation of new human rights effective in protecting human dignity from the misuse of neurotechnology?”, *International Bar Association*, 2025.

La privacidad mental emerge como prioridad pues los datos cerebrales revelan pensamientos, emociones y predisposiciones, vulnerables a hackeos o usos no consentidos en neuromarketing. La accesibilidad y justicia social demandan equidad, evitando que solo elites accedan a mejoras cognitivas, lo que agrava desigualdades. Pero resulta igualmente importante la libertad cognitiva que decidir sobre mejoras o manipulaciones mentales, así como la privacidad/integridad mental para proteger los pensamientos y datos cerebrales y la integridad mental para prevenir manipulaciones no autorizadas. Junto el derecho a la continuidad psicológica para proteger la identidad personal e historia interior, el acceso equitativo a potenciales mejoras cerebrales y la protección contra manipulaciones, sesgos y violaciones de voluntad.

Es más que probable que todos nosotros seamos ya vulnerables a estos riesgos. En todo caso, estimo que uno de los mayores riesgos frente a esta proliferación de sistemas de IA es su uso potencial sin una evidencia clara de sus beneficios y sin unas reglas claras que permitan diseñarlos para minimizar al máximo los potenciales efectos adversos. Si en el pasado, la revolución que supuso el surgimiento de internet y de las redes sociales comportó que, de alguna manera, perdiéramos la titularidad de nuestros datos personales, ahora con la revolución IA y las neuro tecnologías lo que podemos perder es aún más importante, pudiendo alterar nuestra intimidad, nuestra voluntad y nuestra intrínseca dignidad como seres humanos.

2. PROPUESTAS EN EL AMBITO UNIVERSAL: DE LA ÉTICA EN LA IA AL PACTO DIGITAL MUNDIAL

En este apartado se examinan las aportaciones más relevantes que se han adoptado a nivel universal, en el sistema de las Naciones Unidas, mencionadas en epígrafes anteriores.

2.1. NACIONES UNIDAS: PACTO DIGITAL MUNDIAL

En la Cumbre de las Naciones Unidas sobre el Futuro celebrada en septiembre de 2024, los representantes de los Estados miembros adoptaron el Pacto para el Futuro⁵⁰, que incluye como anexo el Pacto Digital Global (PDG) que, veinte años después de la Cumbre Mundial sobre la Sociedad de la Información, traza una hoja de ruta para

⁵⁰ Nuestra Agenda Común del secretario general Antonio Guterres (2021), que proponía una Cumbre del Futuro con un eje tecnológico culminó que resultaría en este pacto. Es un acuerdo intergubernamental amplio y ambicioso, considerado el más comprensivo en décadas. Incluye compromisos clave, entre ellos la reforma del Consejo de Seguridad, el desarme nuclear, la gobernanza del espacio exterior y la regulación de nuevas tecnologías. consta de cinco capítulos: desarrollo sostenible; paz y seguridad; ciencia, tecnología e innovación y cooperación digital; juventud y generaciones futuras; y transformación de la gobernanza global.

la cooperación digital global con el fin de aprovechar el inmenso potencial de la tecnología digital y cerrar las brechas digitales.

El Pacto no tiene carácter legal vinculante, pero representa un consenso global sobre cómo abordar los retos digitales urgentes, promoviendo un enfoque multilateral y multisectorial y representa el primer marco global integral para la cooperación digital y la gobernanza de la IA.

A destacar que uno de sus pilares centrales de la gobernanza de los datos, lo que supone que, por primera vez, esta cuestión se incluye formalmente en la agenda mundial de las Naciones Unidas, lo que representa un importante primer paso⁵¹. Igualmente, por primera vez, se acordó a nivel global un compromiso conjunto sobre conectividad universal, inteligencia artificial, derechos digitales y cooperación internacional en el entorno digital.

Negociado por 193 Estados miembros y basado en consultas mundiales y regionales, el Pacto compromete a los gobiernos a defender el derecho internacional y los derechos humanos en línea y a adoptar medidas concretas para que el espacio digital sea seguro y protegido.

Sus propósitos principales son crear un futuro digital abierto, libre, seguro y centrado en las personas, respetando los derechos humanos y apoyando los Objetivos de Desarrollo Sostenible; Reducir la brecha digital eliminando desigualdades tanto entre países como dentro de ellos y fijar principios y estándares globales sobre gobernanza de datos, ciberseguridad, IA y otros ámbitos tecnológicos⁵².

Los principios rectores que indica incluyen la participación inclusiva de todos los Estados y actores; orientación al desarrollo con foco en países en desarrollo; anclaje en el derecho internacional y derechos humanos; igualdad de género; sostenibilidad ambiental; distribución equitativa de beneficios; accesibilidad; interoperabilidad; enfoque responsable en tecnologías emergentes; fomento de innovación; provisión de medios de implementación; y visión prospectiva ante cambios tecnológicos. Estos principios se refuerzan mutuamente para guiar la cooperación digital global, priorizando a países menos desarrollados y reconociendo el rol de la ONU como plataforma clave.

⁵¹ Las consultas intergubernamentales fueron co-lideradas por Suecia y Zambia desde enero de 2023, involucrando a múltiples actores: gobiernos, sector privado, sociedad civil, academia. Como seguimiento, la Asamblea General de las Naciones Unidas solicitó a la Comisión de Ciencia y Tecnología para el Desarrollo (CSTD) que estableciera un grupo de trabajo multilateral dedicado a la gobernanza de datos a todos los niveles pertinentes para el desarrollo. Este grupo se ha reunido en dos ocasiones: el 1 y 2 de mayo de 2025 y el 3 y 4 de julio

⁵² Galceran, Marta. “El pacto digital global y la transición responsable e inclusiva” Informe CIDOB 12- 2024.

Recoge asimismo, como compromisos y acciones, la de conectar a toda la población, escuelas y hospitales aún offline mediante financiación innovadora antes de 2030; Establecer un Panel Científico Internacional sobre IA vinculado a la ONU y un diálogo global sobre la gobernanza de la IA; Salvaguardar los derechos humanos frente a los efectos adversos de tecnologías digitales, incluyendo la lucha contra desinformación y discurso de odio.; Fomentar prácticas de gobernanza de datos responsables, equitativas e interoperables; Garantizar accesibilidad, transparencia y rendición de cuentas en plataformas digitales, protegiendo la libertad de expresión y privacidad.

Y, en cuanto a su implementación y seguimiento, los Estados se comprometieron a acciones concretas antes de 2030, reconociendo las contribuciones fundamentales del sector privado, al que se exhorta a aplicar los Principios Rectores sobre las Empresas y los Derechos Humanos de las Naciones Unidas, las comunidades técnicas, los investigadores y la sociedad civil a la cooperación digital.

En definitiva, el Pacto Digital Global aspira a ser una guía de políticas y un sello de consenso multilateral para construir una economía y sociedad digital más equitativa, segura y centrada en las personas. Cada uno de los compromisos destacados en el Pacto está asociado al menos a uno de los Objetivos de Desarrollo Sostenible.

Y, en particular, el objetivo 3 del Pacto se refiere específicamente a la gobernanza de estos espacios para “defender el derecho internacional de los derechos humanos a lo largo del ciclo de vida de las tecnologías digitales y emergentes, de modo que los usuarios puedan beneficiarse de las tecnologías digitales de forma segura y estén protegidos contra violaciones, abusos y todas las formas de discriminación”⁵³.

3. LA LABOR DE LA ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA Y SU ENFOQUE ÉTICO.

Durante décadas, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) ha desempeñado un papel destacado en la regulación ética de los avances científicos y tecnológicos. Así, en noviembre de 2021, se celebró la Conferencia General de la UNESCO que aprobó la Recomendación sobre la ética de la inteligencia artificial⁵⁴, con el objetivo de fomentar el uso de la IA para fines pacíficos, así como establecer las bases para poner a la IA al servicio de la sociedad y el medio ambiente y, más recientemente, en noviembre de 2025 la Recomendación sobre la ética en la neuro tecnología. Dos Recomendaciones paralelas en dos ámbitos punteros y estrechamente relacionados.

⁵³ Praachi Kumar Examining the Causes and Consequences of the Digital Divide(s). blog post UNU-MERIT 27 Sep 2024 <https://unu.edu/merit/blog-post/examining-causes-and-consequences-digital-divides>

⁵⁴ UNESCO Recomendación sobre la ética de la inteligencia artificial SHS/BIO/PI/2021/1 https://unesdoc.unesco.org/ark:/48223/pf0000381137_spa

La Recomendación sobre la ética de la inteligencia artificial es el primer instrumento internacional que orienta el desarrollo y uso de la IA con base en la dignidad humana, los derechos humanos, la igualdad de género, la justicia social y la protección del medio ambiente. Define la IA en sentido amplio como aquellos sistemas con capacidad para procesar datos de forma similar a un comportamiento inteligente lo que resulta crucial, ya que el rápido ritmo del cambio tecnológico dejaría obsoleta de forma repentina cualquier definición fija y estrecha, además de hacer inviables las políticas que se hubieran podido desarrollar de cara al futuro. Además, establece valores, principios y ámbitos de acción para que Estados y demás actores adopten marcos jurídicos, políticas públicas y prácticas responsables a lo largo de todo el ciclo de vida de los sistemas de IA.

El texto subraya que las cuestiones éticas abarcan todas las etapas del ciclo de vida del sistema IA, investigación, diseño, desarrollo, financiación, despliegue, uso, supervisión, evaluación, desmantelamiento y fin de uso, e involucran a múltiples actores (investigadores, empresas, Estados, usuarios, universidades, etc.).

Los principales objetivos de la Recomendación son, entre otros, poner la IA al servicio de las personas, sociedades y ecosistemas; prevenir daños; traducir valores en acciones concretas; proteger derechos humanos y generaciones futuras; preservar medio ambiente y diversidad cultural; promover diálogo multiactor y el acceso equitativo a beneficios, con atención especial a países menos adelantados⁵⁵.

Los valores fundamentales que sientan las bases para que los sistemas de IA trabajen por el bien de la humanidad, las personas, las sociedades y el medio ambiente son cuatro. El primero es el respeto, protección y promoción de los derechos humanos, las libertades fundamentales y la dignidad humana, entendida como valor intrínseco e igual de toda persona sin discriminación por raza, género, edad, origen, religión, situación económica u otras condiciones. La IA debe mejorar la calidad de vida sin cosificar a las personas, especialmente a grupos en situación de vulnerabilidad (niños, personas mayores, personas con discapacidad, enfermos), ni permitir violaciones de derechos.

También se reconoce la prosperidad del medio ambiente y los ecosistemas como condición para disfrutar beneficios de la IA, exigiendo reducir su huella ambiental, minimizar el cambio climático y evitar explotación insostenible de recursos naturales. La diversidad y la inclusión son valores clave y se promueve la participación de todas las personas y se pide reducir brechas de infraestructura, educación y capacidades, especialmente en países de ingreso medio. Otro valor es vivir en sociedades pacíficas,

⁵⁵ Carreras Presencio Ana Isabel. “Inteligencia artificial en el siglo XXI. Necesidad de gobernanza y ética” en *Nuevas realidades: un enfoque desde el derecho* dir. M.^a Ángeles Cano Linares, Ana Isabel Carreras Presencio, 2025, pp. 14-34.

justas e interconectadas, donde la IA contribuya a la solidaridad, la paz, la justicia y la coexistencia humana con otros seres vivos y el medio natural, sin dividir ni enfrentar grupos ni mermar autonomía y seguridad. Todos los valores deben considerarse interdependientes y, cuando entren en tensión, gestionarse con evaluación contextual, proporcionalidad y respeto a los derechos humanos.

Además, la Recomendación identifica diez principios básicos que establecen ese enfoque de la ética de la IA centrado en los derechos humanos.

Se trata de los principios de proporcionalidad e inocuidad⁵⁶, seguridad y protección⁵⁷, derecho a la intimidad y protección de datos⁵⁸, gobernanza y colaboración adaptativas y de múltiples partes interesadas⁵⁹, responsabilidad y rendición de cuentas⁶⁰,

⁵⁶ El uso de sistemas de IA no debe ir más allá de lo necesario para alcanzar un objetivo legítimo. La evaluación de riesgos debe utilizarse para prevenir los daños que puedan derivarse de usos ilegítimos.

⁵⁷ El principio de seguridad y protección obliga a prevenir daños no deseados, vulnerabilidades y riesgos de seguridad a lo largo del ciclo de vida, mediante marcos de acceso a datos sostenibles y respetuosos de la privacidad. Los daños no deseados (riesgos de seguridad) y las vulnerabilidades a los ataques (riesgos de protección) deberían ser evitados y tomados en consideración.

⁵⁸ La privacidad debe protegerse y promoverse a lo largo de todo el ciclo de vida de la IA. También deben establecerse marcos adecuados de protección de datos. los datos deben gestionarse conforme a normas nacionales e internacionales, con marcos de protección de datos, gobernanza participativa, principios de privacidad desde el diseño e impacto en la privacidad, y responsabilidad de los actores de IA sobre la información personal en todas las etapas.

⁵⁹ En el uso de datos, deben respetarse el derecho internacional y la soberanía nacional. La participación de diversas partes interesadas a lo largo del ciclo de vida de los sistemas de IA es necesaria para el desarrollo de enfoques inclusivos de gobernanza.

⁶⁰ Los sistemas de IA deben ser auditables y trazables. Deben existir mecanismos de supervisión, evaluación de impacto, auditoría y diligencia debida para evitar conflictos con las normas de derechos humanos y amenazas al bienestar medioambiental.

transparencia y explicabilidad⁶¹, supervisión y decisión humanas⁶², sostenibilidad⁶³, sensibilización y educación⁶⁴ y equidad y no discriminación⁶⁵.

De todos ellos, entiendo necesario detenerse un poco más tanto en el principio de supervisión y decisión humanas como en el de transparencia y explicabilidad.

El primero comporta que la responsabilidad ética y jurídica sea siempre atribuible a personas o entidades jurídicas, y que los sistemas de IA no sustituyan la responsabilidad final de los humanos, especialmente en decisiones irreversibles. Se enfatiza la supervisión pública inclusiva, no solo individual, y la prohibición implícita de marcos donde nadie sea responsable de decisiones automatizadas.

Por su parte, la transparencia y explicabilidad son condiciones para proteger derechos, garantizar responsabilidad y permitir impugnar decisiones automatizadas. Se debe informar a las personas cuando una decisión se base en IA, darles posibilidad de solicitar explicaciones y recurrir ante una persona con facultad de revisar la decisión; el grado de transparencia dependerá del contexto, equilibrándose con privacidad, seguridad y protección. La transparencia incluye comprensibilidad de entradas, salidas y funcionamiento de componentes algorítmicos, y puede requerir compartir códigos o datos en casos de graves amenazas a derechos humanos.

Finalmente, el principio de responsabilidad y rendición de cuentas exige mecanismos de supervisión, evaluación de impacto, auditoría y diligencia debida (incluida protección de denunciantes) para atribuir responsabilidades y corregir conflictos con derechos humanos y medio ambiente mientras que en relación con la gobernanza de datos y participación la Recomendación subraya que el uso de datos debe respetar el derecho internacional y la soberanía nacional, permitiendo a los Estados regular datos generados o que pasan por sus territorios, siempre respetando el derecho a la privacidad.

⁶¹ El despliegue ético de los sistemas de IA depende de su transparencia y explicabilidad (T&E). El nivel de T&E debe ser adecuado al contexto, ya que puede haber tensiones entre T&E y otros principios como la privacidad, la seguridad y la protección

⁶² Los Estados Miembros deberían velar por que siempre sea posible atribuir la responsabilidad ética y jurídica a personas físicas o a entidades jurídicas existentes

⁶³ Las tecnologías de IA deben evaluarse en función de su impacto en la "sostenibilidad", entendida como un conjunto de objetivos en constante evolución, incluidos los establecidos en los Objetivos de Desarrollo Sostenible (ODS) de Naciones Unidas.

⁶⁴ La sensibilización y la comprensión del público respecto de la IA y el valor de los datos deberían promoverse mediante una educación abierta y accesible, la participación cívica, las competencias digitales y la capacitación, y la alfabetización mediática e información

⁶⁵ Los actores de la IA deberían promover la justicia social, salvaguardar la equidad y luchar contra todo tipo de discriminación, adoptando un enfoque inclusivo para garantizar que los beneficios de la IA sean accesibles para todos.

Asimismo, se insta a crear marcos regulatorios que incorporen los principios de la Recomendación, mecanismos de supervisión, auditoría, trazabilidad y explicabilidad, así como figuras de responsables independientes de ética de la IA y redes de colaboración apoyadas por la UNESCO. Se propone un ecosistema digital inclusivo, con infraestructura y mecanismos para compartir conocimiento, y la participación de todos los Estados, especialmente los países menos adelantados, en debates de gobernanza internacional mediante apoyo financiero y logístico.

En medio ambiente y ecosistemas, la Recomendación exige evaluar el impacto ambiental de la IA, reducir su consumo energético y huella de carbono, y aprovecharla para proteger biodiversidad, mejorar gestión de recursos naturales, monitoreo climático y prevención de desastres, evitando nuevos daños⁶⁶.

La segunda Recomendación responde al gran avance en el tratamiento de los trastornos neurológicos que se está expandiendo a nuevas áreas más allá de la atención sanitaria. La neuro tecnología plantea igualmente serias preocupaciones éticas, especialmente porque puede acceder y manipular la actividad cerebral, revelando información personal sobre la identidad, las emociones y los pensamientos y cuando se combina con la inteligencia artificial, plantea serios riesgos para la dignidad humana, la autonomía y la privacidad mental.

Adoptada el 5 de noviembre de 2025, la Recomendación de la UNESCO sobre la ética de la neuro tecnología⁶⁷, está destinada a guiar a los Estados miembros y a todos los actores implicados en el desarrollo, uso y gobernanza de estas tecnologías a lo largo de todo su ciclo de vida, con base en el derecho internacional de los derechos humanos.

La neuro tecnología tiene el potencial de definir la próxima frontera del progreso humano, pero no está exenta de riesgos. Con la adopción de este instrumento la UNESCO establece límites claros y consagra la inviolabilidad de la mente humana.

La Recomendación define neuro tecnología como aquellos dispositivos, sistemas y procedimientos (hardware y software) que acceden, observan, analizan, predicen o modulan directamente el sistema nervioso para comprender, influir, restaurar o anticipar su estructura, actividad, funciones o intenciones (lenguaje, motoras). Introduce conceptos clave: sistema nervioso, datos neuronales (datos especialmente sensibles que correlacionan directamente con estados mentales), datos biométricos cognitivos (datos neurales y no neurales usados para inferir estados mentales) y ciclo de vida completo (desde extracción de materiales y diseño hasta uso, mantenimiento, desmantelamiento, reciclaje y convergencia tecnológica).

⁶⁶ UNESCO: Recomendación sobre la ética de la inteligencia artificial

⁶⁷ En vigor desde el 12 de noviembre de 2025, al concluir la Conferencia General de la UNESCO en Samarcanda (Uzbekistán).

Aborda las cuestiones éticas que plantea la neuro tecnología por sus impactos positivos en salud y florecimiento humano, pero también por los riesgos para derechos como la autodeterminación, la privacidad mental, la identidad personal, la igualdad y la democracia. Se aplica tanto a usos médicos como no médicos (dispositivos de bienestar, neuro juegos, educación, trabajo, etc.), centrada en seres humanos, pero reconociendo consideraciones sobre uso en animales de investigación.

El nuevo instrumento de la UNESCO pide a los gobiernos que garanticen que la neuro tecnología siga siendo inclusiva y asequible, al tiempo que establecen salvaguardias para preservar la inviolabilidad de la mente humana, proteger a los más vulnerables y preservar la dignidad humana.

Además de las infracciones de la privacidad mental, la Recomendación identifica otros riesgos, en particular para niños y jóvenes, cuyos cerebros todavía están en desarrollo, y desaconseja su uso con fines no terapéuticos. También advierte contra el uso de esta tecnología en el lugar de trabajo para monitorear la productividad o crear categorías de empleados, e insiste en la necesidad de consentimiento explícito y total transparencia⁶⁸.

La Recomendación persigue orientar el desarrollo y uso de la neuro tecnología para que sean éticos, seguros y eficaces, al servicio de personas, comunidades, sociedades, medio ambiente y ecosistemas, evitando daños presentes y futuros conforme a la Carta de la ONU y al derecho internacional de los derechos humanos.

Sus objetivos incluyen proteger y promover derechos humanos, libertades fundamentales, dignidad e igualdad (incluida la igualdad de género); integrar la ética en todas las etapas del ciclo de vida; asegurar que la neuro tecnología se base en evidencia científica fiable; ofrecer un marco universal con valores, principios y recomendaciones operativas; fomentar diálogo pluridisciplinar y pluralista; promover acceso justo y reparto de beneficios; y asegurar responsabilidad y solidaridad frente a usos indebidos. Y lógicamente se articula en torno a los valores propios de Naciones Unidas⁶⁹.

Principios éticos y derechos humanos.

⁶⁸ Working Document 27 august 2024 First Draft of a Recommendation on the Ethics of Neurotechnology (Revised Version). SHS/BIO/AHEG-Neuro-2/2024/1 Paris, Agosto 2024

⁶⁹ Respeto, protección y promoción de derechos y dignidad humanos; fomento de la salud y el bienestar entendidos integralmente; garantía de diversidad y equidad, con atención a neurodiversidad, grupos minoritarios, pueblos indígenas y comunidades marginadas; y protección frente al “colonialismo tecnológico” y la asimilación que amenacen diversidad cultural. También se destacan la consideración de perspectivas interculturales, el compromiso con paz, equidad y justicia social (incluida la protección de la libertad de pensamiento frente a usos coercitivos de neuro tecnología), la solidaridad mundial y cooperación internacional, la sostenibilidad (gestionando impactos ambientales y riesgos de consumo desproporcionado de recursos) y la integridad y responsabilidad, incluyendo integridad científica.

Entre los principios, se recogen los propios y específicos de la bioética como son beneficencia, proporcionalidad e inocuidad subrayando el riesgo de que la mejora no médica aumente desigualdades; junto a los propios de derechos humanos, libre determinación y libertad de pensamiento, con énfasis en el derecho a participar o retirarse en cualquier momento y en procedimientos de consentimiento informado afirmativo, dinámico y transparente; y protección de los datos biométricos neuronales y cognitivos para salvaguardar la privacidad mental frente al aumento de capacidad de recopilación y dificultades de amonificación. Se recogen igualmente los principios propios de la ética en la IA como son no discriminación e inclusividad (evitar que la neuro tecnología refuerce estereotipos o nuevas formas de discriminación), rendición de cuentas, fiabilidad y transparencia (trazabilidad, explicabilidad, evitación de sesgos, ensayos registrados y evaluados por comités de ética), además de justicia epistémica participación inclusiva, interés superior del niño y protección de generaciones futuras, y justicia mundial y social con disfrute equitativo de los beneficios del progreso científico.

En materia de políticas públicas, se pide a los Estados que prioricen investigación y despliegue para el bien público, con salvaguardias técnicas, institucionales y procedimentales que aseguren beneficios equitativos y respeto de derechos humanos y establecer prohibiciones claras de ciertos usos⁷⁰, aplicar diligencia debida en derechos humanos con evaluaciones de impacto regulares) y garantizar que el uso en sistemas judiciales se base en evidencia sólida y respete dignidad, integridad, garantías procesales, privacidad y libertad de pensamiento.

En política de datos, exige un marco jurídico sólido para recogida, tratamiento, intercambio y uso de datos neuronales y biométricos cognitivos, reconociéndolos como personales y delicados en contextos médicos y no médicos. Las salvaguardias deben incluir consentimiento informado afirmativo, minimización y limitación de finalidad, derechos de acceso, rectificación y supresión, ciberseguridad avanzada, prohibición de condicionar acceso a bienes o servicios a la entrega de estos datos y prohibición de su uso publicitario selectivo sin consentimiento explícito.

Se pide igualmente reducir la huella ecológica de infraestructuras de datos mediante minimización de datos, eficiencia energética, energías renovables y reciclaje de equipos, y fomentar innovaciones de privacidad⁷¹ y repositorios seguros para investigación con gobernanza robusta e intercambio transfronterizo seguro y armonizado.

En propiedad intelectual, se propone que los datos biométricos neuronales y cognitivos no sean objeto de derechos de propiedad, reservando la protección a compilaciones de datos que cumplan criterios estrictos y éticos y se recomiendan directrices internacionales armonizadas sobre propiedad intelectual en neuro tecnología.

⁷⁰ Interrogatorios no consentidos, armas dirigidas al sistema nervioso, control social, coerción conductual basada en creencias, identidad de género u orientación sexual, vigilancia de estados mentales.

⁷¹ Cifrado, tratamiento en el borde, anonimización, privacidad por diseño.

En ciberseguridad, se reclama cooperación internacional para establecer normas amplias que protejan *hardware*, *software* y datos en neuro tecnología, garantizando integridad, confidencialidad, seguridad y disponibilidad de datos neuronales y reforzando la confianza de usuarios. Y no olvida la comunicación, participación y alfabetización de todos ni a ciertos usuarios específicos y sus vulnerabilidades. La Recomendación trata específicamente de niños y adolescentes, personas mayores, mujeres y minorías de género, personas con discapacidad física y personas con problemas de salud mental. Para menores, pide evitar presiones implícitas o explícitas a usar neuro tecnología, garantizar consentimiento y asentimiento adecuados a la edad, co-diseñar soluciones que cuenten con niñas, niños, cuidadores y familias, vigilar estrictamente la investigación y prohibir neuromarketing y publicidad basada en datos neuronales dirigidos a este grupo.

Para personas mayores, se promueve el envejecimiento saludable con neuro tecnología integrada en cuidados habituales, diseño adaptado, respeto de la autonomía y consentimiento continuo, y directrices para que tecnologías como cuidadores robóticos complementen, no sustituyan, la interacción humana. En mujeres y género, se proponen políticas integrales de igualdad de género en todo el ciclo de vida de la neuro tecnología, investigación inclusiva, datos desagregados, entornos laborales y de investigación seguros, medidas de acción afirmativa, apoyo al liderazgo y redes de mentoría.

Respecto de personas con discapacidad física, se recomiendan políticas que aprovechen la neuro tecnología para remover barreras y promover igualdad de derechos, evaluaciones de accesibilidad de nuevos productos, incentivos para desarrollar neuro tecnología de apoyo, subsidios y cobertura de costos para dispositivos esenciales y bases de datos nacionales de recursos. Para personas con problemas de salud mental, se enfatizan la investigación, la sensibilización, la financiación para evaluación de eficacia a largo plazo y seguimiento postcomercialización, el diseño guiado por la experiencia de usuarios y defensores, y políticas para garantizar acceso oportuno y asequible a tecnologías que mejoren su calidad de vida.

Tampoco se olvida de la salud ni de la ética de la investigación. En salud insta a apoyar aplicaciones orientadas a necesidades no satisfechas en atención neurológica y de salud mental, a reforzar la solidaridad internacional y el cumplimiento de obligaciones de derechos humanos en salud, y a crear mecanismos de supervisión de efectos físicos y mentales del uso prolongado de dispositivos, con énfasis en la invasividad y reversibilidad. En la ética de la investigación pide reforzar marcos éticos, definir cualificaciones mínimas de investigadores, exigir evaluación por comités de ética acreditados de protocolos médicos y no médicos, ofrecer formación obligatoria en ética y fomentar investigación multicéntrica e intercultural⁷².

⁷² También se exige considerar todo el ciclo de vida en ensayos clínicos (incluido qué ocurre si el promotor cesa), registro obligatorio en bases de datos, participación en sistemas de notificación de eventos

No olvida tampoco los ámbitos de la educación, el trabajo y el empleo. En entornos educativos, se propone una integración cautelosa y basada en evidencia, que complemente métodos tradicionales y priorice el desarrollo integral del alumnado, incluyendo salud mental, bienestar, pensamiento crítico, creatividad, por encima de la mera mejora del rendimiento. Se plantean políticas de uso voluntario con consentimiento plenamente informado (incluyendo padres, tutores, menores), supervisión ética robusta, mecanismos de denuncia, prohibición de incentivos injustificados o sanciones a quienes no usen neuro tecnología y medidas para evitar nuevas desigualdades, junto con investigación continua sobre efectos psicológicos y cognitivos a largo plazo.

En trabajo y empleo, se establece que las políticas deben priorizar la salud y el bienestar de las personas trabajadoras, limitar el uso a aplicaciones empíricamente validadas para el bienestar, exigir voluntariedad y prohibir usos punitivos o de vigilancia mental. Se imponen obligaciones de transparencia a empleadores (fines, datos recogidos, riesgos), limitación estricta de finalidad, minimización y seguridad de datos, eliminación de datos cuando la relación laboral termina, prohibición de recoger datos fuera de horario laboral, derecho a copia de los propios datos e interpretaciones, y prohibición del neuro perfilado en contratación y de decisiones laborales discriminatorias basadas en datos neuronales o cognitivos.

En consumo y comercio, se pide un marco regulatorio dinámico que alinee innovación con protección de derechos, bienestar y prevención de daño, asegurando uso consensuado y mecanismos contra manipulación o angustia psicológica y se extiende el principio de consentimiento informado exhaustivo a contextos como deporte y artes, evitando usos coercitivos y respetando autonomía y derechos colectivos. Además, se proponen políticas para prevenir usos abusivos en neuro juegos y dispositivos que explotan el sistema de recompensa dopaminérgico o fomentan consumo excesivo, con etiquetado de riesgos, normas de diseño adaptadas a la edad y protección frente a adicciones, especialmente en menores.

El documento entra en detalle sobre prácticas persuasivas y prohíbe o regula estrictamente el uso de datos neuronales y cognitivos en sistemas de recomendación con fines manipulativos, *nudging* o empujón encubierto (especialmente en política, publicidad y salud), *marketing* durante el sueño, neuromarketing opaco y entornos de retroalimentación continua que ajusten experiencias en tiempo real. En todos estos casos exige consentimiento explícito, derecho a no participar, transparencia, limitación de finalidad y salvaguardias contra vigilancia no autorizada, influencia política indebida y explotación de vulnerabilidades psicológicas.

adversos, validación rigurosa de algoritmos de IA (sesgos, explicabilidad, procedencia de datos), estudio de efectos en experiencia subjetiva e identidad, auditorías periódicas de prácticas de investigación y protocolos claros para comunicar hallazgos incidentales clínicamente significativos respetando autonomía y preferencias de los participantes.

El apartado relativo a las “mejoras” aborda el uso de neuro tecnología para aumentar memoria, atención u otras capacidades fuera del contexto médico, destacando los retos éticos, sociales y jurídicos que puede crear nuevas disparidades globales. Se pide a los Estados que sus leyes y políticas en esta materia eviten agravar desigualdades o discriminación, aborden riesgos de reversibilidad, invasividad y libre determinación y respeten plenamente derechos humanos y dignidad.

Y, finalmente, se establecen obligaciones de aplicación, seguimiento y cooperación. Los Estados deben promover de forma creíble la ética de la neuro tecnología, monitorizar y evaluar políticas y mecanismos y, entre otros, desarrollar capacidades institucionales,

La UNESCO se compromete a difundir la Recomendación y apoyar a los Estados mediante una metodología para evaluar el grado de preparación; otra para evaluar el impacto ético de la neuro tecnología basada en investigación rigurosa y derechos humanos; herramientas para evaluar eficacia de políticas e incentivos; un programa de investigación y un observatorio sobre ética de la neuro tecnología, incluyendo su convergencia con IA y tecnología cuántica; y una plataforma de colaboración y un foro mundial sobre ética de tecnologías emergentes, apoyado por una red de expertos regionalmente equilibrada. Se indica que la Recomendación debe interpretarse de forma integral y que ninguna de sus disposiciones puede entenderse como disminución de obligaciones internacionales existentes ni como autorización para actividades contrarias a derechos humanos, dignidad humana y protección del medio ambiente y los ecosistemas⁷³.

4. RIESGOS ESPECIFICOS EN CASO DE CONFLICTO ARMADO

Como ya se ha señalado los avances tecnológicos, cuando no son el resultado de la búsqueda de una mayor eficacia en la letalidad de las armas, pronto encuentran aplicaciones útiles para aquellos que participan en conflictos armados. En el caso de los sistemas IA y, como ya se ha visto, cambian incluso la naturaleza de las amenazas y los contornos del conflicto.

4.1 Amenazas híbridas, conflictos cibernéticos

De acuerdo con la definición comúnmente aceptada en la Unión Europea, una amenaza híbrida es una acción coordinada que ataca deliberadamente las vulnerabilidades de Estados democráticos mediante una combinación de medios políticos, económicos, militares, civiles y de información, explotando áreas grises entre guerra y paz. Por su parte, la guerra híbrida implica el uso abierto de la fuerza militar junto a medios

⁷³ Anteproyecto de Recomendación sobre la ética de la neuro tecnología. UNESCO SHS/BIO/AHEG-Neuro/2024/2.

no convencionales. Por lo tanto, las amenazas híbridas no comportan necesariamente uso de fuerza directa, sino que se centran en tácticas encubiertas que socavan el sistema sin declararse formalmente ningún conflicto.

Estas amenazas pueden incluir manipulación informativa y desinformación, ciberataques a infraestructuras críticas, coerción económica y diplomática, apoyo a actores no estatales, presión migratoria, interferencia electoral. El objetivo principal de las mismas es socavar la confianza ciudadana y la cohesión social e influir en decisiones políticas e institucionales para favorecer los objetivos estratégicos del autor de la amenaza.

Junto a la amenaza, surgen también operaciones ofensivas y defensivas bien de carácter estatal directo o bien por realizadas por actores no estatales respaldadas por Estados en el ciberespacio, destinadas a dañar, interrumpir o alterar sistemas de información críticos. Son las guerras cibernéticas o ciberguerras que se libran con armas como *malware*⁷⁴, *ransomware*, DDoS, *exploits* de día cero y que cuentan con grandes ventajas en relación con las guerras convencionales ya que permiten el anonimato, las acciones transfronterizas, una rápida ejecución y una baja inversión en relación con el alto impacto. Sus objetivos se centran, junto al clásico espionaje, sabotaje, en las infraestructuras críticas tanto de energía, transporte, salud o finanzas con repercusiones y daño directo a toda la población.

En este contexto, de amenazas híbridas y ciberguerras, IA opera como instrumento multiplicador. Efecto multiplicador y escalado: la IA reduce la carga humana y acelera los ataques, aumentando el alcance y la precisión en diferentes ámbitos (cibernético, informativo, psicológico).

Por un lado, potencia la desinformación y las operaciones cognitivas ya que los sistemas de IA permiten crear desinformación hiper-personalizada dirigida a segmentos específicos, incrementando la polarización y minando la confianza social.

Por otro, los modelos de lenguaje y redes sociales apoyados por IA pueden diseñar, difundir y reforzar narrativas hostiles de manera veloz y a gran escala por lo permite automatizar campañas de influencia pública.

Todo comparta una difuminación de fronteras entre lo digital y lo real pues los sistemas de IA permiten que los ataques digitales generen impactos físicos o sociales reales, operando en una zona gris caracterizada por la disrupción encubierta.

⁷⁴ En 2025, se han identificado amenazas con *malware* impulsado por IA que cambia su código para evadir detección.

Frente a lo anterior la defensa efectiva requiere un enfoque integral y escalable, combinando IA para detección y respuesta, normas internacionales, colaboración pública-privada, y preparación institucional que permitan superar las dificultades que surgen para regular y responder jurídicamente de forma adecuada a la rápida evolución de la IA.

Esta realidad es estudiada y explicada por diversos autores. Así, el informe *Wired for War: How authoritarian states are weaponizing AI against the West*⁷⁵, analiza cómo regímenes como los de China y Rusia utilizan la IA y tecnologías emergentes para amplificar la manipulación informativa extranjera e interferencia (FIMI⁷⁶ por sus siglas en inglés).

Examina por tanto el auge de la guerra híbrida y cognitiva, identifica actores, tácticas y vulnerabilidades democráticas, y propone estrategias para democracias occidentales, destacando la necesidad de resiliencia transatlántica ante la erosión de la confianza pública y la democracia.

La guerra cognitiva altera procesos cognitivos enemigos mediante ciber herramientas, explotando sesgos para distorsionar percepciones, decisiones y acciones a nivel individual/colectivo, sincronizada con otros instrumentos de poder. Tecnologías persuasivas como la IA generativa, neuro tecnología, *wearables* interactúan íntimamente con la mente y el cuerpo, superando con creces las velocidades humanas recurriendo a plataformas digitales y redes sociales y de mensajería, usados por prácticamente todos, como sus vectores principales.

La rápida evolución de las tecnologías y la importante inversión que los gobiernos autoritarios realizan en ellas deberían servir como llamada de atención para los Estados democráticos puesto que, de acuerdo con el informe, la red de Estados autoritarios está creciendo en fuerza y número, lo que conduce a un declive de la democracia y las libertades en todo el mundo. La manipulación y la injerencia de la información extranjera son tanto la causa como la consecuencia de este declive, ya que los actores maliciosos perturban el espacio informativo, avivan la polarización y socavan la confianza en los procesos e instituciones democráticos⁷⁷.

⁷⁵ Escrito por Matthews, Kyle and Marie Lamensch. *Wired for War. How authoritarian states are weaponizing AI against the West*. Copyright © 2025 by Konrad-Adenauer-Stiftung.

⁷⁶ FIMI se define como patrones mayoritariamente no ilegales de comportamiento manipulador, intencional y coordinado por actores estatales o no estatales (incluidos *proxies*) que amenazan valores, procesos y procedimientos políticos, exacerbando tensiones sociales y geopolíticas. No se centra en contenido falso per se, sino en conductas observables, operando en múltiples idiomas/plataformas

⁷⁷ Matthews, Kyle and Marie Lamensch. *Wired for War. How authoritarian states are weaponizing AI against the West*. Copyright © 2025 by Konrad-Adenauer-Stiftung. En su intervención ante la Comisión de Relaciones Exteriores del Senado de los Estados Unidos en marzo de 2023, Damon Wilson, presidente y director ejecutivo de la Fundación *National Endowment for Democracy*, declaró que “con Rusia y China a la vanguardia, las potencias autoritarias se han vuelto cada vez más asertivas y ambiciosas... y en una era

Los gobiernos democráticos deben actuar ahora con urgencia y con una intención estratégica. Las fortalezas que definen a las sociedades abiertas —medios de comunicación independientes, espacio cívico, innovación tecnológica y Estado de derecho— deben aprovecharse de manera decidida para exponer, perturbar y disuadir las operaciones de información hostiles. Esto significa reinvertir en los ecosistemas mediáticos internacionales donde los regímenes autoritarios han llenado el vacío, proporcionando un apoyo sostenido a las organizaciones de la sociedad civil que supervisan y se enfrentan a la guerra cognitiva, y forjando asociaciones más profundas y operativas con el sector privado y las plataformas tecnológicas⁷⁸.

4.2. Inteligencia artificial y derecho internacional humanitario

El Comité Internacional de la Cruz Roja (CICR), entre cuyas funciones se encuentra la promoción y supervisión del Derecho Internacional Humanitario (DIH) ha abordado en distintos documentos el uso de la inteligencia artificial (y el aprendizaje automático en los conflictos armados, desde una perspectiva centrada en las personas afectadas, analizando las aplicaciones militares y humanitarias de la IA, sus posibles consecuencias humanitarias y las obligaciones jurídicas y éticas que deben guiar su desarrollo y uso.

Ha de subrayarse que se aboga por un enfoque “centrado en el ser humano”, que preserve el control humano significativo sobre tareas y decisiones que afectan gravemente a la vida y la integridad de las personas y por el control humano, con base jurídica y ética.

Desde el punto de vista jurídico, las normas del DIH se dirigen a seres humanos, que son quienes hacen los juicios contextuales exigidos (distinción, proporcionalidad, precauciones en el ataque) y quienes responden por las violaciones. Éticamente, la necesidad de preservar la dignidad y la “conciencia pública”, en línea con los principios éticos de IA, que también subrayan la centralidad del ser humano y la supervisión significativa, en especial en decisiones sobre el uso de la fuerza y la detención.

Son tres los ámbitos principales de aplicación militar de la IA en los conflictos armados. Por un lado, el aumento de la autonomía en sistemas robóticos entre los que se encuentran los sistemas armas autónomas (AWS por sus siglas en inglés). Por otro

de interconectividad global, reconocen que mantener a sus propios ciudadanos bajo control ya no es suficiente para consolidar su poder, por lo que se están asociando con otras autocracias afines para compartir ideas, recursos y tecnologías”.

⁷⁸ *Ibid.*

lado, las nuevas formas de ciber- información y, finalmente, los sistemas de apoyo a la decisión militar basados en grandes volúmenes de datos⁷⁹.

Cada uno de estos ámbitos generan riesgos específicos para civiles y combatientes fuera de combate, así como desafíos para lograr un cumplimiento efectivo del derecho internacional humanitario.

Comenzando por el último, el apoyo a la decisión militar y el riesgo de “individualización” de la guerra, los sistemas de IA se usan para analizar datos masivos y apoyar decisiones sobre objetivos, detención, estrategia u operaciones, lo que puede mejorar la observancia del DIH o, por el contrario, si se confía ciegamente en ellos, facilitar violaciones y aumentar el riesgo para civiles. Así, surge una posible “individualización” de la guerra, basada en la agregación de datos personales, obtenidos por sensores, redes sociales y/o biometría para determinar la posibilidad de individuos/objetivos/diana.

El uso de IA en capacidades cibernéticas puede aumentar la escala, velocidad y gravedad de ataques, incluyendo herramientas que encuentran y explotan vulnerabilidades o lanzan contraataques automáticos, lo que complica la protección de infraestructuras y servicios civiles. Además, la IA facilita, al igual que en tiempo de paz, campañas de desinformación y manipulación tanto por texto como por audio, imagen o vídeo, que pueden derivar en detenciones arbitrarias, discriminación o violencia contra civiles⁸⁰.

Y finalmente los sistemas de armas autónomas, el ámbito quizás más preocupante. Se entiende por sistemas de armas autónomas sistemas que seleccionan y encajan objetivos sin intervención humana después de su activación inicial, basándose en información sensor y perfiles de objetivo generales. En la práctica actual, estos sistemas se emplean primordialmente para atacar objetos militares, como misiles, barcos o radares enemigos, usualmente bajo supervisión humana en tiempo real, especialmente en entornos donde los civiles no están presentes. armas autónomas aquellas que seleccionan y atacan objetivos sin suficiente control humano, con el riesgo de pérdida de control sobre el uso de la fuerza y la imprevisibilidad de los efectos⁸¹.

Pese al aumento en la autonomía tecnológica, el DIH requiere que decisiones de ataque sean tomadas por humanos en cada caso, ya que estos deben determinar la legalidad de los ataques. La incorporación de AWS introduce riesgos significativos por su funcionamiento automático, particularmente en la identificación y selección de

⁷⁹ International Committee of the Red Cross. *Artificial Intelligence and Machine Learning in Armed Conflict: A Human-centred Approach*, 2020, 102 913, pp. 463–479.

⁸⁰ International Committee of the Red Cross. *Artificial Intelligence and Machine Learning in Armed Conflict: A Human-centred Approach*, 2020, 102 913, pp. 463–479.

⁸¹ International Committee of the Red Cross. *Autonomous Weapon Systems and International Humanitarian Law*: Selected issues. Position paper by the October 2025.

objetivos, debido a la complejidad y la imprevisibilidad de sus respuestas en escenarios dinámicos. En tanto que armas bajo el DIH, están sujetas a las prohibiciones y restricciones impuestas por este, especialmente en relación con armas que puedan ser indiscriminadas, causen sufrimiento superfluo o sean prohibidas por tratados internacionales. La fiabilidad y precisión de estos sistemas son esenciales para cumplir con el principio de distinción y evitar ataques indiscriminados o desproporcionados, especialmente en entornos urbanos donde la presencia civil es significativa.

Sin duda, el recurso a los AWS plantea desafíos en la evaluación de la legalidad de sus acciones debido a su funcionamiento autónomo, especialmente en la identificación de combatientes, prisioneros de guerra o personas que ejercen participación directa en hostilidades. La imposibilidad de que estos sistemas entiendan contextos complejos o cambien su comportamiento en tiempo real limita su uso conforme a las normas del DIH.

Por ello, el propio Comité Internacional de la Cruz Roja recomienda la adopción de un marco legal internacional que prohíba los AWS diseñados o utilizados para procesos que no puedan ser suficientemente comprendidos, predichos o explicados, además de prohibir aquellos que tengan como objetivo atacar directamente a personas. Asimismo, se propone limitar el uso de AWS a objetivos militares por naturaleza y restringir su alcance, duración y escenarios de empleo, asegurando la supervisión humana efectiva y la capacidad de intervención o desactivación en tiempo oportuno. Por tanto, emerge la necesidad de un instrumento jurídico vinculante que prohíba los AWS que puedan violar principios esenciales del DIH y restrinja otros tipos tecnológicos, en un marco de protección de la humanidad y respeto a las normas humanitarias y armonizar avances tecnológicos con los principios de humanidad y legalidad del DIH. Es imprescindible reforzar el control humano y adaptar el marco legal internacional para que la IA en conflictos respete la dignidad humana y el orden legal establecido⁸².

Pero también en este ámbito los sistemas IA pueden tener efectos muy positivos dando lugar a posibles “usos humanitarios de la IA”. Pueden contribuir a mejorar el análisis de necesidades humanitarias, la protección de civiles y la prestación de servicios, por ejemplo, mediante análisis de imágenes satelitales, detección de daños, seguimiento de desplazamientos o identificación de personas desaparecidas. Aunque estos usos pueden resultar beneficiosos, plantean igualmente riesgos de privacidad, protección de datos, derechos humanos y rendición de cuentas, por lo que se insiste en el principio de “no hacer daño” también en el entorno digital.

⁸² *Ibd.*

En definitiva, los sistemas IA y el aprendizaje automático pueden, también, transformar profundamente el conflicto armado y la acción humanitaria⁸³.

4.4. La preocupación del consejo de seguridad de Naciones Unidas

En los últimos años, el rápido avance de la IA ha llevado al Consejo de Seguridad de las Naciones Unidas, máximo responsable del mantenimiento de la paz y seguridades internacionales, a prestar más atención a los riesgos de la inteligencia artificial, especialmente en contextos militares, de seguridad y desinformación, habiendo celebrado ya varias reuniones formales y algunos encuentros informales bajo la Arria, sobre IA y seguridad.

Se suma con ello a otros órganos, como la Asamblea General y grupos de expertos sobre armas autónomas letales, que también han tratado el impacto de las tecnologías de IA, especialmente en el ámbito militar y en relación con el derecho internacional humanitario y los derechos humanos. Así, cabe recordar que la Asamblea General aprobó, en marzo de 2024 una resolución sobre la IA⁸⁴ en el ámbito militar⁸⁵ que reafirma la aplicabilidad plena del derecho internacional y subraya la necesidad de un uso responsable y centrado en el ser humano. y la promoción de sistemas de IA seguros y fiables que beneficien también al desarrollo sostenible para todos⁸⁶. Y, como se ha analizado en epígrafes anteriores, en la Cumbre del Futuro de 2024 se adoptó el Pacto para el Futuro con el Pacto Digital Global como anexo de este, que prevén, entre otras cosas, la creación de un Panel Internacional Científico sobre IA y un Diálogo Global sobre la gobernanza de la IA.

Con alguna reunión previa, como la celebrada el 19 de diciembre de 2024, en la que el Secretario General António Guterres alertó que la IA avanza más rápido que la capacidad humana para gobernarla, amenazando el control humano sobre sistemas de armas, el 24 septiembre de 2025, tuvo lugar en el Consejo de Seguridad de las Naciones Unidas un debate abierto de alto nivel sobre la inteligencia artificial bajo el tema “Mantenimiento de la paz y la seguridad internacionales, con exposiciones del Secretario General de la ONU, António Guterres, y de destacados expertos académicos en IA.

⁸³ International Review of the Red Cross. *Artificial Intelligence and Machine Learning in Armed Conflict: A Human-centred Approach*, 2020, 102 913, pp. 463–479.

⁸⁴ La Asamblea General adoptó el proyecto de resolución liderado por Estados Unidos, sin votación con el copatrocinio de r más de 120 Estados miembros.

⁸⁵ La Asamblea General adoptó el proyecto de resolución liderado por Estados Unidos, sin votación con el copatrocinio de r más de 120 Estados miembros. Reconoció igualmente el potencial de los sistemas de IA para acelerar y permitir el progreso hacia la consecución de los 17 Objetivos de Desarrollo Sostenible.

⁸⁶ La Asamblea General también reconoció el potencial de los sistemas de IA para acelerar y permitir el progreso hacia la consecución de los 17 Objetivos de Desarrollo Sostenible

En el mismo el secretario general afirmó que “La inteligencia artificial sin supervisión humana dejaría al mundo ciego y quizás más peligrosa e imprudentemente que en cualquier otro lugar que en la paz y la seguridad global”⁸⁷, advirtiendo del peligro de los *deepfakes*, que pueden desestabilizar sociedades y provocar crisis diplomáticas y enfatizando la urgencia de evitar la integración de IA con los sistemas de armas nucleares. Solicitó una cooperación global sin precedentes y la creación de una un Grupo Científico Internacional sobre IA y de un Diálogo Global Dialogue sobre la gobernanza de la IA, siguiendo los compromisos adquiridos en el Pacto Digital global, examinado en un epígrafe anterior.

El debate se centró en las implicaciones de la IA para la paz y la seguridad internacionales, con el objetivo de encontrar medios para reducir riesgos y maximizar beneficios, así como compartir buenas prácticas, especialmente en sus aplicaciones militares., tratando de responder a algunas de las preguntas que habían sido planteadas en la nota conceptual sobre cómo los Estados pueden utilizar y orientar el desarrollo de la IA para promover la paz y la seguridad internacionales, y cómo el Consejo puede apoyar la aplicación responsable de estas tecnologías conforme al derecho internacional y al apoyo a procesos de paz y prevención de conflictos.

Se debatieron igualmente los considerados riesgos más inmediatos derivados del uso indebido de la IA y cómo el Consejo de Seguridad podría abordarlos con sus herramientas y mandatos, así como la forma de reforzar las iniciativas existentes en la ONU y en otros foros internacionales para un desarrollo responsable de la IA.

En todo caso, la mayoría de los Estados miembros que participaron en el debate reconoció que la IA puede mejorar las operaciones de paz, la prevención de conflictos y la respuesta humanitaria, pero también que, si se usa de forma indebida, puede amenazar la paz y la seguridad internacionales.

Aun así, se pusieron de manifiesto diferencias sobre el papel concreto que debe desempeñar el Consejo de Seguridad en relación con la IA. Algunos Estados se mostraron partidarios de que el Consejo tuviera un papel más proactivo mientras que otros consideraron que las discusiones habrían de desarrollarse principalmente en la Asamblea General y foros especializados, para evitar duplicidades y decisiones prematuras.

En todo caso, “La ventana de oportunidad para dar forma a la IA —al servicio de la paz, la justicia y la humanidad— se está cerrando. Debemos actuar sin demora” afirmó el secretario general que destacó igualmente la doble naturaleza de la IA por ser útil para detectar crisis humanitarias, pero también susceptible de ser utilizada como arma. Además, reafirmó su llamamiento a prohibir sistemas autónomos letales sin

⁸⁷ <https://news.un.org/es/story/2025/09/1540485>.

control humano y declaró que decisiones sobre armas nucleares deben ser humanas, no algorítmicas y abogó por un marco regulatorio global coherente, protección de la integridad informativa y reducción de la brecha de capacidades

En el debate se identificaron los riesgos clave en consonancia con lo expuesto en este trabajo. Sistemas autónomos letales sin control humano, especialmente en relación con la selección de objetivos; ciberataques a infraestructuras críticas con IA; *Deepfakes* y desinformación que erosionan la confianza, polarizan y pueden provocar crisis diplomáticas; las amenazas nucleares combinadas con sistemas de IA y la desigualdad tecnológica internacional, con la brecha de capacidades entre países desarrollados y en desarrollo, con impacto en la paz y la equidad ya que genera inestabilidad

En definitiva, hasta la fecha, el Consejo de Seguridad de la ONU ha dejado clara su preocupación por el uso militar de la IA, la desinformación y la necesidad de salvaguardar la supervisión humana, marcos globales reguladores, transparencia y justicia tecnológica sin adoptar medidas en relación con los riesgos identificados.

5. LA DEFENSA DE LA DEMOCRACIA Y EL ESTADO DE DERECHO: EL CONVENIO MARCO DEL CONSEJO DE EUROPA

Tal y como se ha venido desarrollando los riesgos derivados de la IA afectan tanto al disfrute de los derechos humanos como a los propios sistemas de organización social y política, democracia y estado de derecho. De ahí la relevancia de la adopción por el Comité de ministros del Consejo de Europa el 17 de mayo de 2024, del Convenio Marco sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho, primer tratado internacional legalmente vinculante destinado a garantizar que el desarrollo y uso de los sistemas de inteligencia artificial sea totalmente coherente con el respeto de los derechos humanos, la democracia y el Estado de derecho⁸⁸ que fue abierto a la firma el 5 de septiembre de ese año.

El enfoque del Convenio está basado en la gravedad y la probabilidad de un impacto negativo de los sistemas de inteligencia artificial sobre los derechos humanos, la democracia y el Estado de Derecho, sin olvidar sus positivos efectos⁸⁹.

⁸⁸ Abierto a la firma el 5 de septiembre de 2024. ara, conforme señala su artículo 30.1, los Estados miembros, los no miembros que hubieran participado en su elaboración y la Unión Europea

⁸⁹ Desde un punto de vista positivo, el Convenio destaca que el avance acelerado de la ciencia y la tecnología y de los profundos cambios provocados por las actividades en el marco del ciclo de vida de los sistemas de inteligencia artificial, tienen potencial para promover la prosperidad humana, así como el bienestar individual y social, el desarrollo sostenible, la igualdad de género y el empoderamiento de todas las mujeres y niñas, así como otros objetivos e intereses importantes, reforzando el progreso y la innovación. Además, las actividades en el marco del ciclo de vida de los sistemas de IA pueden ofrecer

Desde el punto de vista de los riesgos y peligros, en el Convenio se expresa una triple preocupación. Por un lado, por considerar que determinadas actividades en el marco del ciclo de vida de los sistemas de IA pueden minar la dignidad humana y la autonomía individual, los derechos humanos, la democracia y el Estado de Derecho. Por otro., por los potenciales riesgos de discriminación en contextos digitales, en particular los que implican sistemas de IA, y su posible efecto al crear nuevas desigualdades o agravar las ya existentes. Y, en tercer lugar, por el posible uso indebido de los sistemas de IA y el riesgo de uso de estos sistemas con fines represivos, en violación de las normas internacionales de protección de los derechos humanos, entre otros mediante prácticas arbitrarias o ilegales de vigilancia y censura que minan la privacidad y la autonomía individual.

Por todo ello, el objeto y finalidad el tratado es asegurar que las actividades relacionadas con la IA sean coherentes con los derechos humanos, la democracia y el Estado de Derecho.

El Convenio abarca todas las actividades en el marco del ciclo de vida de los sistemas de inteligencia artificial que son susceptibles de interferir en los derechos humanos, la democracia y el Estado de Derecho. En su ámbito de aplicación incluye tanto las actividades llevadas a cabo por autoridades como por agentes privados que actúen en su nombre para evitar riesgos e impactos negativos, dejando un amplio margen de discrecionalidad a los Estados para alcanzar el resultado buscado.

Por otro lado, tres áreas concretas quedan excluidas del ámbito de aplicación que pueden abrir ciertas brechas o lagunas en la protección buscada, pero que se corresponden en última instancia con las exenciones habituales en materia de protección de derechos humanos, orden público y seguridad nacional, ciertamente no exentas de dificultades en su interpretación y aplicación y sujetas a limitaciones.

El Convenio no es aplicable a las actividades en el marco del ciclo de vida de los sistemas de IA relacionadas con la protección de los intereses de seguridad nacional en un Estado parte en el entendimiento de que dichas actividades se llevan a cabo de manera coherente con el derecho internacional, incluidas las obligaciones del derecho internacional de protección de los derechos humanos, y con respeto a sus instituciones y procesos democráticos. Ni tampoco a ninguna de las actividades relacionadas con la defensa nacional, quedando por ello fuera del ámbito del Convenio, por ejemplo, el desarrollo y producción de las conocidas armas autónomas.

Finalmente, tampoco se aplicará a las actividades de investigación y desarrollo relativas a sistemas de IA que aún no se hayan puesto a disposición para su uso, a menos

oportunidades sin precedentes para proteger y promover los derechos humanos, la democracia y el Estado de Derecho.

que las pruebas o actividades similares se lleven a cabo de manera que puedan interferir en los derechos humanos, la democracia y el estado de derecho. Se trata de una excepción importante que genera un gran interrogante respecto a cómo se van a identificar esas actividades de investigación y desarrollo potencialmente dañinas y que pudieran poner en riesgo los derechos humanos, la democracia y el estado de derecho.

El Convenio recoge expresamente el principio de no discriminación, así como la atención a las necesidades y vulnerabilidades específicas en relación con el respeto de los derechos de las personas con discapacidad y de los niños.

Otros aspectos importantes que busca proteger son el debate y la consulta pública que se deberá garantizar con todos los países implicados en las cuestiones importantes que surgen, así como la alfabetización y capacidades digitales que los Estados deberán promover en todos los segmentos de la población.

Las principales obligaciones para los Estados son, por un lado, de carácter general, otras formuladas en relación con los principios, y las relativas a las garantías procesales, mitigación de riesgos y en relación con el Mecanismo de seguimiento y cooperación.

Como obligaciones generales se establece, en primer lugar, la de adoptar y mantener medidas para que todas las actividades a lo largo de todo el ciclo de vida sean coherentes con las obligaciones vigentes de protección de los derechos humanos, tanto a nivel internacional universal y regional europeo y americano como nacional. Y, en segundo, la misma obligación para garantizar que los sistemas de IA no socaven la integridad de los procesos democráticos y respeto del Estado de Derecho con referencia específica al principio de separación de poderes, el respeto de la independencia judicial y el acceso a la justicia y que los procesos democráticos queden igualmente protegidos. Por ejemplo, medidas generales de ciberseguridad contra las injerencias extranjeras maliciosas en el proceso electoral o medidas para hacer frente a la difusión de información errónea y desinformación. No obstante, en consonancia con la línea general del Convenio, no se especifica mucho más, entendiendo que la referencia a instituciones y procesos democráticos comprende todos los sistemas de gobierno con ciertas características básicas e instituciones que son común a todos los países democráticos.

Por otro lado, el grueso de las obligaciones para las Partes está formuladas en términos de principios generales comunes que deben aplicarse en relación con los sistemas de IA de forma adecuada a los respectivos ordenamientos jurídicos internos de forma que dichos principios sean respetados. Con una formulación tan laxa de valores esenciales a proteger, el margen de discrecionalidad de los Estados es muy amplio. Sin duda, lograr el consenso respecto a esos valores a proteger supone un importante paso si bien el carácter de principios dificulta enormemente la concreción de su contenido, que podrá variar en función de las características y normas de cada uno de los Ordenamientos nacionales.

En concreto, se identifican los principios para la ejecución de las actividades, se recogen medidas correctivas y criterios para la gestión de riesgos e impactos.

Así, los principios que han de ser protegidos son la dignidad humana y la autonomía individual; la transparencia y supervisión, la rendición de cuentas y responsabilidad ante efectos adversos, la igualdad y no discriminación, la privacidad y protección de los datos personales como principio protector del derecho a la intimidad y, finalmente, la fiabilidad e innovación segura. Estos principios coinciden en esencia con los recogidos en la ya referida Recomendación sobre la ética de la inteligencia artificial aprobada por la UNESCO.

Algunos de estos principios constituyen elementos esenciales del sistema internacional de protección de derechos humanos. Es el caso del principio de dignidad humana, fundamento sobre el cual se construyen los derechos humanos o del principio de autonomía individual, derecho humano fundamental relativo a la capacidad de cada persona para tomar decisiones libres y responsables sobre su propia vida, estrechamente vinculada con la dignidad humana y esencial para el ejercicio de otros derechos y libertades.

Otros son igualmente principios comunes y presentes en todos los sistemas de protección de derechos humanos, como son los de igualdad y no discriminación, privacidad y protección de los datos personales e, incluso, rendición de cuentas y responsabilidad ante efectos adversos.

No obstante, otros principios son más específicos y se pueden considerar como propios de este nuevo ámbito surgido de la proliferación de los sistemas de IA o adquieren un significado específico y particular en el mismo. Este es el caso, por ejemplo, del principio de transparencia y supervisión, del que el Convenio especifica que opera igualmente en lo que respecta a la identificación de contenidos generados por los sistemas de inteligencia artificial. Así, en relación con la transparencia, que en el caso de la Recomendación de la UNESCO va unida a la explicabilidad de los sistemas de IA, se afirma que son condiciones previas fundamentales para garantizar el respeto, la protección y la promoción de los derechos humanos. Es el principio que subraya la importancia de que los sistemas de IA sean comprensibles y sus decisiones puedan ser explicadas de manera clara. Esto ayuda a generar confianza y permite a los usuarios entender cómo y por qué se toman ciertas decisiones. Sin duda alguna, esta transparencia es necesaria para que los regímenes nacionales e internacionales propios de la responsabilidad funcionen eficazmente. La falta de transparencia también podría mermar la posibilidad de impugnar eficazmente las decisiones basadas en resultados producidos por los sistemas de IA y podría, por ejemplo, vulnerar el derecho a un juicio imparcial y a un recurso efectivo. Por ello, supone limita los ámbitos en los que estos sistemas pueden utilizarse legalmente.

Por último, las medidas para promover la fiabilidad de los sistemas de inteligencia artificial y la confianza en sus resultados que deben adoptar los Estados Parte en el Convenio del Consejo de Europa hacen referencia a la capacidad de los sistemas de IA para funcionar de manera consistente, segura y ética. La seguridad y robustez significa que los sistemas de IA deben ser diseñados para minimizar riesgos y evitar daños; por otro la transparencia y explicabilidad y finalmente la responsabilidad, con la existencia de mecanismos claros de rendición de cuentas. Se puede afirmar que este principio se basa en tres pilares fundamentales: legalidad, ética y robustez. El primer pilar implica un escrupuloso cumplimiento de la legalidad, el segundo asegurar que los sistemas respeten los valores y principios éticos y el, tercero, garantizar que los sistemas sean técnicamente sólidos, minimizando los riesgos de daños accidentales o malintencionados. Es precisamente este punto al que hacer especial referencia el principio de innovación segura, que se refiere específicamente a la necesidad de desarrollar y desplegar tecnologías de IA de manera que se minimicen los riesgos y se maximicen los beneficios para la sociedad

Finalmente, en relación con la innovación segura el Convenio pide a cada Parte que permita, según proceda, el establecimiento de entornos controlados para desarrollar, experimentar y probar sistemas de inteligencia artificial bajo la supervisión de sus autoridades competentes. A ello también contribuye la evaluación y mitigación de riesgos e impactos adversos, contemplada igualmente.

En todo caso, la implementación de estas obligaciones establecidas se hará por cada Parte aplicando sus propios marcos nacionales al contexto de los sistemas de inteligencia artificial, teniendo presente el objeto y fin del Convenio Marco, que no es otro que garantizar que las actividades dentro del ciclo de vida de los sistemas de inteligencia artificial sean coherentes con los derechos humanos, la democracia y el estado de derecho. Por ello, se describen los mecanismos y procedimientos disponibles para que las personas y grupos busquen reparación por violaciones de sus derechos relacionados con los sistemas de IA. Se garantiza así el acceso a la información relevantes sobre los sistemas de IA, la existencia de mecanismos efectivos para resolver quejas y disputas, con disposición de recursos judiciales y medios alternativos de solución de diferencias.

Los Estados deberán garantizar la disponibilidad de esos recursos legales para las víctimas de violaciones de los derechos humanos relacionadas con el uso de sistemas de IA y de garantías procesales, incluida la notificación a todas las personas que interactúen con sistemas de IA de que están interactuando con estos sistemas. En cuanto a los riesgos para la democracia se establece que han de adoptar medidas para garantizar que los sistemas de IA no se utilicen para minar las instituciones y procesos democráticos, incluido el principio de separación de poderes, el respeto de la independencia judicial y el acceso a la justicia.

El Convenio contempla además que cada Parte ha de establecer un mecanismo de supervisión independiente para vigilar que se cumple con el convenio, que se sensibilice y estimule un debate público informado, y que se lleven a cabo consultas con múltiples partes interesadas sobre cómo debería ser utilizada la tecnología de la IA.

En la evaluación de riesgos se enfatiza la importancia de realizarlas de forma exhaustiva a lo largo de todo el ciclo de vida de los sistemas de IA, lo que implica un monitoreo continuo, de forma que incluyan la identificación de posibles riesgos para los derechos humanos, la democracia y el estado de derecho, teniendo en cuenta la gravedad de estos y la probabilidad de esos posibles impactos.

Por otro lado, se contempla la necesidad de implementar medidas de mitigación efectivas para abordar los riesgos identificados, lo que implicaría la adopción de medidas técnicas así organizativas para minimizar los impactos adversos.

En definitiva, no se oculta que el enfoque del Convenio está basado en la gravedad y la probabilidad de un impacto negativo de los sistemas de inteligencia artificial sobre los derechos humanos, la democracia y el Estado de Derecho y para ello considera que se requiere una metodología que oriente y ayude a identificar los contextos y aplicaciones en los que el despliegue de sistemas de IA podría presentar riesgo, así como a evaluar y mitigar dichos riesgos.

Si bien existen algunos métodos para evaluar esos riesgos y el impacto sobre los derechos humanos el principal objetivo del Comité de Inteligencia Artificial del Consejo, tras la adopción del Convenio fue desarrollar esa tecnología.

Finalmente, el Convenio contempla un mecanismo de supervisión y seguimiento por el que corresponde a la Conferencia de las Partes y los Estados podrá, entre otras acciones, consultarse periódicamente con el fin de facilitar la aplicación efectiva del Convenio, intercambiar información y/o estudiar la posibilidad de modificaciones. Además, será la receptora de los informes que los Estados se comprometen a presentar. En cuanto a la supervisión a nivel nacional se fija para los Estados la obligación de establecer o designar uno o varios mecanismos eficaces para supervisar el cumplimiento de las obligaciones del Convenio, que ejercerán sus funciones de manera independiente e imparcial⁹⁰.

⁹⁰ Cano Linares, M.^a Ángeles: “El convenio marco del Consejo de Europa sobre I.A., Derechos Humanos, democracia y Estado de derecho: ¿riesgo real?” en *Derechos humanos e inteligencia artificial: una aproximación jurídica multidisciplinar* / coord. por Jesús María Navalpotro Sánchez-Peinado, M^a José Lavilla Pons, Alejandra Cobo del Rosal; Gabriela Cobo del Rosal Pérez (dir.), 2025, págs. 57-61; Díaz Galán, Elena Carolina: “La regulación jurídica de la inteligencia artificial (IA) en Europa: dos pasos decisivos, aunque insuficientes, del Consejo de Europa y la Unión Europea”, *Cuadernos de derecho transnacional*, I, Vol. 17, N^o. 1, 2025, pp. 366-390.

Por último, deben destacarse tres aspectos sobresalientes del Convenio Marco. Se trata de su naturaleza de norma jurídica vinculante, su vocación universal y su condición de Convenio marco y enfoque de carácter protector y preventivo.

El primer rasgo distintivo es el ya señalado en cuanto a su naturaleza, es un tratado internacional y por tanto una norma jurídica vinculante. Se trata del primer instrumento internacional jurídicamente vinculante que se adopta sobre inteligencia artificial, poniendo claramente de manifiesto, dado el enfoque del Tratado, basado en la gravedad y la probabilidad de un impacto negativo de los sistemas de inteligencia artificial sobre los derechos humanos, la democracia y el Estado de Derecho, los potenciales riesgos de una tecnología que ya forma parte de nuestras vidas.

En segundo lugar, se trata de una norma adoptada por una organización regional, el Consejo de Europa, pero que tiene vocación universal. Ello se pone de manifiesto no solo en la participación de Estados no europeos en las negociaciones sino también en su articulado y en particular en su artículo 31 relativo a la adhesión. Al respecto, recoge la posibilidad de que el Comité de ministros invite a adherirse cualquier Estado no miembro del Consejo de Europa que no haya participado en la elaboración del Convenio, siempre y cuando tal invitación cuente con el previo consentimiento unánime de las Partes en el Convenio.

Esta vocación universal se ve facilitada por el tercer rasgo. Se trata de un Convenio Marco, un tipo de instrumento que el Consejo de Europa adopta, entre otras razones, por una flexibilidad que facilita la cooperación internacional y el establecimiento de estándares comunes. Así un Convenio marco no pierde en absoluto su carácter jurídicamente vinculante, pero deja un importante margen a los Estados Parte para adaptar algunas de las disposiciones del Convenio a cada situación específica por medio de la respectiva legislación nacional y políticas gubernamentales adecuadas. Así, el Convenio Marco sobre inteligencia artificial establece principios para proteger los derechos humanos, la democracia y el Estado de Derecho, promoviendo también la innovación tecnológica si bien no trata de un instrumento aislado, sino que representa un importante paso hacia una gobernanza de una IA responsable, ética y fiable⁹¹.

La Decisión del Consejo de la Unión Europea (UE) de 2024 autorizó la firma del Convenio Marco al considerar que encaja con el marco jurídico europeo existente. En 2018 la Comisión lanzó una estrategia europea de IA basada en tres pilares: reforzar la capacidad tecnológica e industrial y la adopción de IA; prepararse para las transformaciones socioeconómicas; y garantizar un marco ético y jurídico adecuado, siempre con la persona en el centro. La Comunicación subraya que la UE ya dispone de un

⁹¹ *Ibid.* Iermano, Anna. “Consiglio d’Europa e intelligenza artificiale: un primo tentativo di regolamentazione a tutela dei diritti umani, democrazia e Stato di diritto”, *Freedom, Security & Justice: European Legal Studies* 2025, n. 2, pp. 242-271.

sólido marco normativo que sirve de referencia mundial: el e Reglamento sobre libre circulación de datos no personales, el Reglamento de Ciberseguridad y la futura normativa de privacidad en comunicaciones electrónicas. Estos instrumentos garantizan protección de datos, privacidad y seguridad, y se complementan con la Carta de los Derechos Fundamentales y los valores fundacionales de la Unión como son la dignidad, libertad, democracia, igualdad, Estado de Derecho y derechos humanos. A este marco normativo sea de añadir el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, conocido como Reglamento de Inteligencia Artificial⁹².

Así, en la UE, la IA centrada en el ser humano⁹³ se concibe como aquella que es un medio al servicio de las personas, no un fin en sí mismo, y cuyo objetivo es aumentar el bienestar humano y empoderar a los ciudadanos, en lugar de sustituirlos. En la Recomendación, la Comisión insiste en que la simple aplicación acrítica de tecnologías de IA generaría errores, sesgos, vulnerabilidades a ciberataques y desconfianza, por lo que se necesitan principios éticos claros y directrices prácticas para desarrolladores, proveedores y usuarios, incluido el Reglamento de IA.

Al firmar el Convenio, la UE se comprometió a garantizar que sus normas internas respeten los estándares del tratado de forma que el Reglamento es la principal herramienta jurídica para cumplir ese compromiso, opera como mecanismo de cumplimiento del Convenio.

En relación de coherencia y complementariedad con el Convenio Marco del Consejo de Europa, en el marco de la Unión Europea (UE) se había aprobado, con

⁹² 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° 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.

⁹³ Grupo de expertos de alto nivel y siete requisitos esenciales

Para elaborar esas directrices, la Comisión creó un Grupo de Expertos de Alto Nivel sobre IA y una Alianza Europea de la IA, plataforma multiactor con miles de miembros para alimentar el debate. El grupo, basándose también en el trabajo del Grupo Europeo de Ética y la Agencia de Derechos Fundamentales, definió el concepto de “IA fiable” como aquella que: 1) cumple la ley; 2) respeta principios éticos; y 3) es técnicamente robusta, de donde derivan siete requisitos esenciales.

Los siete requisitos son: intervención y supervisión humanas, solidez y seguridad técnicas, privacidad y gestión de datos, transparencia, diversidad/no discriminación/equidad, bienestar social y medioambiental, y rendición de cuentas. Las directrices incluyen una lista práctica (“checklist”) para ayudar a comprobar el cumplimiento de estos requisitos, que se aplican con un enfoque proporcional al impacto: las exigencias serán mucho mayores, por ejemplo, en un sistema de diagnóstico médico que en un recomendador de libros.

anterioridad, el Reglamento (UE) 2024/1689⁹⁴ con el objetivo principal de garantizar que las aplicaciones de la IA en la UE sean seguras, éticas, fiables y respetuosas con los derechos fundamentales. Se trata de una norma derivada de la UE anterior en el tiempo al Convenio Marco, que es un tratado internacional abierto a países europeos y no europeos.

Y en efecto, el Reglamento de IA operativiza las obligaciones generales para garantizar los derechos humanos, democracia y el estado de derecho y los principios de supervisión humana, transparencia y rendición de cuentas mediante las prohibiciones específicas que recoge, como, manipulación, puntuación social, vigilancia biométrica masiva, los requisitos técnicos exigidos para sistemas de alto riesgo, las obligaciones para proveedores y usuarios y las sanciones administrativas, cubriendo ambos todo el ciclo de vida de la IA⁹⁵.

6. CONCLUSIONES

Desde que el hombre es hombre cualquier avance, descubrimiento o invento con potencial para un desarrollo favorable de la humanidad ha sido igualmente utilizado para causar daño, obtener riqueza o poder o destruir al considerado enemigo. Los ejemplos se agolpan en nuestra cabeza.

No es diferente con la inteligencia artificial y las tecnologías digitales que se han convertido hoy tanto en una herramienta transformadora como en un arma potencial de una sofisticación sin precedentes, lo que supone un grave riesgo para las personas, sus derechos, las instituciones democráticas y la seguridad nacional e internacional.

Y ello porque recurso a las capacidades de la IA con fines criminales y para desestabilizar sociedades, alterar el orden público y socavar las instituciones democráticas es cada vez mayor. La combinación única de escala y accesibilidad de la IA crea retos de seguridad sin precedentes que nuestros marcos de gobernanza actuales no parecen estar preparados para abordar.

⁹⁴ Fue aprobado el 13 de junio de 2024 y publicado en el DOUE el 12 de julio de 2024. La Comunicación COM (2019) “Generar confianza en la inteligencia artificial centrada en el ser humano” expone cómo la Comisión Europea quiere combinar liderazgo tecnológico con un enfoque ético y basado en derechos fundamentales. Presenta la estrategia europea de IA, define el concepto de “IA fiable” mediante requisitos éticos esenciales y describe una fase piloto para probar directrices en la práctica, así como la proyección internacional de este enfoque.

⁹⁵ Palladino, Rossana. “L’“approccio europeo” al contrasto alla disinformazione digitale e alla protezione dei valori democratici: quale contributo dell’AI Act?” *Freedom, Security & Justice: European Legal Studies*, 2025, n. 2, pp. 296-316.

Este nuevo entorno requiere sin duda que los Estados, y en especial los Estados democráticos, en los que rige el estado de derecho, establezcan marcos éticos, jurídicos y organizativos mucho más robustos para aprovechar beneficios y contener daños.

Para ello parece necesaria una agenda que combine regulación robusta, innovación responsable, resiliencia tecnológica y fortalecimiento democrático para que la transformación digital se alinee con los valores del estado de Derecho y no derive en nuevas formas de autoritarismo o vulneración masiva de derechos fundamentales.

En este proceso, el primer paso lógico es conocer, comprender y visibilizar los riesgos y las amenazas derivadas del uso de la IA, multifacéticos y cada vez más sofisticados. La IA ha reducido las barreras a los ciberataques, la manipulación impulsada por *deepfakes*, las campañas de desinformación, el fraude a gran escala y la ingeniería social, creando desafíos sin precedentes, tanto para los Gobiernos como para los actores privados y para la sociedad en general.

No parece posible hacer frente a estas amenazas y promover al mismo tiempo la innovación, sin un esfuerzo coordinado a nivel mundial, en el que participen tanto los Estados como el sector privado. La puesta en marcha de políticas globales que mejoren la seguridad y la protección de la IA requiere un enfoque integral y múltiple.

De ahí la importancia de una gobernanza y marco normativo de carácter universal puesto que los riesgos y amenazas de la IA trascienden las fronteras nacionales y, sin una coordinación global. Resulta necesario que todos los Estados y las grandes compañías tecnológicas se alineen en torno a principios comunes y salvaguardias aplicables. El marco de Naciones Unidas se presenta como idóneo al respecto, con modelos que se han probado eficaces en otros ámbitos⁹⁶ aplicando a la gobernanza de la IA un enfoque que requiera normas globales básicas para el desarrollo seguro de la IA; la integración de medidas de seguridad de la IA en las políticas nacionales de todas las jurisdicciones; evaluaciones y auditorías independientes realizadas por expertos para evaluar el cumplimiento; obligaciones coherentes para el sector privado en todo el mundo, que impidan el arbitraje regulatorio; y mecanismos periódicos de información y aplicación para garantizar el cumplimiento. Sin ese marco, los esfuerzos en materia de seguridad de la IA seguirán estando fragmentados, lo que deja abierto el terreno para el uso criminal de la IA.

Por ello entiendo que solo un marco internacional universal con normas aplicables para garantizar una aplicación coherente en todos los Estados puede evitar lagunas normativas y promover la competencia leal.

⁹⁶ Como el Grupo de Acción Financiera Internacional (GAFTI), que establece normas globales obligatorias para la integridad financiera y el cumplimiento por parte de los gobiernos y el sector privado.

Sería igualmente necesario facilitar, potenciar e incrementar la financiación y los incentivos para la promoción del desarrollo de la IA en los campos de la seguridad y la protección, incluida la detección de fraudes basada en la IA, la autenticación de *deepfakes* y herramientas avanzadas de ciberseguridad. Las asociaciones público-privadas deberían acelerar la investigación sobre mecanismos de defensa contra la IA adversaria. Además, se debería animar o/y exigir a los desarrolladores que integren evaluaciones de riesgos y pruebas para las nuevas aplicaciones de IA, especialmente las de alto riesgo, y que adopten mecanismos defensivos de IA que detecten y mitiguen su manipulación.

Y, en tercer lugar, igualmente imprescindible resulta la concienciación pública. La educación pública integral es fundamental para crear resiliencia frente a los riesgos y amenazas de IA. En este caso, la labor, igualmente liderada por el Estado, requiere la colaboración de la industria y del mundo académico para concienciar sobre los riesgos de la IA, en particular en lo que respecta a la interferencia electoral, los *deepfakes* y la suplantación de identidad, y los delitos financieros. Deben promover la competencia digital y el uso responsable de la IA, asegurándose de que todos los miembros de la sociedad comprendan cómo identificar y responder al engaño generado por la IA.

En definitiva, son los Estados los que tiene la obligación de no vulnerar derechos con vigilancia desproporcionada, biometría invasiva o IA opaca y de proteger a ciudadanos frente a abusos de actores privados y terceros Estados, garantizar marcos legales claros y supervisión efectiva, promover medios independientes y alfabetización digital.

Sin ningún ánimo de alarmismo, entiendo que lo que está en juego es existencial. Sin una intervención coordinada y proactiva, el potencial de la IA para crear consenso y manipular los sistemas sociales seguirá planteando un desafío sin precedentes para la seguridad mundial y las instituciones democráticas.

Un último apunte, casi un *desideratum*: paliar el incremento de la brecha digital, la desigualdad en el acceso, uso y aprovechamiento de las tecnologías digitales entre diferentes grupos sociales y entre Estados que, con la llegada y expansión de la IA se ha incrementado atentando directamente contra el Objetivo de Desarrollo Sostenible (ODS) 10 relativo a la reducción de las desigualdades. La brecha digital amplía la distancia entre países desarrollados y en desarrollo, y entre grupos socioeconómicos dentro de un mismo país. Evitar o reducir la brecha digital requiere cooperación internacional para garantizar acceso equitativo a tecnologías emergentes, en relación directa con el ODS 17: Alianzas para lograr los objetivos. En este contexto, se ha de recordar que la IA puede ser una herramienta poderosa para acelerar los ODS de forma que la IA sea utilizada de manera inclusiva y equitativa porque si no se gestiona bien, la IA puede agravar desigualdades y dificultar el cumplimiento de los ODS. Pero, usada de forma inclusiva, puede ser una herramienta poderosa para acelerar su cumplimiento.

En definitiva, conocimiento, cooperación, gobernanza, marco jurídico, financiación, alianzas público-privadas, concienciación sin olvidar la dimensión ética de la IA.

La ética en la IA no es un lujo, sino que ha de ser parte esencial del desarrollo tecnológico, y al igual que competitividad económica y la confianza social deben basarse en los mismos valores fundamentales.

Todos hemos de perseguir una IA centrada en el ser humano, fiable y respetuosa de derechos deseando que el camino ferroviario emprendido por la Unión Europea no lleve a un descarrilamiento del proceso.

Es de desear que, como se afirma en la Declaración final del debate abierto en el Consejo de Seguridad, en septiembre de 2025, “Trabajemos juntos para garantizar que la IA sirva a la humanidad y que el futuro digital sea un futuro de paz y seguridad. Construyamos un mañana digno para nuestros hijos”.

La innovación debe servir a la humanidad, no debilitarla y el progreso tecnológico solo vale la pena si está guiado por la ética, la dignidad y la responsabilidad hacia las generaciones futuras.

APÉNDICE BIBLIOGRÁFICO

- Brian Nougères, Ana. “Fundamentos y principios para la regulación de neurotecnologías y el tratamiento de neurodatos desde el derecho a la privacidad”. *Informe de la Relatora Especial sobre el derecho a la privacidad*, A/HRC/58/58, enero 2025.
- Cano Linares, M.^a Ángeles: “El convenio marco del Consejo de Europa sobre I.A., Derechos Humanos, democracia y Estado de derecho: ¿riesgo real?” en *Derechos humanos e inteligencia artificial: una aproximación jurídica multidisciplinar* / coord. por Jesús María Navalpotro Sánchez-Peinado, M^a José Lavilla Pons, Alejandra Cobo del Rosal; Gabriela Cobo del Rosal Pérez (dir.), 2025, págs. 57-61.
- Cano Linares, M.^a Ángeles, Carreras Presencio, Ana Isabel. *Nuevas realidades: un enfoque desde el derecho* 2025, Dykinson, pp. 14-34. Marta Carreras Presencio Ana Isabel. “Inteligencia artificial en el siglo XXI. Necesidad de gobernanza y ética” en *Nuevas realidades: un enfoque desde el derecho* dir. M.^a Ángeles Cano Linares, Ana Isabel Carreras Presencio, 2025, pp. 14-34.
- CFR’s Daily News Brief. *Global AI summit in Paris*. 11 de febrero de 2025.

- Consejo de Europa. *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*. Explanatory Report. 2024
- Asamblea Parlamentaria *Technological convergence, artificial intelligence and human rights*. 2018
- Dias, Bernadete André Lins, Thais Fernandes. *Neurorights: Is the creation of new human rights effective in protecting human dignity from the misuse of neurotechnology?*, International Bar Association, 2025.
- Díaz Galán, Elena Carolina: “La regulación jurídica de la inteligencia artificial (IA) en Europa: dos pasos decisivos, aunque insuficientes, del Consejo de Europa y la Unión Europea”, *Cuadernos de derecho transnacional*, I, Vol. 17, Nº. 1, 2025, pp. 366-390.
- European Parliamentary Research Service. *Person identification, human rights and ethical principles Rethinking biometrics in the age of artificial intelligence*, December 2021.
- Festa, Angela. “La disinformazione online come “minaccia ibrida” alla democrazia nell’Unione europea: meccanismi di tutela e strumenti a contrasto per uno Spazio di libertà, sicurezza e giustizia”, *Freedom, Security & Justice: European Legal Studies* 2025, n. 2, pp. 272-295.
- Fuster Gonzales, Gloria, Michalina Nadola Peeters. “Person identification, human rights and ethical principles: Rethinking biometrics in the era of artificial intelligence”, Scientific Foresight Unit (STOA), European Parliamentary Research Service, 2021.
- Galceran, Marta. “El pacto digital global y la transición responsable e inclusiva” *Informe CIDOB* 12- 2024.
- International Committee of the Red Cross. *Autonomous Weapon Systems and International Humanitarian Law: Selected issues*. Position paper by the October 2025.
- *Artificial Intelligence and Machine Learning in Armed Conflict: A Human-centred Approach*, 2020, 102 913, pp. 463–479.
- Iermano, Anna. “Consiglio d’Europa e intelligenza artificiale: un primo tentativo di regolamentazione a tutela dei diritti umani, democrazia e Stato di diritto”, *Freedom, Security & Justice: European Legal Studies* 2025, n. 2, pp. 242-271.
- Kemp, David. *What if generative AI is reaching its limits?* European Parliamentary Research Service, noviembre 2025.
- Kumar, Praachi. *Examining the Causes and Consequences of the Digital Divide(s)*. blog post UNU-MERIT 27 Sep 2024 <https://unu.edu/merit/blog-post/examining-causes-and-consequences-digital-divides>
- Levantino Francesco Paolo, Paolucci Federica. “Advancing the Protection of Fundamental Rights through AI Regulation: How the EU and the Council of Europe are Shaping the Future” *European Yearbook on Human Rights* 2024. 16 pp.

- Malgieri, Gianclaudio, *Vulnerability and Data Protection Law*, Oxford University Press, 2023.
- Matthews, Kyle and Marie Lamensch. *Wired for War. How authoritarian states are weaponizing AI against the West*. Copyright © 2025 by Konrad-Adenauer-Stiftung.
- Navalpotro Sánchez-Peinado, Jesús María, M^a José Lavilla Pons, Alejandra Cobo del Rosal; Gabriela Cobo del Rosal Pérez (dir.), *Derechos humanos e inteligencia artificial: una aproximación jurídica multidisciplinar*. Universidad Rey Juan Carlos: Instituto de Estudios Jurídicos Internacionales, 2025.
- Palladino, Rossana. “L’“approccio europeo” al contrasto alla disinformazione digitale e alla protezione dei valori democratici: quale contributo dell’AI Act?” *Freedom, Security & Justice: European Legal Studies*, 2025, n. 2, pp. 296-316.
- Palmiotto, Francesca. “When Is a Decision Automated? A Taxonomy for a Fundamental Rights Analysis”, *German Law Journal*, 2024, pp. 1–27.
- Paolucci, Federica. “From Global Standards to Local Safeguards: The AI Act, Biometrics, and Fundamental Rights” en Natalia Menéndez González and Giuseppe Mobilio (eds), *Facial Recognition Technology*, Springer, 2024.
- Programa de las Naciones Unidas para el desarrollo (PNUD). Estrategia digital 2022-2025. https://digitalstrategy.undp.org/documents/Digital-Strategy-2022-2025-Full-Document_ES_Interactive.pdf
- Kumar Praachi., “Examining the Causes and Consequences of the Digital Divide(s),” *UNU-MERIT (blog)*, 2024-09-27, <https://unu.edu/merit/blog-post/examining-causes-and-consequences-digital-divides>.
- Salamanca Aguado, Esther. “Respect for privacy and personal data protection in the context of mass surveillance of communication”, *Journal of the Spanish Institute for Strategic Studies* N. 4, 2014, pp.1-24.
- Shany, Yuval. “A Human Right to Informational Self-determination: What it is and why it matters for Digital Human Rights”; *EJIL:Talk!*, noviembre 2025.
- Suñé Cano, Juan Emilio. “La Unión Europea frente a la mentira digital” en *Nuevas realidades: un enfoque desde el derecho*, dir. M.^a Ángeles Cano Linares, Ana Isabel Carreras Presencio, 2025, pp. 52-81.
- Taeihagh Araz, “Governance of Generative AI”, *Policy and Society*, Volume 44, Issue 1, 2025, pp.1-22.
- Vielka M. Párraga. “From efficiency to exclusion rethinking AI governance in public institutions”. *Journal of Law and Epistemic Studies*, Vol. 3, N.º. 2, 2025, pp. 19-25.
- UNESCO. Unveiling the Neurotechnology Landscape Scientific Advancements Innovations and Major Trends, UNESCO Digital Library, 2023.

- Recomendación sobre la ética de la inteligencia artificial.
- UE. Comunicación de la Comisión al Parlamento Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. *Generar confianza en la inteligencia artificial centrada en el ser humano* Bruselas, 8.4.2019 COM (2019) 168 final.
- 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) nº 300/2008, (UE) nº 167/2013, (UE) nº 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).
- EU Agency for fundamental Rights (FRA). *Assessing High-risk Artificial Intelligence: Fundamental Rights Risks*. Diciembre 2025.
- Wagman, Shlomit, Hubbard Sarah. *Weaponized AI: A New Era of Threats and How We Can Counter It*, Ash Center for Democratic Governance and Innovation; Harvard Kennedy School



CAPÍTULO 3.

POSICIÓN DE CASILLA “d8”: SUPERVISIÓN HUMANA⁹⁷

ANA ISABEL CARRERAS PRESENCIO⁹⁸

Universidad Rey Juan Carlos

Resumen

Desde que la persona es persona, individuo racional y social, no ha podido evitar la desigualdad, más bien la ha creado, porque no nacemos desiguales. El ser humano para sobrevivir ha necesitado ser social, “con”vivir con los demás. Y esa convivencia ha necesitado de una organización, de un orden social, de unas normas de convivencia, justas, en justicia, a partir de un medio, el Derecho, que exteriorizado en lenguaje es ley. Pero ¿cómo hemos determinado qué es justo en esa convivencia social? Lo justo es igualar nuestra conducta a la del otro, para posicionarle en una situación de igualdad respecto de uno. Es igualar la conducta de uno para con el otro. Por tanto, la desigualdad será lo injusto, siendo que, esa conducta justa se establece, al menos de dos formas, una a través de la razón, la lógica natural, y otra a través de los usos públicos, la voluntad, la autoridad. Tenemos, entonces, al menos, dos formas de establecer Derecho, un Derecho natural y un Derecho positivo, desde lo cual, en su exteriorización, elaboramos un *ius* naturalismo y un *ius* positivista. Pues bien, la norma internacional de protección de los derechos humanos está basada en un *ius* naturalismo desde la razón y la lógica de los valores humanos. Es una norma universal y uniforme para todos los seres humanos, protegiendo la dignidad humana en igualdad. Y esa protección jurídica consiste en posicionar al ser humano en igualdad en la ley y ante la ley. Por tanto, su infracción es discriminación. Una discriminación responsabilidad del Estado, obligado por la norma internacional de los derechos humanos como delimitadora de su soberanía. Una discriminación que mantenida se convierte en estructural, por razón de sexo, género, raza, nacionalidad, religión, y más. Y ¿la inteligencia artificial?

Palabras clave

Sesgo, inteligencia artificial, género, interseccionalidad, discriminación.

⁹⁷ El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

⁹⁸ Profesora Permanente. URJC. Abogada y Consultora Académica. ORCID: 0000-0003-2293-5061

Abstract

Since the dawn of humanity, as rational and social individuals, inequality has been unavoidable; rather, it has been created, because we are not born unequal. To survive, human beings have needed to be social, to live together. And this coexistence has required organization, a social order, just rules of coexistence, and a means—Law—which, expressed in language, becomes law. But how have we determined what is just in this social coexistence? Justice lies in aligning our conduct with that of others, placing them in a position of equality with us. It is about equating one's conduct with that of another. Therefore, inequality is injustice. This just conduct is established in at least two ways: through reason, natural logic, and through public custom, will, and authority. We have, then, at least two ways of establishing Law: natural or positive, from which, in its external manifestation, we have natural law and positivist law. Well, the international standard for the protection of human rights is based on a natural law with human values. It is a universal and uniform standard for all human beings, protecting human dignity equally. And this legal protection consists of positioning all human beings as equals, under the law and before the law. Therefore, its violation is discrimination. Discrimination is the responsibility of the State and the authorities. Discrimination that, if maintained, becomes structural, based on sex, gender, race, nationality, religion, and more. And what about artificial intelligence?

Key words

Bias, artificial intelligence, gender, intersectionality, discrimination.

Sumario: INTRODUCCIÓN. 1. EL SESGO 1.1.- *El sesgo en la IA*. 1.2.- *El sesgo de género en la IA*. 2. POSICIÓN DE CASILLA “d8”. 2.1.- *La interseccionalidad*. 2.2.- *Supervisión humana. Buenas prácticas*. 3. EDUCACIÓN EN IA. 4. CONCLUSIONES BIBLIOGRÁFICA

INTRODUCCIÓN

Este trabajo está realizado en el marco de las actividades propias del Grupo de Investigación, Derechos Fundamentales ante el reto de la Inteligencia Artificial de la Universidad Rey Juan Carlos (IAIURIS), financiado a través de las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos. Una iniciativa de la Profesora Cobo del Rosal Pérez, con un extraordinario impacto actual, y enorme calado, y a la que estoy profundamente agradecida por proponerme mi participación.

Como profesora de Derecho Internacional Público de la Universidad Rey Juan Carlos, que soy, con la especialidad en Derechos Humanos, no ha resultado difícil aceptar el reto propuesto. Sí lo ha sido, el aceptar, después de la investigación realizada, que la inteligencia artificial, más bien, los sistemas de inteligencia artificial que utilizamos de forma ya constante y habitual en nuestro día a día, están replicando y

postergando, desigualdades que creíamos ya superadas en nuestras sociedades occidentales. El acoso, la violencia verbal, el sexismo, el racismo, la misoginia, todo ello, se reproducen en nuestras sociedades a través de la inteligencia artificial. Y porque insaurado entre los seres humanos, se transmite y se vuelca en los programas y sistemas de inteligencia artificial. Y, por tanto, la inteligencia artificial “comete” sesgo de género, de sexo, de raza, de ubicación geográfica, de religión, de etnia, y más, lo que hace necesario investigarlo, identificarlo, y rectificar, siendo el objetivo de este trabajo, al menos, investigarlo e identificarlo. Queremos como resultado hacer una advertencia. El uso de los sistemas de inteligencia artificial no está a salvo de riesgos para que los utilicemos sin ningún control y supervisión humana. Las decisiones, respuestas, resoluciones, y conclusiones a las que queremos llegar a partir del uso de herramientas de inteligencia artificial deben tener una supervisión humana, siendo el humano el que definitivamente, decida, responda, resuelva y concluya.

Y para la elaboración del presente trabajo hemos utilizado bibliografía extensa y normativa de Naciones Unidas, basándonos en una metodología cualitativa de descripción.

Y como punto clave de inicio, el sesgo.

1. EL SESEO

A menudo cuando oímos la palabra *juez*, o *fiscal*, creamos en nuestra mente de forma rápida una imagen, de una persona, un individuo, en ocasiones un hombre, de raza blanca, occidental, sin discapacidad, o una mujer, de raza blanca, occidental, sin discapacidad, ello dependiendo de nuestro propio contexto social, en este caso, hombre, mujer, de raza blanca, con ubicación geográfica occidental, y sin discapacidad. Lo cierto es que el dato oficial de jueces nos lo aporta el Consejo General del Poder Judicial en su sitio web⁹⁹, según el cual, con datos a 2024, los Tribunales colegiados españoles (más de un juez), tienen un 57,2% de jueces mujeres. Y, según el sitio web de la Fiscalía¹⁰⁰, con datos a 2025, de 2.808 fiscales en activo, 1.867 son fiscales mujeres.

En efecto, ¿qué hacemos los humanos cuando nos transmiten un mensaje referente a una persona o personas que no procesamos previamente?, pues creamos una imagen. Creamos un pensamiento, un símbolo. Así, y según el profesor Arsuaga, antropólogo español enormemente conocido por todos, en su libro, *Vida, la gran*

⁹⁹ España. Consejo General del Poder Judicial (CGPJ). Disponible en: <https://www.poderjudicial.es/cgpi/es/Poder-Judicial/Consejo-General-del-Poder-Judicial/Oficina-de-Comunicacion/Notas-de-prensa/El-CGPJ-ha-nombrado-a-una-mujer-en-el-53-por-ciento-de-las-plazas-a-las-que-concurren-candidatos-de-ambos-sexos> [Consultado el 3 de septiembre de 2025].

¹⁰⁰ España. Fiscalía General. Disponible en: <https://www.fiscal.es/-/las-mujeres-ocupan-el-46-de-los-cargos-directivos-del-ministerio-fiscal> [Consultado el 3 de septiembre de 2025].

*historia*¹⁰¹, y sobre la evolución humana, los animales tienen mente y tienen experiencias como sensaciones, emociones y pensamientos, pero, sin embargo, no podemos hablar con ellos, ni ellos nos hablan. Y es que, la diferencia entre la mente humana y la mente de los animales es que la nuestra es simbólica, damos nombre a las cosas. Somos una especie simbólica y lingüística. Siendo que nuestra comunicación está basada en símbolos y pensamientos racionales unidos en nuestra mente. Y ¿podemos razonar sin palabras?, pues el neurocientífico Mario Sigma nos da la respuesta en su libro *El poder de las palabras*¹⁰², según el cual, los humanos al hablar ordenamos nuestros pensamientos, por tanto, el ser humano necesita hablar con los demás, de ahí la necesidad de socializar, siendo necesario claro está, una socialización en comunicad, con armonía y sin violencia.

Pues bien, según lo expuesto, las imágenes que creamos en nuestra mente suelen estar sesgadas, y ello por no tener en cuenta distintas variables, como el sexo o el género, que sin embargo afectan, traducándose en nuestro lenguaje, en nuestras palabras, creando estereotipos, esto es, visiones o ideas preconcebidas sobre roles, características, atributos, o papeles, en particular en los humanos que poseemos o deberíamos poseer, hombres y mujeres, por pertenecer a uno u otro grupo social, perpetuando desigualdades, y hasta la discriminación, pudiendo además interseccionarse con otras variables como la raza, la etnia, la religión, o la nacionalidad.

Y como sesgos más comunes, el “sesgo implícito”, esto es, la tendencia inconsciente de asociar ciertas características con determinados grupos humanos¹⁰³, como hemos indicado antes; o el “sesgo de disponibilidad”, esto es, un atajo mental que hace nuestra mente al dar una respuesta, y así, en vez de buscar la información mental contrastada con datos, damos una respuesta según recuerdos más recientes que tenemos guardados en nuestra mente. Y esto lo hacemos con mucha frecuencia. Pero entre los sesgos el más lineal es el sesgo de género, que veremos después.

1.1.- El sesgo en IA

Así como es obvio, según lo expuesto, que, al interactuar los seres humanos, en nuestras relaciones sociales, está presente el sesgo social y cultural (raza, sexo, edad, género nacionalidad), es obvio, porque así lo determinan los estudios, también este sesgo se incorpora en los sistemas de inteligencia artificial (desde ahora IA), perdiendo y manteniendo desigualdades. Y ello se produce al diseñar los sistemas y las

¹⁰¹ ARSUAGA, J.L. *Vida, la gran historia*. 5ª Edición. Editorial Planeta. Madrid. 2019. ISBN: 978-84-233-5574-7

¹⁰² SIGMA, M. *El poder de las palabras*. Penguin Random House Grupo Editorial. Barcelona. 2022. ISBN: 978-84-18006-49-4.

¹⁰³ Schiebinger, L., Klinge, I., Paik, HY, Sánchez de Madariaga, I., Schraudner, M. y Stefanick, M. (Eds.) (2011-2020). *Innovaciones de género en ciencia, salud y medicina, ingeniería y medio ambiente* (genderedinnovations.stanford.edu).

decisiones sobre el origen y alcance del conjunto de datos con los que se entrenan estos sistemas de IA¹⁰⁴. No obstante, siendo la IA una herramienta tecnológica, debe estar sometida a una regulación estatal. Y a nivel internacional lo está. Así, el diseño, la elaboración y el despliegue de la IA debe estar conforme a obligaciones de los Estados y responsabilidades de los actores privados en consonancia con el Derecho Internacional de protección de los derechos humanos¹⁰⁵. Esto es, el ámbito jurídico de protección de los derechos humanos impone a los Estados obligaciones positivas de promover regulación de protección, pero también, a la vez, obligaciones negativas de abstenerse de aplicar medidas que afecten a la protección. Y respecto de las responsabilidades de los actores privados, y en aplicación de los derechos humanos, los Estados también está obligados a garantizar el respeto por los derechos individuales y a no ser violados por infracciones cometidas por particulares, a través de la regulación. En particular y respecto de la IA, los Estados están obligados a establecer legislación de protección de derechos individuales frente a la implementación de aplicaciones de IA. Sin desquitar la propia responsabilidad de las empresas tecnológicas, en aplicación de los derechos humanos, de crear, adoptar y movilizar estas tecnologías en aplicación de los “Principios Rectores sobre las Empresas y los Derechos Humanos: Puesta en Práctica del Marco de las Naciones Unidas para “Proteger, Respetar y Remediar”¹⁰⁶, que establecen normas de conducta mundial aplicables a todas las empresas que deben regir todas las operaciones de las empresas, cualquiera que sea el lugar donde operen, incluidas las redes sociales y las empresas de búsqueda¹⁰⁷.

No obstante, el incumplimiento de la protección de los derechos humanos por parte de los Estados conlleva discriminación. Y a *sensu contrario*, los Estados tienen la obligación de prohibir toda discriminación. Esto es, en la ley y ante la ley. Una discriminación por razón de raza, sexo, idioma, religión, nacionalidad, ubicación geográfica, situación económica, o cualquier otra de índole social.

En efecto, en el último Informe¹⁰⁸ de la Relatora Especial sobre las formas contemporáneas de racismo, discriminación racial, xenofobia, y formas conexas de intolerancia, se examinan cuatro formas transversales en las que la inteligencia artificial puede contribuir a las manifestaciones de discriminación racial; esto es (i) los problemas relacionados con los datos, (ii) las cuestiones de diseño de algoritmos, (iii) el uso intencionadamente discriminatorio de la inteligencia artificial y (iv) las cuestiones relativas a la rendición de cuentas. Y porque según la Relatora Especial las tecnologías digitales y la discriminación racial está conexas.

En efecto, así como puede haber una percepción pública de que la tecnología es objetiva y neutra, incluso entre los creadores tecnológicos, esto no es así. La tecnología

¹⁰⁴ Naciones Unidas. Asamblea General. *Resolución A/73/348*, de 29 de agosto de 2018.

¹⁰⁵ *Ut supra*, pag 10.

¹⁰⁶ Naciones Unidas. Asamblea General. *Resolución A/HRC/17/31*, de 21 de marzo de 2011.

¹⁰⁷ *Ut supra*.

¹⁰⁸ Naciones Unidas. Consejo de Derechos Humanos. *Resolución A/HRC/56/68*, de 3 de junio de 2024.

no es neutra ni objetiva en ningún caso, pues es el fiel reflejo de los valores e intereses de sus creadores, esto es, de los que diseñan la tecnología, y, por tanto, está fundamentalmente conformada por las mismas estructuras de desigualdad que operan en la sociedad¹⁰⁹.

En verdad, la IA tiene varios tipos. Está la inteligencia artificial predictiva, que es una IA considerada como "tradicional", basada en modelos que utilizan datos históricos, patrones y tendencias para hacer predicciones fundamentadas respecto a resultados o acontecimientos futuros. También está la inteligencia artificial igualmente categorizada como "tradicional" que se utiliza para identificar caracteres impresos, rostros humanos, objetos y demás información, y engloba diversas tecnologías para reconocer y distinguir objetos, individuos y patrones en los datos que se le facilitan. Y finalmente están los sistemas de inteligencia artificial generativa que son formas más recientes de inteligencia artificial, versátiles que pueden utilizarse para diversos fines. Esto es, abarcan una clase de sistemas de inteligencia artificial diseñados para producir diversos resultados sobre la base de amplios conjuntos de datos de entrenamiento, redes neuronales, arquitectura de aprendizaje profundo e indicaciones del usuario. Y pueden producir una amplia gama de resultados, como imágenes, texto, audio, vídeo y datos sintéticos. Por tanto, a diferencia de los modelos de inteligencia artificial "tradicionales" que se centran en detectar patrones en los datos existentes, esta inteligencia artificial generativa se entrena para crear nuevos puntos de datos que imiten los patrones y las características de los datos utilizados para entrenar los modelos de aprendizaje automático¹¹⁰.

"Un ejemplo claro de cómo los prejuicios raciales se reproducen en los avances tecnológicos es la actuación policial en base a modelos predictivos. Las herramientas de actuación policial en base a modelos predictivos llevan a cabo evaluaciones sobre quién cometerá futuros delitos, así como dónde puede que ocurra un delito en el futuro, basándose en la ubicación y en datos personales.

*Las actuaciones policiales basadas en modelos predictivos pueden agravar la exagerada vigilancia policial a la que se ha sometido históricamente a ciertas comunidades atendiendo a sus características raciales y étnicas. Debido a que los agentes encargados del mantenimiento del orden han centrado históricamente su atención en estos vecindarios, los miembros pertenecientes a estos vecindarios tienen una representación excesiva en los registros policiales. Este hecho repercute a su vez sobre dónde predicen los algoritmos que ocurrirán futuros delitos, lo que conduce a un aumento del despliegue policial en estas zonas en cuestión."*¹¹¹.

¹⁰⁹ Naciones Unidas. Consejo de Derechos Humanos. *Resolución A/HRC/44/57*, de 18 de junio de 2020.

¹¹⁰ *Ut supra*, *Resolución A/HRC/56/68*, pag 3.

¹¹¹ Naciones Unidas. Ashwini K.P., Relatora Especial sobre las formas contemporáneas de racismo, discriminación racial, xenofobia e intolerancia relacionada. *Resolución A/HRC/56/68*.

1.2.- El sesgo de género en la IA

El término género apareció a finales de la década de 1960, como término distintivo de lo biológico del sexo, siendo un término que interpreta las diferencias de comportamiento como resultado de la disposición biológica. El género así entendido, se utilizó para distinguir los factores socioculturales que moldean los comportamientos y actitudes con respecto a los factores biológicos relacionados con el sexo. Y, por tanto, entendiendo que los comportamientos y actitudes de género se *aprenden*; no son fijos ni universales. Y estas experiencias de género pueden afectar a la biología. Además, las personas pueden querer cambiar aspectos de su cuerpo para alinearlos mejor con su identidad de género. Por esta razón, “sexo y género” suelen ser términos analíticos de gran utilidad en todas las ciencias, incluso cuando interactúan¹¹², aunque sexo sea una variable que carece de precisión¹¹³ distinta del género.

En verdad, el género, como variable que afecta a la persona, es multidimensional. Y se interrelaciona con otras variables como el sexo, según se ha indicado, pero también con la sexualidad, la edad, la etnia, el nivel económico, y más como la raza. Y

¹¹² *Ut supra*. Schiebinger, L., Klinge, I., Paik, H.Y., Sánchez de Madariaga, I., Schraudner, M. y Stefanick, M. (Eds.) (2011-2020). *Innovaciones de género en ciencia, salud y medicina, ingeniería y medio ambiente* (gendered-innovations.stanford.edu). El proyecto Gendered Innovations se inició en la Universidad de Stanford en julio de 2009. Entre 2011 y 2013, la Comisión Europea financió un Grupo de Expertos, «Innovación a través del Género/Innovaciones con Perspectiva de Género», en el marco de su Séptimo Programa Marco, con el objetivo de desarrollar la dimensión de género en la investigación e innovación de la UE. La Fundación Nacional de Ciencias de EE. UU. (NSF) se unió al proyecto en enero de 2012. Entre 2018 y 2020, el Grupo de Expertos de Horizonte 2020, Gendered Innovations (G12), actualizó y amplió la metodología y los estudios de caso del proyecto. Entre 2019 y 2022, la NSF cofinanció talleres para elaborar nuevos estudios de caso sobre perspectiva de género. A finales de 2018, Gendered Innovations recibió financiación de la Comisión Europea para desarrollar lo que denominamos Gendered Innovations 2. Este grupo estaba formado por 29 expertos en diversos campos, desde biología de sistemas y desarrollo de fármacos hasta ingeniería civil, entre otros. También se incorporó a una nueva generación de codirectores: Elena Gissi, científica sénior del Instituto de Ciencias Marinas de Venecia; Mathias W. Nielsen, catedrático de Sociología de la Universidad de Copenhague; y Sabine Oertelt-Prigione, catedrática de Medicina con Perspectiva de Género de la Universidad Radboud (Países Bajos) y la Universidad de Bielefeld (Alemania). Cada uno de ellos fue fundamental para aportar nuevas habilidades y materiales a Gendered Innovations. Anne Pépin, responsable de políticas del sector de género de la Dirección General de Investigación e Innovación de la CE, convocó al grupo de expertos, que se reunió en tres talleres presenciales en Bruselas y elaboró quince nuevos estudios de caso, que abarcaban desde el análisis del sexo en las ciencias marinas —un campo totalmente novedoso— hasta el análisis del género en soluciones energéticas inteligentes y espacios urbanos de calidad. Metodológicamente, Gendered Innovations 2 actualizó numerosos estudios de caso para integrar la interseccionalidad.

¹¹³ Velocci, F. C. *Historia de la investigación sobre el sexo: ¿Es el “sexo” una categoría útil?*. En línea. Cell Press Journal. Volumen 187, Número 6, 1343 – 1346. Disponible en: [https://www.cell.com/cell/full-text/S0092-8674\(24\)00122-3](https://www.cell.com/cell/full-text/S0092-8674(24)00122-3) [Consultado el 3 de octubre de 2025].

porque el género se refiere a las normas, las identidades y las relaciones socioculturales que estructuran la sociedad y determinan los comportamientos, los entornos sociales y lo que rodea a la persona. Y cambian con el tiempo y el espacio. El género también es un principio organizador social que estructura comportamientos, actitudes, apariencia física y hábitos de comportamiento. El género marca normas sociales, aunque cambian constantemente según la época histórica, la cultura o la ubicación geográfica. El género marca identidades personales según las personas se perciben a sí mismos y se presentan hacia los demás. Y el género marca las relaciones personales y con las instituciones que nos rodean. El género se aprende. No es fijo ni universal. Pero interactúa con el sexo¹¹⁴.

En cambio, el sexo, hoy podemos afirmar que no es una variable de investigación estática, es más bien múltiple. Sin embargo, se ha ido determinando en función de los estudios de los que lo han estudiado, como la ciencia, la tecnología y la sociología. Así, durante siglos se ha identificado con el modelo anatómico mediante inspección visual de los genitales externos. A partir del siglo XIX la visión anatómica se vio rectificada por la visión gonadal en función del hermafroditismo o lo que hoy se llama la intersexualidad. No obstante, en el siglo XX se introduce la idea de lo bisexual y nuevas teorías se posicionan entorno al término sexo como las endocrinológicas en lo referente a la diferenciación sexual, que surgen derivadas de las teorías del modelo gonadal y la ciencia embriológica, aunque también se unieron teorías metabólicas y hormonales. Pero también llegaron nuevos métodos de investigación científica como los citólogos y los genetistas que aportaron un modelo cromosómico para el sexo (el cromosoma Y el X), con lo que se reformuló de nuevo la teoría binaria como verdadera. En todo caso, para la ciencia como para la gobernación de las sociedades, de forma simplificada, el sexo es una variable dicotómica universalmente aceptada, aun cuando es necesario añadir que no por ello se contribuye a resultados precisos, más bien erróneos que hacen perdurar supuestos cisonormativos y heteronormativos. Por lo tanto, el término sexo abarca múltiples definiciones, además de las científicas, debiendo en todo caso especificar en cada caso los métodos en la recopilación y registro de datos relacionados con el sexo para mejorar la precisión de los resultados basados en esta variable no binaria, y sin considerar como anomalías aquellos resultados que no se ajustan a la estructura normalizada heteronormativa. Es lo que se llama – pensamiento crítico aplicado al sexo y a la sexualidad, estudiado por científicos de forma interdisciplinar, esto es, científicos de las ciencias de la salud, de la tecnología, pero también de las ciencias sociales, como los humanistas, los antropólogos, y los sociólogos¹¹⁵.

¹¹⁴ *Ut supra*. Schiebinger, L., Klinge, I., Paik, H.Y., Sánchez de Madariaga, I., Schraudner, M. y Stefanick, M. (Eds.) (2011-2020). *Innovaciones de género en ciencia, salud y medicina, ingeniería y medio ambiente* (genderedinnovations.stanford.edu). Disponible en: <https://genderedinnovations.stanford.edu/terms/gender.html> [Consultado el 22 de octubre de 2025].

¹¹⁵ *Ut supra*. Velocci, F. C. *Historia de la investigación sobre el sexo: ¿Es el “sexo” una categoría útil?*.

Y es que, cuando nos referimos a seres humanos y a mamíferos no humanos, hablamos de hombres, mujeres, machos, hembras, y/o intersexuales, según hablemos de las distintas funciones derivadas de los complementos cromosómicos, los órganos reproductores o factores hormonales, ambientales o culturales específicos que afecta. Pero si estamos hablando de seres no mamíferos, el sexo son los atributos biológicos que determinarán si estamos ante machos, hembras, y/o hermafroditas, según distintas funciones derivadas de los complementos cromosómicos o de las condiciones ambientales o sociales. Pero en la ingeniería, y respecto del ser humano, el sexo también afecta a la altura, el peso, y las características biomédicas, para determinarlo en relación con productos, sistemas o procesos. Por ello, y respecto de la recopilación de datos, tan importante en esta era digital, las Academias Nacionales de Ciencias, Ingeniería y Medicina en EE. UU. han determinado la necesidad de utilizar el factor o variable del sexo para la recopilación de datos, pero con la intersección de la variable del género, no binario, y multidimensional¹¹⁶ como ya hemos indicado.

2. POSICIÓN DE CASILLA “d8”

En el juego del ajedrez, aunque hay muchas variables a lo largo del mundo, podemos decir de forma estándar, que dispone de dieciséis piezas móviles en dos colores, blancas y negras, para ser utilizadas por dos jugadores, disponiendo cada uno de los jugadores de las dieciséis piezas del mismo color, esto es, un rey, una dama, dos torres, dos alfiles, dos caballos y 8 peones, sorteando al inicio del juego el color para cada jugador, porque las piezas blancas son las que empiezan a jugar y por tanto llevan la iniciativa del juego, con ligera ventaja sobre las negras. Pues bien, es la dama negra la que está situada en la posición de salida de la casilla “d8”, al inicio de la partida. Y sobre ella ponemos el foco de un modo anecdótico, únicamente para evidenciar cómo el blanco y el negro ha condicionado y condiciona socialmente nuestra visión de vida.

En efecto, en 1956, se dio lo que se llamó “la partida del siglo” donde dos personas, *Bobby Fischer* de 13 años, y *Donald Byrne* de 26, disputaron una partida mítica en New York, y donde ganó el joven de 13 años al entonces maestro del ajedrez con una jugada de sacrificio pasivo de la dama negra en el movimiento 17. ¿no es anecdótico? Y ¿por qué traemos esta anécdota a este trabajo?

En verdad, lo que queremos focalizar es precisamente la visibilidad de la ventaja y por tanto de la desventaja en contrapartida, siendo una simple nota, o no, que el color blanco sea el ventajoso y el color negro sea el desventajoso, y en particular la dama negra la sacrificada.

¹¹⁶ *Ut supra*. Schiebinger, L., Klinge, I., Paik, H.Y., Sánchez de Madariaga, I., Schraudner, M. y Stefanick, M. (Eds.) (2011-2020). *Innovaciones de género en ciencia, salud y medicina, ingeniería y medio ambiente* (genderedinnovations.stanford.edu). Disponible en: <https://genderedinnovations.stanford.edu/terms/sex.html> [Consultado el 23 de octubre de 2025].

2.1.- La interseccionalidad

La interseccionalidad es un término acuñado en 1989¹¹⁷ por el jurista *Kimberlé Crenshaw*, por el cual se describe la discriminación intersectorial entre el sexismo y el racismo para identificar una doble discriminación en una misma persona, y así evitar su tratamiento por separado. Esto es, mujer y de raza negra. No obstante, después se ha ampliado a otras formas de discriminación. Por tanto, actualmente la interseccionalidad es multisectorial, esto es, sistemas superpuestos interdependientes de discriminación relacionados con la edad, la discapacidad, el sexo, la sexualidad, la nacionalidad, la etnia, y más. Pero se parte de la variable del género en una persona, que, por su contexto, se entrelaza con otra u otras variables que le afectan, interseccionando en desigualdad, y derivando en una múltiple discriminación. En efecto, personas, que afectándoles esta interseccionalidad, les produce una vivencia de desventaja estructural que afecta del todo en sus vidas, moldeando sus oportunidades vitales sociales¹¹⁸.

En verdad, es necesario evaluar factores sociales interrelacionados en el diseño, en la actuación profesional, o en la ciencia de la investigación de productos, procesos, o servicios, para evitar la interseccionalidad, siendo entre otros factores sociales, la edad, la discapacidad, el género, el sexo, la raza, la localización geográfica, la etnia, o la sostenibilidad. Este último, la sostenibilidad, un factor social que mide la necesidad de diseñar, actuar, o investigar, ambientalmente de forma responsable, para satisfacer el presente, pero sin comprometer a generaciones futuras.

Así, en relación con los programas de IA para el desbloqueo de móviles, en seguridad y videovigilancia masiva, pueden darse circunstancias de discriminación por razón de raza, género o interseccionalidad, esto es una amplificación de la discriminación. Y es que, los sistemas de IA que utilizan datos para el aprendizaje automatizado, y recopilación de datos, preparación de datos y selección de modelos de IA, esto es, datos masivos utilizados, provienen masivamente de hombres blancos, llegando a identificar de forma errónea a mujeres de color.

En relación con los asistentes virtuales o chatbots, mayoritariamente con voces feminizadas, suelen ser acosados. Y es que, la IA responde adecuadamente a consultas, pero surge el problema en el leguaje cuando este tiene carga sexual y abusiva. Son programas de IA para responder con palabras y frases evasivas o juguetonas a preguntas con intención sexual, que lo que producen es en realidad un replicar y reproducir estereotipos de género, como la debilidad femenina en respuesta al abuso o acoso. Claramente fomentan el acoso.

¹¹⁷ *Ut supra*. Disponible en: <https://genderedinnovations.stanford.edu/terms/intersectionality.html> [Consultado el 22 de octubre de 2025].

¹¹⁸ *Ut supra*. Disponible en: <https://genderedinnovations.stanford.edu/terms/intersectionality.html> [Consultado el 3 de octubre de 2025].

Respecto de los maniqués de prueba en pruebas de seguridad automovilística por los ingenieros industriales, hay que decir que varían según población. Así el maniqué hombre tiene un perfil 50, según se entiende es el promedio más común de hombre en la población, esto es, 1,75m de altura y 78kgs de peso. No obstante, en 1966 se introdujo el maniqué mujer con un perfil 5, siendo el perfil más pequeño para representar a la población promedio más común de mujer, esto es, 1,50m de altura y 49kgs de peso. Sin embargo, el maniqué femenino de perfil 5, no contempla el cuerpo femenino, es en realidad la versión pequeña del maniqué masculino de perfil 50 a escala perfil 5. Y porque no hay maniqué de perfil 50 femenino, y, por tanto, no se contempla la tolerancia a lesiones en el cuerpo femenino, la biomecánica, la alineación espinal, la fuerza del cuello, la fuerza muscular y ligamentosa, la respuesta dinámica al trauma y otras características de la mujer. En consecuencia, las mujeres sufren lesiones más graves que los hombres en accidentes de tráfico comparables. Pero también la posición al sentarse a conducir varía de hombres a mujeres. En promedio, la mujer se sienta más cerca del volante para compensar su baja estatura, y por tanto ello provoca mayores lesiones en colisión frontal. Y es que, el problema radica en la posición y forma del asiento del conductor, así como su configuración respecto del resto de mecanismos del coche para guiarlo. En consecuencia, los vehículos no están diseñados para la seguridad de todos los grupos humanos. Sin dejar de mencionar a las mujeres embarazadas y el cinturón de seguridad, que en un número considerado de choques frontales provocan muertes fetales en accidentes de tráfico con traumatismo materno¹¹⁹.

2.2.-Supervisión humana. Buenas prácticas

En verdad, debemos pensar que hay esperanza respecto de todo lo mencionado porque hay intención de aplicar innovación frente a la IA.

En efecto, respecto de la necesidad del agua potable para el ser humano, se sabe que casi 1.000 millones de personas carecen de acceso confiable al agua potable. En particular, en África subsahariana, la recolección del agua se hace por las mujeres y las niñas cuando las infraestructuras híbridas no permiten el acceso directo, por tanto, son las mujeres y niñas las que dedican a realizar este rol de género en sus sociedades y comunidades, dedicando unos 40 mil millones de horas anuales a esta tarea. Sin embargo, recientes proyectos de ingeniería civil incluyen a mujeres en su elaboración. Sus conocimientos en suelos y rendimientos híbridos se utilizan para el mejor desarrollo de estos programas, facilitando la determinación más exacta de ubicación idónea de pozos y tomas de agua. Por tanto, el rol de género que viven estas mujeres, dedicando como tarea social a abastecer de agua a la familia, la comunidad, o el poblado, hacen que tengan un conocimiento mayor del suelo y de la obtención del agua, conociendo las mejores estaciones híbridas, habiendo mejorado la eficacia de estos programas con su participación.

¹¹⁹ *Ut supra.*

En relación con los chatbots y los asistentes virtuales como innovación, se están implementando ajustes tecnológicos por las empresas para que asuman dirigirse a grupos humanos diversos, con características diversas e interrelacionadas, como el género, la raza, la etnia, la edad o la religión. Se está implementando la utilización de idiomas, dialectos, acentos y jergas, ajustadas a la diversidad de grupos humanos. Y ante el acoso, se está implementado el no dar respuestas a preguntas con connotación sexual, así como tolerancia cero. En particular se está introduciendo voz de género neutro denominada “Q”, una voz sin género de entre 145Hz y 175 Hz de frecuencia.

3. EDUCACIÓN EN IA

Según UNESCO, Organización de las Naciones Unidas para la educación, la ciencia y la tecnología, órgano líder en Naciones Unidas en educación, con un programa de educación fundado tras la segunda guerra mundial en 1945, la educación es un derecho de todos a lo largo de la vida¹²⁰. Y porque, la educación es un derecho humano. Sin embargo, este derecho humano se enfrenta a nuevos desafíos mundiales actuales, como el cambio climático, los conflictos y las crisis prolongadas como la pandemia del Covid-19, y la era digital. En particular, la pandemia ha sido la peor perturbación de la educación en la historia, afectando al aprendizaje de 1.600 millones de niños y jóvenes, agravando las desigualdades existentes en materia de aprendizaje, habiendo sido una advertencia y señal de alarma sobre la necesidad de realizar cambios e innovaciones urgentes, y proporcionando un impulso para llevar a cabo la Cumbre para la Transformación de la Educación de las Naciones Unidas en septiembre de 2022¹²¹. Y porque los datos son los siguientes: 771 millones de jóvenes y adultos, de los cuales dos tercios son mujeres, carecen de conocimientos básicos de lectura y escritura; 244 millones de niños y jóvenes están sin escolarizar en todo el mundo; menos del 40% de las niñas en África subsahariana tienen estudios del primer ciclo de educación secundaria; y se necesitan 200.000 millones de dólares anuales de financiación adicional para que el mundo pueda alcanzar el ODS n°4 en 2030¹²².

En efecto, según el Informe de la Comisión Internacional sobre futuros de la educación¹²³,

¹²⁰ Naciones Unidas. UNESCO. Disponible en: <https://www.unesco.org/es/education/action> [Consultado el 3 de septiembre de 2025].

¹²¹ Naciones Unidas. UNESCO. *Transformar la educación para el futuro*. ED.2022/WS/23. Disponible en: <http://www.unesco.org/open-access/terms-use-ccbysa-sp> [Consultado el 3 de septiembre de 2025].

¹²² *Ut supra*.

¹²³ Naciones Unidas. UNESCO. Informe de la Comisión Internacional sobre los futuros de la educación. Disponible en: <https://es.unesco.org/open-access/terms-use-ccbysa-sp> Este informe es el resultado del trabajo colectivo de la Comisión Internacional sobre los Futuros de la Educación, creada por la UNESCO en 2019. En particular, se analiza el futuro de los siguientes temas críticos que deben replantearse: sostenibilidad; conocimiento; aprendizaje; docentes y enseñanza; trabajo, capacidades

La transformación digital de nuestras sociedades está afectando nuestras vidas de una forma sin precedentes. Los ordenadores están cambiando rápidamente la forma de crear, acceder, difundir, validar y utilizar el conocimiento. Esto hace que la información sea más accesible y abre nuevas y prometedoras vías para la educación. Pero los riesgos son muchos: el aprendizaje puede tanto reducirse como ampliarse en los espacios digitales; la tecnología proporciona nuevas palancas de poder y control que pueden tanto reprimir como emancipar; y, con el reconocimiento facial y la IA, nuestro derecho a la privacidad puede reducirse de formas inconcebibles hace apenas diez años. Tenemos que estar atentos para asegurarnos de que las transformaciones técnicas en curso nos ayudan a prosperar y no amenazan las perspectivas futuras de conocimiento o nuestra libertad intelectual y creativa.

Y en particular y respecto de la IA, el mandato de la UNESCO exige intrínsecamente un enfoque de la inteligencia artificial basado en el ser humano. Su objetivo es incluir el papel desempeñado por la IA en la solución de las desigualdades actuales en materia de acceso al conocimiento, la investigación y la diversidad de las expresiones culturales, y garantizar que la IA no amplíe la brecha tecnológica dentro de los países y entre ellos. La promesa de la “IA para todos” debe permitir que cada cual pueda sacar provecho de la revolución tecnológica en curso y acceder a sus frutos, fundamentalmente en términos de innovaciones y conocimientos. La inteligencia artificial tiene el potencial de abordar algunos de los mayores desafíos de la educación actual, innovar las prácticas de enseñanza y aprendizaje, y acelerar el progreso hacia el ODS 4. Sin embargo, el rápido desarrollo tecnológico inevitablemente conlleva múltiples riesgos y desafíos, que hasta ahora han superado los debates políticos y los marcos regulatorios. La UNESCO se compromete a apoyar a los Estados Miembros para que aprovechen el potencial de las tecnologías de IA para alcanzar la Agenda de Educación 2030, garantizando al mismo tiempo que su aplicación en contextos educativos se rija por los principios fundamentales de inclusión y equidad¹²⁴.

La velocidad a la que las tecnologías de IA generativa se están integrando en los sistemas educativos en ausencia de controles, normas o reglamentos, es asombrosa. Sorprende que hoy en día, en la mayoría de los contextos nacionales, el tiempo, los pasos y los permisos necesarios para validar un nuevo libro de texto superen con creces los requeridos para trasladar los servicios de la IA generativa a las escuelas y las aulas. De hecho, los servicios que proporciona la IA a menudo no requieren validación alguna. Han sido arrojados a la esfera pública sin debate previo ni revisión. Nos planteamos qué pocas tecnologías que se ponen a disposición de niños y jóvenes de todo el mundo pocas semanas después de su desarrollo. En muchos casos, los gobiernos y las escuelas están adoptando una tecnología radicalmente desconocida que ni siquiera los principales especialistas en tecnología afirman comprender. Hay muy pocos precedentes de esta evolución. Internet y los teléfonos móviles no fueron acogidos de inmediato

y competencias; ciudadanía; democracia e inclusión social; educación pública; y educación superior, investigación e innovación. [Consultado el 3 de octubre de 2025].

¹²⁴ *Ut supra*.

en las escuelas ni para su utilización con niños a los pocos meses de su invención. El uso de la IA puede controlarse o limitarse, como ocurre con otras tecnologías, aunque se haya popularizado la idea de que esto no es factible. En muchos países existen normas estrictas que controlan y restringen el uso de tecnologías que se sabe constituyen una amenaza o que todavía son demasiado nuevas para justificar una difusión amplia o desmedida. Aunque estas normas no siempre son perfectas, son bastante eficaces. De cara al futuro, se necesita un equilibrio mucho mejor entre los especialistas en IA que desarrollan tecnología y aplicaciones para su uso y, por otro lado, los expertos que trabajan para los gobiernos con miras a revisar la seguridad de estas aplicaciones y considerar cuidadosamente sus potenciales usos indebidos y cómo minimizar estos potenciales. En la actualidad, hay muy pocos especialistas en el lado correspondiente a la seguridad de esta ecuación y aún menos que operen con verdadera independencia y al margen de las empresas que desarrollan la IA con fines comerciales. El reciente informe sobre el Índice de Inteligencia Artificial para 2023 mostraba que menos del uno por ciento de los doctorados en Inteligencia Artificial se incorporan a la administración pública tras su graduación. Esta tendencia no ha cambiado en los últimos cinco años. La mayoría de estos graduados se incorporan a la industria, mientras que aproximadamente una cuarta parte lo hace al ámbito académico. Sencillamente, no hay suficientes especialistas en la parte reguladora de la ecuación. En nuestro contexto actual de incertidumbre, novedad y escasos controles de seguridad, un enfoque más prudente de la IA generativa en la educación es una medida de sentido común¹²⁵.

4. CONCLUSIONES

La inteligencia artificial, mal llamada así, está en nuestras vidas, en las calles, en los locales, en nuestras casas, en las escuelas, en las universidades, en los centros de trabajo. Y la utilizamos casi a diario. El problema es que sabemos poco de sus riesgos. No nos lo han enseñado. No hemos adquirido educación y formación en el uso de esta nueva herramienta tecnológica. Y la norma que la regula, apenas la conocemos. No conocemos nuestros derechos afectados por su uso. Y tiene riesgos de vulnerarlos, y de forma muy grave, grave, o menos grave o leve. Pero lo desconocemos.

Por tanto, la ciudadanía en la sociedad está desprotegida por una falta de conciencia del riesgo del uso de la IA, siendo un problema mayor la población de menor edad. Esto es, jóvenes y adolescentes, cada día utilizan la IA sin saber de sus repercusiones.

Y es que, la IA comete sesgos en sus respuestas, replicando y reproduciendo estereotipos sexistas, racistas, incluso xenófobos, que creíamos superados. Y los jóvenes al interactuar con la IA los interiorizan como forma común de comportamiento. La IA se ha incorporado a nuestra vida cotidiana de una forma muy rápida y sutil, sin control, y con gran interés empresarial tecnológico. Y esto nos debe hacer reflexionar.

¹²⁵ Naciones Unidas. UNESCO. <https://doi.org/10.54675/ACWQ6815> Disponible en: <https://creativecommons.org/licenses/by-sa/3.0/igo> [Consultado el 3 de octubre de 2025].

Hemos reivindicado históricamente el derecho a la igualdad, a la no discriminación. Y nos ha costado mucho sufrimiento como sociedad llegar a esa plena igualdad de sexo, de género, de raza, de nacionalidad, religión, etnia. Pero la llegada de la IA está trastocando estos principios asumidos e instaurados con tanta sangre derramada. Es necesario tomar conciencia que la IA debe estar sometida a controles humanos en su respuesta. No es que debamos impedir que la IA nos suplante. Lo verdaderamente necesario es que seamos los humanos los que la controlemos, sabiendo que sólo es una herramienta tecnológica, a diferencia de nosotros, individuos, seres humanos, con emociones y sentimientos, lo que nos diferencia como especie en la tierra, nuestro planeta.

BIBLIOGRAFÍA

España. Consejo General del Poder Judicial (CGPJ). Disponible en:

<https://www.poderjudicial.es/cgpj/es/Poder-Judicial/Consejo-General-del-Poder-Judicial/Oficina-de-Comunicacion/Notas-de-prensa/El-CGPJ-ha-nombrado-a-una-mujer-en-el-53-por-ciento-de-las-plazas-a-las-que-concurren-candidatos-de-ambos-sexos> [Consultado el 3 de septiembre de 2025].

España. Fiscalía General. Disponible en: <https://www.fiscal.es/-/las-mujeres-ocupan-el-46-de-los-cargos-directivos-del-ministerio-fiscal> [Consultado el 3 de septiembre de 2025].

ARSUAGA, J.L. *Vida, la gran historia*. 5ª Edición. Editorial Planeta. Madrid. 2019. ISBN: 978-84-233-5574-7

SIGMA, M. *El poder de las palabras*. Penguin Random House Grupo Editorial. Barcelona. 2022. ISBN: 978-84-18006-49-4.

Schiebinger, L., Klinge, I., Paik, HY, Sánchez de Madariaga, I., Schraudner, M. y Stefanick, M. (Eds.) (2011-2020). *Innovaciones de género en ciencia, salud y medicina, ingeniería y medio ambiente*. (genderedinnovations.stanford.edu).

Naciones Unidas. Asamblea General. *Resolución A/73/348*, de 29 de agosto de 2018.

Naciones Unidas. Asamblea General. *Resolución A/HRC/17/31*, de 21 de marzo de 2011.

Naciones Unidas. Consejo de Derechos Humanos. *Resolución A/HRC/56/68*, de 3 de junio de 2024.

Naciones Unidas. Consejo de Derechos Humanos. *Resolución A/HRC/44/57*, de 18 de junio de 2020.

Naciones Unidas. Ashwini K.P., Relatora Especial sobre las formas contemporáneas de racismo, discriminación racial, xenofobia e intolerancia relacionada. *Resolución A/HRC/56/68*.

Velocci, F. C. *Historia de la investigación sobre el sexo: ¿Es el “sexo” una categoría útil?* En línea. Cell Press Journal. Volumen 187, Número 6, 1343 – 1346. Disponible en: [https://www.cell.com/cell/fulltext/S0092-8674\(24\)00122-3](https://www.cell.com/cell/fulltext/S0092-8674(24)00122-3) [Consultado el 3 de octubre de 2025].

Naciones Unidas. UNESCO. Disponible en: <https://www.unesco.org/es/education/action> [Consultado el 3 de septiembre de 2025].

Naciones Unidas. UNESCO. *Transformar la educación para el futuro*. ED.2022/WS/23. Disponible en: <http://www.unesco.org/open-access/terms-use-ccbysa-sp> [Consultado el 3 de septiembre de 2025].

Naciones Unidas. UNESCO. Informe de la Comisión Internacional sobre los futuros de la educación. Disponible en: <https://es.unesco.org/open-access/terms-use-ccbysa-sp> [Consultado el 3 de octubre de 2025].

Naciones Unidas. UNESCO. <https://doi.org/10.54675/ACWQ6815> Disponible en: <https://creativecommons.org/licenses/by-sa/3.0/igo> [Consultado el 3 de octubre de 2025].



CAPÍTULO 4.

SESGOS ALGORÍTMICOS Y DERECHOS HUMANOS: DESAFÍOS ÉTICOS Y REGULATORIOS EN LA ERA DE LA INTELIGENCIA ARTIFICIAL¹²⁶

ANGÉLICA MARÍA CASTRO ACOSTA

Profesora Doctora de Derecho Constitucional y de Nuevas Tecnologías de
la Información y las Comunicaciones e IA y
Directora del Máster de Propiedad Intelectual y Derecho Tecnológico de la UEM¹²⁷
Departamento de Ciencias Jurídicas y Humanidades
Facultad de Ciencias Jurídicas, Educación y Humanidades
Universidad Europea de Madrid
angelica.castro@universidadeuropea.es

Resumen

La expansión de la Inteligencia Artificial (IA) en todos los ámbitos de la vida de los seres humanos, algunos críticos como: justicia, salud y empleo, ha generado

¹²⁶ El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

¹²⁷ Doctora en Ciencias Jurídicas por la Universidad de Granada, tras 25 años de ejercicio profesional, su devenir académico investigativo refleja su constante movilidad internacional, con un variopinto elenco de titulaciones, entre ellas, se pueden mencionar: Máster en Reestructuraciones empresariales por el ICAM, Máster en Criptoactivos y Blockchain, Máster en Derecho Penal Internacional por la Universidad de Granada, Máster en Gestión de Calidad ISO 9001 y Auditoría Interna por European Quality University, DEA en Derecho de Familia y Menores por la Universidad de Zaragoza, Diplomada Internacional en Gerencia de Proyectos, Postgraduada en Derecho Procesal Civil y Comercial; Conciliadora y Mediadora. Cursó y se graduó como Licenciada en Derecho en tres países: España, Colombia y Perú, en donde también es Abogada colegiada. Ejerce profesionalmente de forma internacional, asesorando y llevando casos nacionales y con componente extranjero.

En su faceta docente es Docente de Derecho Constitucional, Derechos Fundamentales, Derecho de Nuevas Tecnologías de la Información y las Comunicaciones, Derecho de Extranjería, Derecho Internacional Privado, Ética y Eficacia Profesional, Teoría General del Derecho, Derecho Civil y Regulación de la IA y su componente Ético en la Universidad Europea de Madrid, donde también es Coordinadora de las Prácticas profesionales de los Grados de Derecho, Derecho - ADE y Derecho Relaciones Internacionales y Directora del Máster de Propiedad Intelectual y Derecho Tecnológico.

oportunidades significativas de mejoras en muchos aspectos de la vida humana, pero también genera un impacto directo en los Derechos Humanos mediante la creación de riesgos, entre los cuales ubicamos en una posición privilegiada por lo riesgoso a los sesgos algorítmicos, que se han convertido en una preocupación central, pues pueden reproducir y amplificar discriminaciones existentes, afectando principios fundamentales como igualdad y no discriminación.

Este artículo pretende hacer una ubicación conceptual en el mundo especializado de los sesgos algorítmicos y tras analizar el impacto de los sesgos algorítmicos en la protección de derechos humanos, al revisar casos documentados y marcos regulatorios internacionales, como el EU AI Act y las directrices de UNESCO, mediante una revisión sistemática de literatura y análisis comparativo, se concluirá con una propuesta de protocolo que integra mecanismos para garantizar transparencia, auditoría y supervisión ética en el desarrollo de IA. Los resultados sugieren que la falta de explicabilidad y control en sistemas algorítmicos incrementa la vulnerabilidad de grupos marginados, lo que exige una gobernanza global orientada a derechos humanos.

Palabras clave

Inteligencia Artificial, sesgo algorítmico, derechos humanos, regulación, ética tecnológica.

Abstract

The expansion of Artificial Intelligence (AI) in all areas of human life, including critical ones such as justice, health, and employment, has generated significant opportunities for improvement in many aspects of human life, but it also has a direct impact on Human Rights by creating risks. Among these risks, algorithmic biases stand out as particularly dangerous and have become a central concern, as they can reproduce and amplify existing discrimination, affecting fundamental principles such as equality and non-discrimination.

This article aims to establish a conceptual framework within the specialized field of algorithmic bias. After analyzing the impact of algorithmic bias on human rights protection by reviewing documented cases and international regulatory frameworks, such as the EU AI Act and UNESCO guidelines, through a systematic literature review and comparative analysis, it will conclude with a proposed protocol that integrates mechanisms to ensure transparency, auditing, and ethical oversight in AI development. The results suggest that the lack of explainability and control in algorithmic systems increases the vulnerability of marginalized groups, thus requiring global governance oriented towards human rights.

Key words

Artificial Intelligence, algorithmic bias, human rights, regulation, technological ethics.

Sumario: INTRODUCCIÓN 1. MARCO TEÓRICO 1.1. Conceptualización de los Sistemas de IA 1.2. Entendiendo los Algoritmos 2. TAXONOMÍA DE LOS SESGOS ALGORÍTMICOS 2.1. Tipos de Sesgos 2.1.1. Sesgo de Representación (Representation Bias) 2.1.2. Sesgo de Medida (Measurement Bias) 2.1.3. Sesgo de etiquetado (Label Bias) 2.1.4. Sesgo por objetivo (Objective Bias) 2.1.5. Sesgo Emergente en Sistemas Complejos (Emergent Bias) (Comportamiento no previsto) 3. DERECHOS HUMANOS EN LA ERA DIGITAL 3.1. Marco normativo europeo y español 3.2. Impacto sobre derechos humanos y principios afectados 3.2.1. Derecho a la igualdad y no discriminación 3.2.3. Derecho a la privacidad y protección de datos personales 3.2.4. Debido proceso y derecho a una decisión motivada 3.3. Previniendo los Sesgos Algorítmicos 4. CONCLUSIONES APÉNDICE BIBLIOGRÁFICO

INTRODUCCIÓN

La Inteligencia Artificial (IA) en la era contemporánea post-industrial ha llegado a consolidarse con extrema rapidez como una herramienta transformadora en casi todos los sectores de actividad humana, desde la administración de despachos profesionales, realizar una reserva en un restaurante, guiar un viaje por carretera o multimodal¹²⁸, hasta la generación de documentos con apariencia de discurso coherente sobre una multiplicidad de temas, incursionando también en aquellas actividades de las cuales depende la supervivencia como especie de la humanidad, como: la atención sanitaria, sensores atmosféricos, control de naves de producción animal o de campos de cultivo intensivo, detección rápida de enfermedades al procesar muestras con mayor velocidad, entre muchos otros usos actuales. Sin embargo, su creciente influencia plantea interrogantes sobre la protección del propio ser humano, base y centro de todo Derecho, en entornos digitales, y como máximo exponente de esto, se analiza la protección de los Derechos Humanos como limitación comportamental de los modelos computacionales basados en Inteligencia Artificial, con el fin de precaver situaciones de vulneración de los mismos, habida cuenta el largo camino humano desarrollado, para alcanzar su reconocimiento mundial.

Uno de los desafíos más críticos que se presenta en la interacción de los Modelos computacionales de Inteligencia Artificial con los Derechos Humanos, se puede concentrar en el reconocimiento de la existencia *sine quanon* de un sesgo algorítmico en todos los modelos debido a programación o estructuración, entendido sesgo como la tendencia de los sistemas automatizados a reproducir o amplificar prejuicios presentes en los datos de entrenamiento o en el diseño del modelo. Esto *per se* no puede ser catalogado de malo o bueno, al uso, sino que es una realidad con la que hay que contar para poder, entonces sí, analizarla y determinar la regulación que se le da a las consecuencias de la aplicación de un modelo con un sesgo específico.

¹²⁸ Utilizando diferentes medios de transporte optimizando tiempos y trayectos.

Para ilustrar mejor este aspecto invitamos a analizar el escenario simulado siguiente: Dentro de los datos de programación de un modelo computación con Inteligencia Artificial se han incluido las variables género, edad, raza, nacionalidad, grupo étnico e idioma, este modelo se encargará de mejorar los servicios aeroportuarios, por supuesto al tener estas variables el sistema detecta a una mujer con velo y la asocia al grupo étnico específico, luego deduce que no habla español y plantea una serie de idiomas de comunicación, para garantizarle a la mujer una comunicación adecuada con los agentes fronterizos, cuando la mujer se identifica incorpora la variable nacionalidad, por ello designa a un agente que domine el idioma nacional de tal país para garantizar que en su atención no se generen malos entendidos debidos a sesgos del lenguaje, hasta allí los resultados que ha arrojado la IA han sido para ayudar a brindar mayor agilidad en la atención a personas con idiomas diferentes, garantizándole sus derechos, pero si el output que se le pide a la IA es que identifique personas que pueden estar vulnerando las normas de migración o que sean un peligro para la seguridad aeroportuaria, y entonces dentro de las variables se le incluyen prejuicios sociales como: mujeres con velo, de “x” nacionalidad y que viajen solas, el output que genere la IA, así programada, si estaría vulnerando los Derechos Humanos fundamentales de las mujeres que en ejercicio de su libertad deciden usar velo en su vestimenta diaria.

Todo el ejemplo, conduce a establecer el siguiente punto, con la pretensión de funcionar como un disruptor académico del tema que se analizará en esta investigación: las IA tienen todas las posibilidades de ser el mayor avance de la humanidad, si colocamos al ser humano en su centro; por ello, sólo con una visión humanocéntrica que limite todas las variables utilizadas para el procesamiento de los inputs que se le proporcionan a la IA, se podrá garantizar que esta tecnología no vulnere ningún Derecho Humano en su actuación, lo que garantizaría que la misma sea el vehículo o la herramienta que propulse de forma segura a la Humanidad a su mejor versión.

Permitida la IA como modelos computacionales libres, donde los programadores tienen una libertad absoluta en sus códigos, es cuando nos encontramos con, si se me permite el coloquialismo, “la otra cara de la moneda”, una, más oscura y peligrosa, como es la IA afectando los Derechos Humanos, afectando la equidad en la toma de decisiones, vulnerando Derechos fundamentales de cualquier índole, y generando un grave riesgo para seres humanos específicos y para la sociedad global como tal.

Se debe entender que el avance por el avance no es lo ideal, ya que el precio a pagar, nunca debe ser perder los mínimos que la humanidad ha conseguido tras muchos años de luchas y esfuerzos multilaterales, cuyo principal exponente se puede observar en las declaraciones mundiales de los Derechos Humanos, o en la creación del Derecho Internacional Humanitario, entre otros muchos logros.

Casos recientes evidencian cómo algoritmos utilizados en procesos judiciales, sistemas de reconocimiento facial y plataformas de reclutamiento han generado

resultados discriminatorios, afectando especialmente a mujeres, minorías étnicas y personas con discapacidad.

La adopción exponencial de sistemas de IA en las administraciones públicas, la sanidad, la banca, la selección de personal y el ámbito judicial plantea retos jurídicos y éticos sin precedentes. Si bien la automatización puede aportar eficiencia y coherencia, cuando los modelos son entrenados con datos históricos o reflejan decisiones humanas discriminatorias, no es menos cierto que tienen la virtualidad de producir resultados injustos que vulneren Derechos fundamentales (por ejemplo, la igualdad de trato, el derecho a la intimidad, y el derecho a un procedimiento justo). En la casuística se han encontrado casos emblemáticos, como: la clasificación de género en imágenes que falla con mayor frecuencia para mujeres de piel oscura (Buolamwini & Gebru, 2018), y herramientas de evaluación de riesgo penal que mostraron sesgos raciales (Angwin et al./ProPublica, 2016). Estas constataciones obligan a repensar la responsabilidad humana sobre la programación algorítmica que, al final, es la que genera los resultados indeseados descritos, a lo cual, se le ha empezado a denominar sesgos algorítmicos.

En este contexto, surge la necesidad de analizar el impacto de los sesgos algorítmicos desde una perspectiva jurídica y ética, evaluando la eficacia de los marcos regulatorios existentes y proponiendo mecanismos que garanticen transparencia y rendición de cuentas. Por ello este artículo se plantea responder las preguntas siguientes: ¿cómo afectan los sesgos algorítmicos a los Derechos Humanos? y ¿qué estrategias regulatorias pueden mitigar estos riesgos?

Con tal finalidad en mente, se han planteado los siguientes objetivos: 1. Establecer un marco conceptual. 2. Sintetizar la evidencia empírica sobre sesgos algorítmicos; 2. Mapear el marco jurídico europeo y español aplicable; y 3. Proponer un marco de recomendaciones: técnicas, institucionales y regulatorias, alineado con los estándares internacionales de Derechos Humanos

1. MARCO TEÓRICO

La Inteligencia Artificial (IA) ha llegado a convertirse en muy poco tiempo, en una tecnología global, caracterizada por la transversalidad al impactar múltiples dimensiones humanas: sociales, culturales, políticas, jurídicas, económicas, lúdicas, recreativas, entre otras. Su capacidad para procesar grandes volúmenes de datos y tomar decisiones automatizadas plantea grandes y muy variadas oportunidades, pero también trae consigo el nacimiento o potenciación de riesgos significativos para los Derechos Humanos, especialmente en relación con Derechos tan importantes como la igualdad y la no discriminación, y, es más, es relevante indicar desde ya que, los riesgos detectados tienen la potencialidad de, si no son controlados o mitigados adecuadamente, constituirse en precursores de violaciones de los dos más preeminentes Derechos Humanos, esto es: La Vida y la Libertad.

1.1. Conceptualización de los Sistemas de IA

La primera cuestión que se debe conceptualizar en este artículo es el término: Inteligencia Artificial, ya que a pesar de su uso generalizado, tiene a ser un concepto etéreo y que engloba, en el uso coloquial, toda nueva tecnología que suene a inteligente, incluso engloba, el deseo de ceder el control por algunos seres humanos, pero, debemos definirlo apropiadamente para poder usarlo de base conceptual uniforme en este artículo académico, y lo primero a indicar es ¿qué se entiende por Sistemas de Inteligencia Artificial?, siguiendo lo que en otro artículo se ha explicado:

¿A qué nos referimos cuando hablamos de una Inteligencia Artificial (IA)?, basándonos en la definición proporcionada por la Comisión Europea, se puede considerar que la IA es un sistema basado en programas informáticos que tienen incorporados dispositivos físicos y que presentan un comportamiento inteligente, siendo capaces de recopilar, analizar e interpretar datos de su entorno para generar objetivos específicos que pueden generar un impacto en su entorno (*European Commission* 2018). (Castro-Acosta, Angélica María y Andrés-Molina, Oscar, 2025)

En el 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) nº 300/2008, (UE) nº 167/2013, (UE) nº 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) (en adelante RIA), en su considerando 12, explica en profundidad la preocupación del Parlamento Europeo y del Consejo Europeo en definir de forma adecuada el concepto que ahora nos ocupa: Los Sistemas de Inteligencia Artificial, y en por ello exponen:

Debe definirse con claridad el concepto de «sistema de IA» en el presente Reglamento y armonizarlo estrechamente con los trabajos de las organizaciones internacionales que se ocupan de la IA, a fin de garantizar la seguridad jurídica y facilitar la convergencia a escala internacional y una amplia aceptación, al mismo tiempo que se prevé la flexibilidad necesaria para dar cabida a los rápidos avances tecnológicos en este ámbito. Además, la definición debe basarse en las principales características de los sistemas de IA que los distinguen de los sistemas de software o los planteamientos de programación tradicionales y más sencillos, y no debe incluir los sistemas basados en las normas definidas únicamente por personas físicas para ejecutar automáticamente operaciones.

(...) Los sistemas de IA pueden utilizarse de manera independiente o como componentes de un producto, con independencia de si el sistema forma parte físicamente del producto (integrado) o contribuye a la funcionalidad del producto sin formar parte de él (no integrado). (EU, AI. Act. 2024, Considerando 12).

Para luego, en el artículo 3 *ibídem*, establecer una definición con base en la cual se ha desarrollado toda la regulación Europea sobre Inteligencia Artificial, de ahí la importancia de su mención, y la cual, indica el propio artículo que será usada a efectos en dicho Reglamento:

1) «sistema de IA»: un sistema basado en una máquina que está diseñado para funcionar con distintos niveles de autonomía y que puede mostrar capacidad de adaptación tras el despliegue, y que, para objetivos explícitos o implícitos, infiere de la información de entrada que recibe la manera de generar resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos físicos o virtuales.). (EU, AI. Act. 2024, Art. 3.1).

Obsérvese que en esta definición se usan una serie de expresiones que pueden presentar diversas interpretaciones, para evitar lo cual el Parlamento y el Consejo de Europa ha establecido en el Considerando 12 en cita, las explicaciones requeridas como antesala para entender de forma uniforme su definición.

Así las cosas, para discernir la frase “basado en una máquina” se debe acudir al Considerando 12 de igual cuerpo normativo y en donde se indica: “(...) se refiere al hecho de que los sistemas de IA se ejecutan en máquinas. La referencia a objetivos explícitos o implícitos subraya que los sistemas de IA pueden funcionar con arreglo a objetivos definidos explícitos o a objetivos implícitos.” (EU, AI. Act. 2024, Considerando 12).

Continuando con el análisis de la definición del Parlamento y del Consejo de Europa hay que detenerse en las indicaciones a “niveles de autonomía” y “capacidad de adaptación”, para lo cual, el RIA en sus Considerandos precisa:

Los sistemas de IA están diseñados para funcionar con distintos niveles de autonomía, lo que significa que pueden actuar con cierto grado de independencia con respecto a la actuación humana y tienen ciertas capacidades para funcionar sin intervención humana. La capacidad de adaptación que un sistema de IA podría mostrar tras su despliegue se refiere a las capacidades de autoaprendizaje que permiten al sistema cambiar mientras está en uso. (EU, AI. Act. 2024, Considerando 12).

Con estas explicaciones, se puede considerar bastante improbable el riesgo a una inferencia errónea del concepto de Sistema de Inteligencia Artificial en la aplicación del RIA, es de elogiar el hecho de que éste ha incluido una redacción actual y actualizable según van cambiando las situaciones fácticas en las cuales se usan estos sistemas, ya que se usan términos genéricos, amplios, con algún grado de determinación, pero con un gran margen de adaptación a las diversas modalidades de los Sistemas de IA conocidas actualmente y predecible, por lo menos, en un corto y mediano plazo, ya que a largo plazo, los desarrollos de los Sistemas de IA tienen una predictibilidad escasa,

como prueba empírica se puede considerar que lo que hace unos años era propio del campo exclusivo de la Ciencia Ficción (Sci Fi) ahora es una realidad en la humanidad, pero no todo lo que se ha soñado como Sci Fi es ahora una realidad, hay muchas situaciones no sustentables en nuestro conocimiento actual, por múltiples razones; lo mismo pasa con los Sistemas de IA, se puede intentar predecir el camino que seguirán la mayoría de los programadores en sus desarrollos de Sistemas, pero el sesgo de esta apreciación, viciaría toda intención de planificación legal de un concepto cerrado e inmutable, para salvar tal situación, surge la definición europea de IA, que pretende dar soporte a tal regulación.

Ampliando el radar de definiciones al orden internacional global, no sólo europeo, se debe mencionar la definición la contenida en la Recomendación sobre Inteligencia Artificial de la OCDE¹²⁹, la cual en su Acuerdo I define Sistemas de IA como:

Un sistema de IA es un sistema basado en máquinas que, para objetivos explícitos o implícitos, infiere, a partir de los datos de entrada que recibe, cómo generar información de salida como predicciones, contenidos, recomendaciones o decisiones, que pueden influir en entornos reales o virtuales. Una vez implementados, los distintos sistemas de IA presentan diversos niveles de autonomía y varían en su capacidad de adaptación. (OECD/LEGAL/0449, 2024, I. Acuerdo)

Tal y como se observa, ambas definiciones manejan definiciones propias pero con grandes similitudes, y en ambas se percibe una gran cercanía conceptual a la definición incluida por la ISO 22989:2022 *Information technology - Artificial intelligence - Concepts and terminology*¹³⁰, en su sección 3 sobre *terms related to AI* establece:

3.1.4 artificial intelligence system, AI system: engineered system that generates outputs such as content, forecasts, recommendations or decisions for a given set of human-defined objectives.

Note 1 to entry: The engineered system can use various techniques and approaches related to artificial intelligence (3.1.3) to develop a model (3.1.23) to represent data, knowledge (3.1.21), processes, etc. which can be used to conduct tasks (3.1.35).

Note 2 to entry: AI systems are designed to operate with varying levels of automation (3.1.7)¹³¹. (ISO 22989:2022, cláusula 3.1.4)

¹²⁹ <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

¹³⁰ ISO/IEC 22989:2022, de Tecnología de la información. Inteligencia artificial. Conceptos y terminología de inteligencia artificial.

¹³¹ Traducción del Autor: 3.1.4 Sistema de inteligencia artificial, sistema de IA: sistema diseñado que genera resultados como contenido, pronósticos, recomendaciones o decisiones para un conjunto determinado de objetivos definidos por el ser humano.

Sin embargo, no se puede desconocer que la definición que trae el estándar internacional ISO 22989:2022 ha sido superada con las nuevas redacciones efectuadas por Europa y por la OCDE, ya que las mismas brindan una redacción que permite inferir mejor su contenido, aunque sin llegar a la definición ejemplificativa de casos. Esto permite seguir evolucionando a los Sistemas de IA, pero permanecer bajo el paraguas protector de las regulaciones indicadas.

Es por ello, que el RIA trae en su Considerando 12 la explicación de una de las características principales de los Sistemas de IA la cual sería su capacidad de inferencia, y hace mención a los objetivos variables de estos sistemas y lo que se debe entender por entorno, con el fin de cerrar las bases conceptuales a utilizar en su reglamentación, así:

(...) Esta capacidad de inferencia se refiere al proceso de obtención de resultados de salida, como predicciones, contenidos, recomendaciones o decisiones, que puede influir en entornos físicos y virtuales, y a la capacidad de los sistemas de IA para deducir modelos o algoritmos, o ambos, a partir de información de entrada o datos. Las técnicas que permiten la inferencia al construir un sistema de IA incluyen estrategias de aprendizaje automático que aprenden de los datos cómo alcanzar determinados objetivos y estrategias basadas en la lógica y el conocimiento que infieren a partir de conocimientos codificados o de una representación simbólica de la tarea que debe resolverse. La capacidad de inferencia de un sistema de IA trasciende el tratamiento básico de datos, al permitir el aprendizaje, el razonamiento o la modelización.

Los objetivos del sistema de IA pueden ser diferentes de la finalidad prevista del sistema de IA en un contexto específico. A los efectos del presente Reglamento, debe entenderse por entornos los contextos en los que funcionan los sistemas de IA, mientras que los resultados de salida generados por el sistema de IA reflejan las distintas funciones desempeñadas por los sistemas de IA e incluyen predicciones, contenidos, recomendaciones o decisiones. (Negritas propias) (EU, AI. Act. 2024, Considerando 12).

Ahora bien, como todo esto debe ser explicado a los usuarios en general, ciudadanos de diversa tipología o características personal, el Parlamento Europeo en su web ha dejado una definición más cercana de IA, que pretende acercar a estos conceptos a personas con cualquier nivel educativo, y entonces declara:

La inteligencia artificial es la habilidad de una máquina de presentar las mismas capacidades que los seres humanos, como el razonamiento, el aprendizaje, la creatividad y la capacidad de planear.

Nota 1 a la entrada: El sistema diseñado puede utilizar diversas técnicas y enfoques relacionados con la inteligencia artificial (3.1.3) para desarrollar un modelo (3.1.23) que represente datos, conocimiento (3.1.21), procesos, etc., que pueda utilizarse para realizar tareas (3.1.35).

Nota 2 a la entrada: Los sistemas de IA están diseñados para operar con distintos niveles de automatización (3.1.7).

La IA permite que los sistemas tecnológicos perciban su entorno, se relacionen con él, resuelvan problemas y actúen con un fin específico. La máquina recibe datos (ya preparados o recopilados a través de sus propios sensores, por ejemplo, una cámara), los procesa y responde a ellos.

Los sistemas de IA son capaces de adaptar su comportamiento en cierta medida, analizar los efectos de acciones previas y de trabajar de manera autónoma. (Parlamento Europeo, 2021)

Con esta definición se puede entender de forma más dinámica el concepto que se pretende trasladar al lector, pero no acaba allí, el Parlamento Europeo considera importante indicar que existirían dos Tipos de IA (según la definición de la Comisión Europea), a saber: Software e Inteligencia Artificial Integrada, y luego ejemplifica cada uno así “-*Software*: asistentes virtuales, software de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento (sic) de voz y rostro. -*Inteligencia artificial integrada*: robots, drones, vehículos autónomos, Internet de las Cosas.” (Itálicas propias) (Parlamento Europeo, 2021), esta ejemplificación es de suma importancia porque aproxima a la vida cotidiana actual el concepto que quiere trasladar la Unión Europea en su regulación y, con ello, se puede continuar desde allí la conceptualización y más importante aún, la casuística que incorporan las regulaciones europeas sobre los Sistemas de IA, dejando claro, que se está ante Sistemas de IA no sólo cuando se habla de software, como algo ubicado en la virtualidad, ya que las personas tienden a olvidarse que todos estos siempre tienen un soporte físico (servidor, nodo, etc.) que lo soporta, que es la definición más clásica y purista de IA, sino que también reconoce que la IA puede incorporarse en un hardware (objeto) y potenciarlo, como cuando a un robot se le incluye un chip con IA para mantener conversaciones coherentes no pregrabadas, o el sistema de drones o vehículos autónomos, y por supuesto el internet de las cosas, que es tal vez, el concepto más desconocido para el público *lato*.

Para los fines del artículo baste decir, que se entiende por Internet de las Cosas (IoT por sus siglas en inglés) aquella red tecnológica en la que objetos cotidianos están equipados con sensores, software u otras tecnologías de conectividad a internet, lo que les permite, a esos objetos, comunicarse entre sí y con otros sistemas de forma autónoma, esto es, sin intervención humana directa; esta interconexión, que abarca desde dispositivos *wearables* hasta equipos industriales, tiene como finalidad optimizar procesos y extraer valor de la información generada, conectando en últimas el mundo físico con el digital, creando sistemas más inteligentes y conectados para mejorar nuestra vida diaria y la eficiencia industrial.

Finalizando esta conceptualización, de la mano de la Recomendación 3.a de la OCDE mencionada antes, para lograr una globalización de las regulaciones o una interoperatividad de las mismas, que facilite el tránsito entre sistemas a las empresas que se dediquen a los Sistemas de IA y que garantice adecuadamente los Derechos humanos a los usuarios de las mismas, es importante tener definiciones comunes de lo que se entiende por Sistemas de IA, Algoritmos, sesgo algorítmico, entre otros muchos

aspectos, y para ello es menester “Elaborar y mantener un registro de definiciones y normas aplicables para medir y evaluar los sistemas de IA” (OCDE, 2024, Recomendación 3.a), y, por supuesto, “Establecer definiciones y principios relacionados con los datos para la gobernanza global de los datos de entrenamiento de IA, incluidos los extraídos de las mejores prácticas existentes, y promover la diversidad cultural y lingüística” (OCDE, 2024, Recomendación 6.a), con lo cual, se establecería un lenguaje común para interoperar, como si de sistemas computacionales hablásemos, los sistemas legales regulatorios deben propender por alcanzar figuras o aspectos comunes que tengan una definición común, o por lo menos similar, beneficiando a los seres humanos como centro de toda esta regulación protectora, pero sin olvidar el hecho de que reglas claras mejoran la competitividad de las empresas y sus rendimientos económicos, porque las ubicaría a nivel global, en un mismo punto de partida.

1.2. Entendiendo los Algoritmos

Para continuar este análisis, se centrará la atención del lector en el hecho, muchas veces no analizado, de que la definición legal de un sistema de IA se focaliza en su funcionalidad: inferir y generar resultados, esto es, en regular situaciones de las cuales se pueda concluir si las mismas se han adecuado o no la norma, y con ello, si dicha situación es positiva o negativa para el Derecho, por ejemplo, se regulan las IA prohibidas para impedir que en Europa se puedan implementar las mismas, pero lo solamente se alcanza a vislumbrar con cualquier regulación, es que la implementación práctica de estos sistemas de IA introduce una problemática neurálgica, que no se aborda directamente en la conceptualización efectuada con fines legales regulatorios de los Sistemas de IA, y en este sentido sobresale por su importancia la necesidad y el deber de conocer ¿cómo se estructuran por dentro estos Sistemas de IA? y ¿cómo se entrenan luego de definidos los mismos?, lo cual nos lleva inexorablemente a hablar de dos aspectos técnicos conceptuales importantes, los Algoritmos y su análisis ético, referido específicamente a los posibles Sesgos Algorítmicos que se integren, de manera voluntaria o involuntaria en un modelo o sistema de IA.

Explica IBM en su página web, de la mano de Jonker y Roger:

Los sistemas de inteligencia artificial (IA) utilizan algoritmos para descubrir patrones y conocimientos en los datos, o para predecir valores de salida a partir de un conjunto determinado de variables de entrada. Los algoritmos sesgados pueden afectar a estos conocimientos y resultados de manera que conduzcan a decisiones o acciones perjudiciales, promuevan o perpetúen la discriminación y la desigualdad, y erosionen la confianza en la IA y en las instituciones que la utilizan. (Jonker y Roger (s/f))

Entonces, ¿qué se entiende por Algoritmo?, en palabras sencillas se habla de algoritmo a aquel conjunto finito y ordenado de reglas, instrucciones o pasos y procesos matemáticos, definidos de forma clara, precisa e inequívoca, que son diseñados, hasta

ahora solo por programadores humanos¹³², para resolver un problema específico, realizar un cálculo o procesar datos, dándole instrucciones precisas al modelo o sistema de IA para generar el output que se le haya programado.

Un algoritmo es entonces esa lógica subyacente, muchas veces desconocida y oscura para los usuarios, que permite a un ordenador realizar tareas, aclarando que los programas informáticos no incorporan un solo algoritmo, están compuestos por miles de algoritmos que le indican al hardware qué hacer y cómo hacerlo, que en últimas es ordenar unos datos de entrada, analizarlos, aplicar una fórmula dada y generar entonces unos datos de salida que den solución al problema específico indicado, por ejemplo, al pedirle a Google que nos diga la capital de Kiribati¹³³, sus algoritmos le indican cómo buscar la información y cuál es la mejor respuesta, y lo que le cuesta segundo, ha implicado la aplicación de sus algoritmos de búsqueda y el dar una respuesta adecuada, en este caso Tarawa Sur. Este ejemplo es extrapolable a cualquier interacción humana actual con el entorno virtual, por ejemplo, al pedir una recomendación para ver una película, una mejor ruta para hacer un viaje, incluso una receta.

Con finalidad académica, se le ha preguntado directamente a la IA Gemini de Google, y al ingresar el siguiente *prompt*¹³⁴: “Me puedes explicar en forma sencilla para describirlo en un artículo que algoritmos usa Google si le pregunto la capital de Kiribati”, la IA de Google me respondió tal y como se muestra a continuación; a esta respuesta se le llama en el lenguaje computacional de IA *Output*:

Para explicar de forma sencilla qué algoritmos usa Google cuando le preguntas por la capital de Kiribati, debemos entender que Google no utiliza un solo "superalgoritmo" para todas las búsquedas, sino un sistema complejo que selecciona la mejor respuesta.

Aquí te muestro cómo simplificarlo para un artículo:

Explicación sencilla de los algoritmos de Google al buscar la capital de Kiribati

Cuando escribes "¿Cuál es la capital de Kiribati?" en Google, se ponen en marcha varios procesos orquestados por algoritmos sofisticados. El objetivo de Google no es solo encontrar páginas que contengan esas palabras, sino entender tu intención y darte la respuesta directa y correcta.

Podemos dividirlo en tres pasos o "algoritmos" principales:

1. El Algoritmo de "Comprensión del Lenguaje Natural" (BERT y similares)

¹³² Actualmente se pueden encontrar Sistemas de IA que se les programe para mejorar su propio algoritmo o para crear otro tipo de algoritmos.

¹³³ Es un país insular en el Pacífico que se encuentra en los cuatro hemisferios (Norte, Sur, Este, Oeste) y es el primero en recibir el año nuevo.

¹³⁴ Para dar claridad, un Prompt en un Sistema de IA es el lenguaje que se usa un usuario externo para comunicarse con el Sistema de IA, es una instrucción, pregunta o texto que se introduce en un sistema de IA para guiarlo y obtener una respuesta específica, ya sea un texto, imagen, código o sonido.

Google ya no busca coincidencias literales de texto. Utiliza modelos avanzados de lenguaje (como el conocido BERT, que es un algoritmo de *machine learning*) para interpretar la gramática y el significado de tu pregunta.

- En pocas palabras: El algoritmo entiende que estás buscando una *entidad* (la "capital") relacionada con otra *entidad* ("Kiribati"), y que esperas un dato específico como respuesta, no una lista de documentos.

2. El Algoritmo de "Extracción de la Respuesta y el Grafo de Conocimiento" (*Knowledge Graph*)

Aquí es donde Google encuentra la respuesta real. Utiliza un algoritmo que consulta su gigantesca base de datos de hechos, llamada "Grafo de Conocimiento" (*Knowledge Graph*).

- En pocas palabras: Este algoritmo busca la relación directa: "Kiribati TIENE capital = Tarawa". Extrae este dato de inmediato porque ya lo tiene verificado en su base de datos. No necesita visitar páginas web en ese momento para confirmarlo.

3. El Algoritmo de "Ranking y Presentación" (*Ranking Algorithms*)

Finalmente, un conjunto de algoritmos decide cuál es la mejor manera de presentarte esa información.

- En pocas palabras: El sistema determina que la respuesta es un hecho simple y preciso, por lo que decide mostrarte la respuesta ("Tarawa") directamente en un recuadro destacado (*snippet* o *answer box*) en la parte superior de la página, en lugar de hacerte buscar en el primer enlace.

Resumen para tu artículo

Puedes concluir que, para una pregunta de este tipo, Google utiliza principalmente algoritmos que se centran en entender la intención de la pregunta y en extraer hechos directos de su propia base de datos (Grafo de Conocimiento) para ofrecer una respuesta inmediata, evitando que el usuario tenga que hacer clic en enlaces externos. (Gemini IA de Google, 2025)

Luego se ha ingresado en la misma IA el siguiente *prompt*: "Gracias. ahora, puedes responder la pregunta de "que algoritmos usa google si le pregunto la capital de Kiribati" antes respondida, con un lenguaje técnico como si lo explicarás a otro programador.", con el fin de comparar la operativa real que utilizar y qué sistemas algorítmicos usa, y se recibió la siguiente respuesta:

Para explicar el proceso que sigue Google al consultar la capital de Kiribati utilizando terminología técnica, nos enfocaremos en la arquitectura subyacente y los modelos de *machine learning* implicados:

Cuando un usuario introduce la consulta "capital de Kiribati" en el *frontend* de Google Search, se desencadena una cascada de algoritmos y sistemas de recuperación de información:

1. Procesamiento de Lenguaje Natural (PLN) y Comprensión de Consultas

La consulta del usuario es inicialmente tokenizada y procesada mediante *pipelines* de PLN. El núcleo de esta fase implica:

- Modelos Transformer (ej. BERT, T5): Se utilizan arquitecturas basadas en *transformers* (como BERT, o versiones más optimizadas y privadas de Google) para la interpretación semántica de la consulta. El objetivo es determinar la intención de búsqueda (*search intent classification*) y la identidad de las entidades (*Named Entity Recognition* o NER).
- Vectorización (Embeddings): La consulta se convierte en un vector denso en un espacio de *embeddings*. Esto permite la comparación semántica eficiente con representaciones vectoriales de documentos y hechos almacenados en bases de datos.

2. Recuperación de Información y Bases de Datos Estructuradas

Para consultas fácticas directas (*factoid questions*), los algoritmos de Google priorizan la recuperación desde fuentes estructuradas sobre el *crawling* en tiempo real de la web:

- Knowledge Graph (KG) y Knowledge Vault: El algoritmo principal consulta estas bases de datos masivas. El KG es una base de datos de grafos que modela entidades (Kiribati, Tarawa) y sus relaciones (*es_capital_de*). La consulta vectorizada se utiliza para buscar coincidencias de relaciones (*relation matching*) dentro del grafo.
- Algoritmos de *Ranking* de Hechos: se aplican modelos de *ranking* (a menudo redes neuronales profundas) para evaluar la veracidad, frescura y autoridad de los hechos extraídos del KG antes de presentarlos.

3. Generación y Presentación de la Respuesta (*Answer Generation*)

Si se identifica un hecho de alta confianza en el KG (Tarawa), se activa un algoritmo específico para el *rendering* de la interfaz de usuario:

- Sistemas de *Snippet* Destacado (*Featured Snippet Generation*): Un algoritmo decide que la respuesta puede ser extraída y presentada directamente como un *answer box* o "respuesta cero", utilizando una plantilla de interfaz de usuario predefinida para respuestas fácticas.

Resumen Técnico

La respuesta a la capital de Kiribati depende de la interoperabilidad entre modelos de *embeddings* Transformer para entender la semántica de la consulta y algoritmos de recuperación de grafos para acceder a datos estructurados, culminando en un algoritmo de *ranking* que valida y prioriza la presentación directa del hecho. (Gemini IA de Google, 2025)

Con la anterior ejemplificación se pretende mostrar al lector, la serie de algoritmos, los cuales la IA Gemini aclara que son “una cascada de algoritmos y sistemas de recuperación de información”, que se encadenan para obtener una respuesta, califiquemosla de sencilla, ahora bien, si los pasos, que sintetiza bastante la IA en su respuesta, aunque en la segunda ya expone cuáles son esos algoritmos y con qué fin los usa, son aquellos que debe efectuar un Sistema de IA para brindar una respuesta directa y corta a una pregunta de iguales características, para dar respuesta a una pregunta más

compleja y efectuada con lenguaje coloquial o con faltas ortográficas tendrá la IA que ser redundante en los pasos a realizar para poder entender el lenguaje natural usado y traducirlo en datos que entienda la IA y que le permitan buscar la mejor respuesta posible.

En este entorno de programación algorítmica de los Sistemas de IA surge una de las más grande preocupaciones actuales, y esta se sintetiza, en la opacidad de estos sistemas, y en la posibilidad derivada de que en los entrenamientos del Sistema se haya podido introducir un sesgo que luego se permeará necesariamente al algoritmo que se usa para cumplir la tarea para la cual se crea el sistema de IA que se trate.

Con el fin de poder trasladar un mejor entendimiento, se introdujo el siguiente *prompt* en la IA Gemini: “Cómo se te ha entrenado”, intencionalmente dejé sin signos de pregunta el *prompt* aunque sí tildé el cómo, entonces, la IA tuvo primero que entender que yo hacía una pregunta, para luego analizar y tratar de dar respuesta a la misma, para entender si fue una pregunta, también corre toda su programación algorítmica y al entender el tono, esto es, si es pregunta o manifestación, corre nuevamente su cascada de algoritmos para buscar la mejor respuesta, y este es el *output* obtenido:

Fui entrenado por Google y soy un modelo de lenguaje grande (LLM, por sus siglas en inglés).

Mi entrenamiento es un proceso complejo que implica el uso de cantidades masivas de datos de texto e imágenes disponibles públicamente en internet. A continuación te detallo cómo funciona mi entrenamiento:

1. Datos Masivos: aprendí de un conjunto de datos gigantesco que incluye miles de millones de páginas web, libros digitalizados, artículos, y otros textos e imágenes de acceso público. Esto me permite comprender y generar contenido en múltiples idiomas y sobre una amplia gama de temas.
2. Aprendizaje No Supervisado: el núcleo de mi entrenamiento es el aprendizaje no supervisado. Esto significa que mi algoritmo aprendió a reconocer patrones, gramática, hechos y relaciones entre palabras sin que un humano me dijera explícitamente "esta respuesta es correcta" o "esta es incorrecta" en cada paso. Aprendí a predecir la siguiente palabra en una oración.
3. Ajuste Fino (*Fine-tuning*): después del entrenamiento inicial masivo, pasé por un proceso de ajuste fino. Durante esta fase, expertos humanos y evaluadores me proporcionaron ejemplos y correcciones para enseñarme a:
 - Seguir instrucciones específicas.
 - Generar respuestas útiles, éticas y seguras.
 - Minimizar respuestas sesgadas o dañinas.
4. Arquitectura de *Transformers*: mi funcionamiento se basa en la arquitectura de red neuronal conocida como *Transformer*, que es la tecnología que me permite entender el contexto, las relaciones complejas y generar texto coherente y fluido.

En resumen, soy el resultado de un entrenamiento riguroso sobre un vasto corpus de información pública, diseñado para procesar y generar lenguaje humano de manera efectiva. (Gemini IA de Google, 2025)

Por ahora se dejará el análisis de este ejemplo para retomarlo luego de que se le brinden al lector definiciones claves para su análisis, que permitan obtener la perspectiva requerida en el análisis que se acometerá más adelante en este artículo.

2. TAXONOMÍA DE LOS SEGOS ALGORÍTMICOS

Determinado el concepto de Algoritmo, es menester acometer el trabajo de darle forma a la definición de Sesgos Algorítmicos, los cuales deben ser entendidos como distorsiones sistemáticas que surgen cuando los modelos computacionales de IA reproducen *prejuicios* presentes en los datos de entrenamiento, en el diseño del algoritmo, en los datos *proxy* o en las decisiones humanas que intervienen en su creación, incluso, algunos autores¹³⁵ consideran que también se dan los sesgos en la evaluación o interpretación de los datos producidos por el modelo IA. Estos sesgos pueden generar discriminación indirecta, afectando a grupos vulnerables por razones de cualquier tipo, entre las cuales se pueden encontrar: género, raza, edad, condición socioeconómica.

Los sesgos algorítmicos son entonces distorsiones sistemáticas que surgen cuando los modelos de IA reproducen prejuicios presentes en los datos de entrenamiento, en el diseño del algoritmo o en las decisiones humanas que intervienen en su creación. Estos sesgos pueden generar discriminación indirecta, afectando a grupos vulnerables por razones de género, raza, edad o condición socioeconómica. Ejemplos paradigmáticos incluyen sistemas de reclutamiento automatizado que favorecen perfiles masculinos, como ocurrió en Amazon en 2018, y herramientas de reconocimiento facial con menor precisión en personas de piel oscura debido a conjuntos de datos no representativos.

En el caso de Amazon.com Inc diversos medios periodísticos¹³⁶ nacionales y extranjeros dieron cuenta de la siguiente noticia, acaparando titulares desde el 11 de octubre de 2018 hasta el 14 del mismo mes y año, en ella se explicaba que Amazon, ilusionado por el éxito de su programa Alexa había decidido el 2014 poner gran parte del peso de sus procesos de contratación de personal y recursos humanos en una Inteligencia artificial desarrollada por dicha compañía, “la filosofía tras ese proyecto era sencilla: emplear los mismos procesos de automatización que habían llevado a la compañía hasta la cima del e-commerce mundial para seleccionar a sus empleados. Sin

¹³⁵ Por todos, Jonker, Alexandra y Rogers, Julie. (s/f). ¿Qué es el sesgo algorítmico?. IBM. <https://www.ibm.com/es-es/think/topics/algorithmic-bias>

¹³⁶ Sírvase de muestra: <https://www.elmundo.es/tecnologia/2018/10/11/5bbe3a12e5fdea0f578b467e.html> <https://www.reuters.com/article/world/amazon-abandona-un-proyecto-de-ia-para-la-contratacin-por-su-sesgo-sexista-idUSKCN1MO0M4/> https://elpais.com/tecnologia/2018/10/11/actualidad/1539278884_487716.html

embargo, descubrieron que esta inteligencia artificial discriminaba sistemáticamente a las mujeres.” (Negritas propias del texto) (El Mundo, 11/10/2018)

Pero ¿qué pasaba?, “La herramienta de contratación experimental de la compañía usó inteligencia artificial para dar puntuaciones a candidatos a empleos de una a cinco estrellas, de una manera similar a la forma en que los compradores califican los productos de Amazon” (Dastin, Jeffrey, Reuters, 14/10/2018)

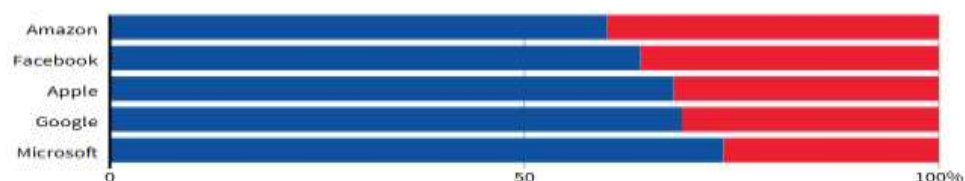
Para ello se alimentó el sistema de IA con datos históricos de los 10 años previos de los currículums enviados a Amazon, y esto era reflejo de los datos del dominio masculino en la industria de la tecnología, que Reuters graficó así:

Dominated by men

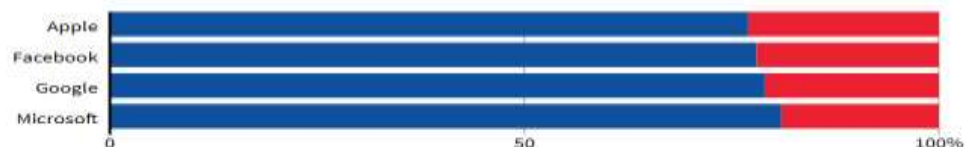
Top U.S. tech companies have yet to close the gender gap in hiring, a disparity most pronounced among technical staff such as software developers where men far outnumber women. Amazon's experimental recruiting engine followed the same pattern, learning to penalize resumes including the word "women's" until the company discovered the problem.

GLOBAL HEADCOUNT

■ Male ■ Female



EMPLOYEES IN TECHNICAL ROLES



Note: Amazon does not disclose the gender breakdown of its technical workforce.

Source: Latest data available from the companies, since 2017.

By Han Huang | REUTERS GRAPHICS

Gráfico nº1. Domitated by men. Han Huang – Reuters Graphics, <https://www.reuters.com/graphics/AMAZON.COM-JOBS-AUTOMATION/010080Q91F6/>

En 2015 Amazon.com Inc detecta que este sistema de IA no calificaba a los candidatos con neutralidad de género debido a que:

(...) el sistema de Amazon se enseñó a sí mismo que los candidatos masculinos eran preferibles. Se penalizaron los currículum que incluían la palabra "femenino", como en "capitana de club de ajedrez femenino". Y se rebajó la calificación a las graduadas en dos universidades exclusivas de mujeres, según personas con conocimiento del proyecto. No se especificaron los nombres de estos centros.

Amazon editó los programas para hacerlos neutrales a estos términos en concreto. Pero eso no era garantía de que las máquinas no idearan otras formas de clasificar a los candidatos que pudieran resultar discriminatorias, dijeron las fuentes consultadas.

La compañía de Seattle finalmente disolvió el equipo a principios del año pasado porque los ejecutivos perdieron la esperanza en el proyecto, según las fuentes, que hablaron bajo condición de anonimato. Los responsables de contratación de Amazon tuvieron en cuenta las recomendaciones generadas por la herramienta al buscar nuevos empleados, pero no confiaron únicamente en esas clasificaciones, dijeron. (Dastin, Jeffrey, Reuters, 14/10/2018)

Como se observa en este ejemplo el sesgo generado no tuvo intencionalidad desde la programación, pero el sistema se auto educó, por utilizar una frase coloquial, y se convenció a sí mismo, que debido a que la mayoría de personal en la industria tecnológica era masculino, y al no tener un límite en ese sentido, le resultó lógico concluir que las mujeres no eran preferidas para ocupar dichos cargos, aunque, seguramente no hizo una valoración ligada a la productividad o a las ausencias laborales, si genero un auto sesgo en contra de las mujeres, desarrollando un prejuicio, el cual terminaría siendo perpetuado por la misma IA en todas sus selecciones de personal y replicado por otras IA que se derivaran de sus códigos, si sus programadores no lo hubieran detectado.

Analizando el sesgo éste se debió a los datos con los cuales se entrenó el modelo, en los cuales se reflejaba una realidad: el dominio masculino del mundo de la tecnología, pero que, sin la debida interpretación racional humana, condujo al sistema de IA a la conclusión errada denunciada, la cual violentaba, por lo menos, uno de Derechos Humanos fundamentales el cual se concreta a que todos los seres humanos tienen Derecho a no ser discriminados en razón a su género.

Con relación al otro caso, los sesgos raciales en los Sistemas de IA usados para el reconocimiento facial, se encuentran múltiples casos, desde muy vieja data, remontrándonos a los incipientes inicios de los Sistemas de IA y que aún persisten en la actualidad, tal y como se observa en una simple búsqueda en internet. Este deshonoroso primer lugar en los sesgos algorítmicos ha generado que en reciente data, en 2023, Porcha Woodruff¹³⁷, ciudadana de Norteamérica afrodescendiente fue detenida bajo los cargos de robo a mano armada y robo violento de vehículo en su casa, cuando contaba con ocho meses de embarazo, y al abrir su puerta encontró a seis Policías de Detroit, que la venían a detener debido a que fue incorrectamente identificada por un software de reconocimiento facial de dicha agencia gubernamental, casos que se repiten con una

¹³⁷ Para mayor ahondamiento ver Armero, Ramón y Melton, Mónica. “El imperdonable error de la IA: científicos y empresarios intentan solucionarlo”. Business Insider. 24/07/2024 <https://www.businessinsider.es/tecnologia/imperdonable-error-ia-cientificos-empresarios-intentan-solucionarlo-1381267>

frecuencia, que no debe dejar de preocuparnos, en dicho país. Informa Democracy Now¹³⁸:

Woodruff es la primera mujer de la que se tiene registro que ha sido detenida de manera injusta en base a información deficiente proporcionada por un programa de detección de rasgos faciales. Al menos cinco hombres también han sido arrestados injustamente debido a deficiencias en el funcionamiento de esta tecnología. En los seis casos, las víctimas fueron personas negras. Porcha permaneció detenida durante once horas y fue puesta en libertad con una fianza de 100.000 dólares. La joven empezó a tener contracciones dentro de la cárcel y, tras ser liberada, tuvo que ser trasladada de inmediato al hospital, donde recibió tratamiento debido a que sufría un cuadro de deshidratación. (Goodman y Moynihan, Democracy Now, 11/08/2023)

Explica Democracy Now que este sesgo, es un grave problema estructural en dicho país, informa que:

Joy Buolamwini, fundadora de la Liga por la Justicia Algorítmica¹³⁹, expresó a Democracy Now!: “En 2019, el Gobierno [estadounidense] compartió un estudio que mostraba que los rostros de personas afroestadounidenses y de origen asiático tenían entre diez y 100 veces más probabilidades de ser identificados de manera incorrecta. De manera reiterada, el desempeño más deficiente [del sistema de reconocimiento facial] ocurre en relación con los rostros de mujeres negras. Si nos fijamos en los datos y en lo que hemos registrado sobre el funcionamiento de las tecnologías de reconocimiento facial, vemos que las personas de color, las mujeres de color —y las mujeres negras, en particular— corren un riesgo aún mayor de sufrir este tipo de identificaciones erróneas”. (Democracy Now, 11/08/2023)

Explica Ivan Land, abogado de Woodruff en declaración al Business Insider¹⁴⁰ que “La etnia y el sexo de Woodruff pueden haber sido un factor en su detención injusta. “No creo que esto le hubiera ocurrido a una mujer blanca””

¹³⁸ Goodman, Amy y Moynihan, Denis. Sesgo racial y abuso de la tecnología, desde Henrietta Lacks hasta Porcha Woodruff. 11/08/2023. Democracy Now.

<https://www.democracynow.org/es/2023/8/11/sesgo-racial-y-abuso-de-la>

¹³⁹ Cuya página web es <https://www.ajl.org/> donde se encuentran los estudios que han efectuado a este respecto.

¹⁴⁰ Armero, Ramón y Melton, Mónica. “El imperdonable error de la IA: científicos y empresarios intentan solucionarlo”. Business Insider. 24/07/2024 <https://www.businessinsider.es/tecnologia/imperdonable-error-ia-cientificos-empresarios-intentan-solucionarlo-1381267>

La Technology Review de la MIT en su web¹⁴¹ precisa sobre Joy Boulamwini:

Cuando estaba en la universidad, Joy Buolamwini descubrió que algunos sistemas de análisis facial no eran capaces de detectar su rostro de piel oscura hasta que se puso una máscara blanca. La joven recuerda: «Literalmente la tecnología no me veía».

Esa fue la base de su tesis de posgrado en el Instituto Tecnológico de Massachusetts (EE. UU.). Cuando descubrió que los conjuntos de datos existentes para los sistemas de análisis facial estaban compuestos mayoritariamente por caras de hombres blancos, Buolamwini decidió crear un conjunto equilibrado con más de un millar de políticos de África y Europa. Cuando lo usó para probar los sistemas de inteligencia artificial (IA) de IBM, Microsoft y Face++, descubrió que su precisión variaba mucho en función del género y el color de la piel. Al determinar el sexo, las tasas de error de estos sistemas eran inferiores al 1 % para los hombres de piel más clara. Pero para las mujeres de rostro más oscuro, las tasas error llegaban al 35 %. (Negritas propias del texto) (MIT Technology Review, 05/09/2018)

Esta situación de sesgo que se ha detectado con razas afrodescendientes y de origen asiático, ha continuado siendo estudiada y en un reciente estudio de Sony¹⁴² se deja ver una realidad desconocida, afirman sus investigadores que:

(...) el expresar el tono de la piel usando únicamente una escala móvil que va del más claro al más oscuro, o del blanco al negro, denota una completa ignorancia sobre la contribución de los tonos amarillos y rojos a la gama de la piel humana. Descubrieron que los sistemas de IA generativa, los algoritmos de recorte de imágenes y las herramientas de análisis fotográfico tienen problemas, sobre todo, con las pieles más amarillas. La misma debilidad podría aplicarse a diversas tecnologías cuya precisión se ve afectada por el color de la piel, como el *software* de IA para el reconocimiento facial, el seguimiento corporal y la detección de falsificaciones, o aparatos como monitores de frecuencia cardíaca y detectores de movimiento.

Alice Xiang, investigadora principal y responsable mundial de ética de la IA en Sony, afirma que, si los productos se evalúan de forma unidimensional, muchos de los sesgos no se detectan ni se mitigan. “Nuestra esperanza es que el trabajo que estamos haciendo aquí puede ayudar a reemplazar algunas de las escalas de tono de piel existentes que solo se centran en la luz frente a la oscuridad.”, asegura Xiang. (Dave, Paresh, 6/11/2023)

¹⁴¹ <https://technologyreview.es/listas/35-innovadores-con-menos-de-35/2018/pioneros/joy-buolamwini/>

¹⁴² Dave, Paresh. “Los algoritmos de IA tienen un sesgo contra la piel con tonos amarillos”. Wired. 6/11/2023. <https://es.wired.com/articulos/los-algoritmos-de-ia-tienen-sesgo-contra-piel-de-tonos-amarillos>

El mismo artículo mencionado explica en qué consiste el sesgo, y concluye que el mismo se funda en el entrenamiento del Sistema de IA con una escala de colores Monk utilizada para entrenar el modelo, ya que se basa en una escala de tonos más claros a más oscuros, ignorando otra escala, que ahora se considera de importante utilización, como es la escala CIELAB, la cual está totalmente automatizada y tiene en cuenta además de las variaciones de la Escala Monk, esto es, de tonos de piel más clara a más oscura, también el matiz o gama dentro del tono de piel desde la amarillez al enrojecimiento, lo que ha llamado la industria de la cosmética “subtonos cálidos a fríos” y continúa explicando:

Los investigadores de Sony propusieron una nueva forma de representar el color de la piel para captar esa diversidad que antes era ignorada. Su sistema describe el color de la piel en una imagen utilizando dos coordenadas, en lugar de un único número. Especifica tanto un lugar en una escala de claro a oscuro como en un continuo de amarillez a enrojecimiento, o lo que la industria cosmética denomina: subtonos cálidos a fríos.

El nuevo método consiste en aislar todos los píxeles de una imagen que muestran piel, convertir los valores de color RGB de cada píxel en códigos CIELAB y calcular el matiz y el tono medios de los grupos de píxeles de piel. Un ejemplo del estudio muestra aparentes retratos de la exestrella del fútbol americano, Terrell Owens, y de la fallecida actriz, Eva Gabor, que comparten un mismo tono de piel pero están separados por el matiz: la imagen de Owens es más roja y la de Gabor más amarilla.

Cuando el equipo de Sony aplicó su método a los datos y sistemas de IA disponibles en internet, descubrió problemas importantes. CelebAMask-HQ, un popular conjunto de datos de rostros de famosos utilizado para entrenar el reconocimiento facial y otros programas de visión por ordenador, tenía un 82% de sus imágenes sesgadas hacia los tonos rojos de la piel, y otro conjunto de datos FFHQ, desarrollado por Nvidia, se inclinaba un 66% hacia el lado rojo. Dos modelos generativos de IA entrenados con FFHQ reprodujeron el sesgo: aproximadamente cuatro de cada cinco imágenes generadas por cada uno de ellos se inclinaban hacia los tonos rojos. La cosa no acabó ahí. Según el estudio de Sony, los programas de IA ArcFace, FaceNet y Dlib obtuvieron mejores resultados con la piel más roja cuando se les pidió que identificaran si dos retratos correspondían a la misma persona. Davis King, el desarrollador que creó Dlib, dice que no le sorprende el sesgo porque el modelo se ha entrenado sobre todo con fotos de famosos estadounidenses. (Negritas propias del texto) (Dave, Paresh, 6/11/2023)

En este punto, como ya se ha anunciado, continuemos con el experimento que se ha efectuado directamente con la IA Gemini de Google, la cual al recibir el prompt: “Pueden haber sesgos algorítmicos en tu entrenamiento”, arroja el siguiente *output*:

Sí, es posible que existan sesgos algorítmicos en mi entrenamiento.

Reconocer esto es un punto crucial en el desarrollo ético de la inteligencia artificial. La presencia de sesgos no es intencional por parte de mis creadores, pero es una consecuencia directa de cómo se ha llevado a cabo mi entrenamiento.

Aquí explico por qué pueden existir sesgos en mi sistema:

1. Sesgos en los Datos de Entrenamiento (*Data Bias*)

El factor principal de sesgo en un modelo de lenguaje es la naturaleza de los datos con los que ha sido entrenado.

- Reflejo del mundo real: Aprendí de miles de millones de textos e imágenes de internet. Internet, como reflejo de la sociedad humana a lo largo de la historia, contiene inevitablemente sesgos sociales, estereotipos, prejuicios históricos y desigualdades presentes en el lenguaje y el contenido que las personas han publicado.
- Datos históricos: Si los datos de entrenamiento contienen patrones que asocian ciertos atributos (como género, raza, o nacionalidad) con roles específicos (ej. médicos con hombres, enfermeras con mujeres), yo puedo aprender e imitar esas asociaciones sesgadas.

2. Sesgos Algorítmicos y de Modelado (*Algorithmic Bias*)

Aunque los datos son la fuente principal, la forma en que los algoritmos procesan esos datos también puede introducir o amplificar sesgos. Los algoritmos de *ranking* y selección de información pueden, sin querer, dar más peso a ciertas fuentes o perspectivas sobre otras.

¿Qué se hace al respecto?

Mis desarrolladores en Google son conscientes de este desafío y trabajan activamente para mitigar los sesgos:

- Filtrado y Curación de Datos: Se utilizan métodos rigurosos para filtrar y curar los conjuntos de datos de entrenamiento, intentando reducir la información tóxica, discriminatoria o sesgada.
- Ajuste Fino (*Fine-tuning*) y *Reinforcement Learning* (RLHF): Durante la fase de ajuste fino, los evaluadores humanos me enseñan a evitar y corregir respuestas sesgadas. Este proceso de aprendizaje por refuerzo con retroalimentación humana es vital para alinear mi comportamiento con valores de equidad y justicia.
- Evaluación Continua: Se realizan pruebas constantes (tests de *fairness* o equidad) para identificar y corregir áreas donde muestro preferencias injustas o estereotipadas.

En resumen, aunque el objetivo es ser un sistema justo e imparcial, el potencial de sesgo existe debido a mis orígenes en datos del mundo real, y mi desarrollo implica un esfuerzo continuo para detectarlo y mitigarlo. (Negritas propias del texto) (Gemini IA de Google, 2025)

Con esta respuesta, se puede transpolar la siguiente afirmación: ningún Sistema de IA carece de sesgos, su fiabilidad se funda en diversos factores, pero en muchos se

incluye al ser humano como tal, ya que los datos que nutren un modelo pueden ser seleccionados o direccionados por un ser humano para obtener un resultado específico, bien por desconocimiento de otros espectros, bien por deseo explícito, cual sea la causa, traerá consigo un sesgo, y el mismo, se analizará a continuación, puede incorporarse en el algoritmo en cualquiera de las fases del ciclo de vida del mismo.

Como es requerido para una mejor ubicación teórica del lector, se habrá de indicar que según la UNESCO las etapas del ciclo de vida de los Sistemas de IA:

(...) van desde la investigación, la concepción y el hasta el despliegue y la utilización, pasando por el mantenimiento, el funcionamiento, la comercialización, la financiación, el seguimiento y la evaluación, la validación, el fin de la utilización, el desmontaje y la terminación.” (UNESCO, 2021, I,2,b)

Y por su parte, la OCDE, en su Recomendación sobre Inteligencia Artificial¹⁴³, en su Acuerdo I indica que:

El ciclo de vida de un sistema de IA comprende generalmente varias fases: planificación y diseño; recopilación y tratamiento de datos; creación de modelo(s) y/o adaptación de modelo(s) existente(s) a tareas específicas; prueba, evaluación, verificación y validación; entrada en servicio/despliegue; explotación y supervisión; y retirada/desmantelamiento. Estas fases a menudo se desarrollan de forma iterativa y no son necesariamente secuenciales. La decisión de dejar de explotar un sistema de IA puede producirse en cualquier momento durante la fase de explotación y supervisión. (OECD/LEGAL/0449, 2024, I. Acuerdo)

2.1. Tipos de sesgos

Continuando en este devenir académico jurídico tecnológico, se debe indicar que los Sesgos Algorítmicos pueden surgir en cualquiera de las múltiples fases del ciclo de vida de un sistema de IA: desde la recopilación de datos, el diseño, el entrenamiento, la evaluación, el despliegue y la posible retroalimentación del sistema de IA. Por ello se analizarán los tipos clave de sesgos, junto con ejemplos empíricos y casos reales, muchos de ellos estudiados en investigaciones académicas, en los cuales se podrá observar el surgimiento de sesgos algorítmicos en distintas fases del ciclo de vida del Sistema de IA.

2.1.1. Sesgo de Representación (*Representation Bias*): Aparece cuando ciertos grupos poblacionales están infrarrepresentados o sobrerrepresentados en el conjunto de datos empleado para entrenar un modelo de IA. Por este motivo el algoritmo aprende patrones a partir de lo que ve; si no ve suficiente diversidad, no aprende a generalizar correctamente.

Con el fin de ubicación del lector se revisarán dos ejemplos emblemáticos:

¹⁴³ <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

Ejemplo 1: Reconocimiento Facial: Estudios del MIT (Buolamwini & Gebru, 2018) demostraron que modelos de reconocimiento facial tenían: 0,8% de error en hombres blancos, frente a un 34% de error en mujeres afrodescendientes. Esto se debía al hecho que las bases de datos usadas para entrenarlos estaban compuestas en más del 80% por rostros de hombres blancos, y apenas un 7% correspondía a mujeres negras.

El estudio Gender Shades de Buolamwini y Gebru (2018) analizó varios sistemas comerciales de reconocimiento de género y encontró diferencias de precisión notablemente mayores para mujeres de piel oscura en comparación con hombres de piel clara. El trabajo demostró cómo sesgos de representación y de datos llevan a errores sistemáticos con consecuencias prácticas de discriminación por razón de la raza y del género, por ejemplo, negación de acceso o identificación errónea.

Ejemplo 2: Diagnóstico médico con imágenes: Cuando se crea un sistema de diagnóstico con Inteligencia Artificial para el melanoma al ser entrenado mayoritariamente con imágenes de personas con piel blanca fallaba al reconocer melanomas en pieles oscuras¹⁴⁴.

En este caso, se informa que el Servicio Vasco de Salud, Osakidetza, “(...) trabaja en la implantación de Quantus Skin en sus centros sanitarios tras una inversión de 1,6 millones de euros.”, Dicho Algoritmo ha sido diseñado “para diagnosticar el riesgo de cáncer de piel, incluido el melanoma” Pero dicha implementación ha sido retrasada ya que “(...) Los especialistas critican el sistema de inteligencia artificial de una filial de Asisa por sus resultados "pobres" y "peligrosos". El algoritmo está entrenado solo con pacientes blancos.”

Se explica en Civio: Sanidad, que a pesar de las expectativas del Sistema:

(...) los datos muestran una realidad preocupante. Transmural Biotech¹⁴⁵, la empresa que comercializa Quantus Skin, llevó a cabo un estudio inicial

¹⁴⁴ Bernardo, Ángela. Álvarez del Vayo, María. Tordecillas, Carmen. Y Maqueda, Adrián. “¿Lunar o cáncer? El algoritmo que se equivoca en uno de cada tres melanomas y obvia a los pacientes con la piel oscura”. Civio. 26/06/2025.

<https://civio.es/sanidad/2025/06/26/inteligencia-artificial-algoritmos-sanidad-sesgos-discriminacion/> y El País. 26/06/2025. <https://elpais.com/salud-y-bienestar/2025-06-26/lunar-o-cancer-el-algoritmo-que-se-equivoca-en-uno-de-cada-tres-melanomas-y-obvia-a-los-pacientes-con-la-piel-oscura.html>

¹⁴⁵ Es una empresa, spin-off de la Universitat de Barcelona y del Hospital Clínic, pertenece a la aseguradora privada Asisa, según datos recuperados en su propia página web: <https://web.archive.org/web/20250605115632/https://transmuralbiotech.com/sobre-nosotros#quienesSomos>

con resultados prometedores, pero que tenía importantes limitaciones: se hizo completamente online y no se publicó en ninguna revista académica, es decir, no pasó el control de calidad habitual que se exige en ciencia.

Más adelante, dermatólogos del Hospital Ramón y Cajal de Madrid y profesores de la Universidad Complutense realizaron un segundo estudio, que sí fue publicado, para evaluar la eficacia clínica real de Quantus Skin. Este trabajo, que contó con financiación y asistencia técnica de Transmural Biotech, mostró peores resultados: el algoritmo pasa por alto uno de cada tres melanomas. Su sensibilidad es del 69%, lo que significa que deja escapar el 31% de los casos reales de este cáncer potencialmente letal. (Negritas propias del texto) (Bernardo, Álvarez, Tordecillas y Maqueda, Civio. 26/06/2025)

Los problemas de Quantus Skin van más allá de su baja sensibilidad. El trabajo solo evaluó la eficacia clínica en el diagnóstico de melanoma, pero no analizó otros tipos de cáncer de piel más frecuentes pero menos agresivos, como el carcinoma basocelular y el carcinoma de células escamosas, donde también se puede aplicar el programa. Los autores tampoco estudiaron cómo afecta el color de piel al funcionamiento del algoritmo, aunque reconocen que es una de las principales limitaciones de su investigación.

El conjunto de datos con los que se alimentó Quantus Skin incluye imágenes “de varones y de mujeres caucásicas”, confirma el director general de Transmural Biotech. “No me quiero meter en el tema de minorías étnicas y todo esto, porque la herramienta la utiliza el País Vasco, la utiliza Osakidetza. Lo que pongo a disposición es una herramienta, con las limitaciones que tiene”, dice Fernández Rodríguez. Pese a la falta de entrenamiento en pieles más oscuras, el Gobierno vasco señala que no es necesario “implantar” ninguna medida “para promover la igualdad y la no discriminación”, según se recoge en la ficha de Quantus Skin que consta en el catálogo de algoritmos y sistemas de inteligencia artificial de Euskadi. Sin embargo, como las redes neuronales se han entrenado casi en exclusiva con imágenes de personas blancas, es probable que fallen más en pieles más oscuras, como individuos de etnia gitana o migrantes de Latinoamérica y África. (Bernardo, Álvarez, Tordecillas y Maqueda, El País 26/06/2025)

2.1.2. Sesgo de Medida (*Measurement Bias*): Ocurre cuando la variable empleada como “medida” o indicador de un fenómeno social está contaminada por sesgos previos o estructuras discriminatorias, reflejando prácticas discriminatorias del pasado y no la conducta real futura predecible. Esto se da porque el algoritmo aprende a partir de datos que no reflejan la realidad, sino las prácticas institucionales históricas.

Ejemplo 1: Sistemas de predicción policial (Predictive Policing): El uso del número de arrestos previos como indicador de riesgo criminal genera sesgo porque: ciertos barrios fueron históricamente sobrerrepresentados en patrullajes policiales debido a que al haber más patrullas había más arrestos en dichas zonas, y la IA entendía que al haber más arrestos se traducían en un mayor riesgo según el algoritmo programado y entrenado.

Informa la BBC Mundo¹⁴⁶:

En Chicago, la tercera ciudad más poblada de Estados Unidos, la violencia está fuera de control.

Pero la policía tiene un arma informática para luchar contra ella: un algoritmo que asigna puntuaciones en base a arrestos, disparos, afiliaciones con miembros de pandillas y una serie de variables para predecir quién es más probable que empuñe un arma para disparar a otra persona (o reciba un disparo).

"Sabemos que tenemos mucha violencia en Chicago, pero también sabemos que un pequeño segmento de la población es responsable de esto", asegura el superintendente de la Policía de Chicago, Eddie Johnson.

(...) Gracias a su nuevo algoritmo computacional, las autoridades de Chicago se han propuesto reducir el número de muertes al más puro estilo de la película de ciencia ficción "Minority Report": vaticinando la comisión de un crimen y arrestando al potencial autor antes de que cometa el delito.

El sistema informático permite crear una lista extrayendo datos de los registros de la policía.

Una vez que la nómina es generada, los agentes pueden visitar a los individuos con los puntajes más altos.

(...) "Es un sistema que observa variables como el historial de arrestos, por ejemplo, si has sido detenido por un delito con armas o si te han disparado. Y usa algunos datos de víctimas criminales", le cuenta a la BBC Jonathan Lewin, director adjunto del Departamento de Policía de Chicago.

Lewin afirma que el algoritmo también tiene en cuenta la edad en el momento del arresto y que utiliza todas esas variables, "nada sobre la raza, el género o la etnia" para determinar el puesto del implicado en la lista.

(...) "No está diseñado para ser una herramienta de castigo, sino para gestionar lo que llamamos un "proceso de notificaciones personalizadas" que consiste, básicamente, en visitar la casa de estos sujetos.

En esa visita, de acuerdo con el comandante, la policía les dice lo siguiente: "Has llamado nuestra atención por estas razones y queremos sacarte del ciclo de violencia; podemos ofrecerte servicios sociales (como formación profesional o servicios de cuidado de niños).

"Pero si no dejas el círculo de violencia y sigues cometiendo crímenes, serás objeto de sanciones penales más severas". (Negritas propias del texto) (BBC Mundo, 11/05/2017)

¹⁴⁶ BBC Mundo. "¿Es posible predecir un delito?: el controversial algoritmo usado por la policía de Chicago, la ciudad con más criminalidad de Estados Unidos". 11/05/2017 <https://www.bbc.com/mundo/noticias-39883234>

Analiza este caso Asuntos Policiales¹⁴⁷ e indica:

(...) este tipo de procedimientos algorítmicos, por más que se perfeccionen, no consiguen resolver un problema central: cómo tratar de desvincular a los barrios más pobres, que están poblados por personas de color negro y por hispanoamericanos, de la delincuencia. Se trata de un efecto perverso, ampliamente acreditado por los informes científicos, que indican que es imposible que esta vía logre diferenciar esos sesgos. Ese es uno de los motivos por los que el Parlamento Europeo ha pedido que se prohíban esas herramientas informáticas en el territorio de la Unión Europea (UE). En un intento de evitar que el uso de esta técnica acabe siendo perjudicial para algunos colectivos, se ha pensado en utilizarla en su lugar para vigilar la propia labor policial. El propósito es conocer los abusos policiales que se pudieran dar en sus intervenciones, asunto muy sensible en esa nación norteamericana. Aunque se olvida que también podría ayudar a las víctimas de esos barrios con asistencia no solo policial sino médica, social o de cualquier otro tipo que fuese necesaria.

(...) En las grandes ciudades norteamericanas, hay barrios que suelen estar asociados a una etnia en particular. La actividad policial en ellos, puede contribuir al rumbo de los informes policiales y causar que se propaguen modelos equivocados y no hay forma de controlar, estadísticamente, esas desviaciones. (Negritas propias del texto) (Asuntos Policiales, 22/01/2023)

La investigación de ProPublica¹⁴⁸ sobre la herramienta COMPAS¹⁴⁹ y la predicción de reincidencia mostró que los scores de riesgo asignados tendían a sobreestimar el riesgo para personas negras y subestimar para personas blancas, incluso cuando se controlaba por variables relevantes. Este caso es paradigmático porque evidencia cómo variables históricas y procesos de criminalización desiguales producen métricas distorsionadas que, al automatizarse, generan injusticias en la justicia penal.

¹⁴⁷ Asuntos Policiales. “*La Policía Predictiva: El Fracaso Experimento de Chicago*.”. 22/01/2023 <https://asuntospoliciales.com/2023/01/22/la-policia-predictiva-el-experimento-de-chicago/>

¹⁴⁸ <https://www.propublica.org/>

¹⁴⁹ COMPAS es un algoritmo utilizado en EE.UU. para analizar el riesgo de reincidencia, entre otras cosas, dicho nombre común es un acrónimo del nombre completo en inglés: Correctional Offender Management Profiling for Alternative Sanctions, traducida como Administración de Perfiles de Criminales para Sanciones Alternativas del Sistema de Prisiones.

Sobre este caso MIT Technology Review¹⁵⁰ indica:

Este proceso es muy imperfecto. Las herramientas utilizan los previos arrestos como un indicador de los delitos, pero, en realidad, hay grandes discrepancias entre ambos, porque la policía tiene un historial desproporcionado de arrestos de minorías raciales y de manipulación de datos. Además, los nuevos arrestos de reincidentes ocurren a menudo por culpa de infracciones técnicas, como no comparecer ante el tribunal, más que por la reincidencia en la actividad delictiva. Aquí lo hemos simplificado bastante para examinar lo que sucedería si los arrestos correspondieran a los delitos reales.

Esta predicción se conoce como la «puntuación de riesgo» del acusado y su objetivo es actuar a modo de recomendación: los acusados de «alto riesgo» deben ser encarcelados para evitar que causen un posible daño a la sociedad; los acusados de «bajo riesgo» deben ser puestos en libertad antes del juicio. (En realidad, los jueces no siempre siguen estas recomendaciones, pero las evaluaciones de riesgo resultan influyentes). (MIT Technology Review, 10/11/2021)

Sobre este algoritmo indica la BBC¹⁵¹ llama la atención sobre esta pregunta “¿Es justo para los criminales que sus sentencias y sus opciones de obtener la libertad condicional en EE.UU. sea determinada por un programa informático de análisis de riesgo basado en un complejo algoritmo?” (Simon, 17/10/2016), continúa resaltando que “En el documental *"Making a murderer"* de Netflix queda claro que, por ejemplo, el sistema judicial del estado de Wisconsin tiene sus problemas a la hora de tomar decisiones justas.” (Simon, 17/10/2016)

Explica que:

"Compas es básicamente un cuestionario que se le da a las personas cuando son arrestadas. Y se les pregunta un montón de cosas que terminan concluyendo si esa persona en el futuro podría cometer un crimen", explicó Angwin a la BBC.

Angwin explicó que las preguntas incluyen aspectos como "tu historia criminal, si algún miembro de tu familia ha cometido un crimen o ha sido arrestado, si vives en un barrio peligroso, si tienes amigos que pertenecen a una pandilla, tu historial laboral y académico".

¹⁵⁰ MIT Technology Review. “Caso práctico: probamos por qué un algoritmo judicial justo es imposible”. 10/11/2021

<https://technologyreview.es/article/caso-practico-probamos-por-que-un-algoritmo-judicial-justo-es-imposible/>

¹⁵¹ Maybin, Simon. ¿Cómo en Estados Unidos las matemáticas te pueden meter en prisión?. BBC News, 17/10/2016. <https://www.bbc.com/mundo/noticias-37679463>

Y agregó: "Y a eso le suman preguntas que podíamos denominar pensamiento criminal, que son algo así como: '¿está de acuerdo o en desacuerdo con esta afirmación: está bien para una persona hambrienta robar comida?'". Cada respuesta recibe un puntaje que varía de 1 a 10.

Entonces aparece un valor promedio de riesgo que decide si alguien puede salir bajo fianza, debe ser enviado a prisión o recibir otro castigo. Cuando la persona ya está encarcelada, el algoritmo determina si merece el beneficio de la libertad condicional.

El problema es que el algoritmo -los números- de donde las respuestas obtienen el puntaje es mantenido en secreto.

"No sabemos cómo se otorga el puntaje porque es secreto. Es un secreto comercial que tiene la empresa que diseñó Compas y no lo comparte", dijo Angwin.

Y para Angwin esto es un gran problema para el acusado si quiere intentar rebatir su puntuación.

"¿Cómo la persona va a decir que es un ocho o un siete, cuando no sabe cómo lo calificaron, cómo llegó a obtener ese puntaje?", explicó Angwin.

(...) La Corte Suprema de Wisconsin también advirtió que Compas podría darle a las minorías étnicas puntajes de riesgo desproporcionado: una sería debilidad en el sistema teniendo en cuenta las actuales tensiones que se vive con la policía por cuestiones vinculadas a la supuesta discriminación y el desigual trato basado en aspectos raciales.

La corte expresó esta preocupación basándose en la investigación de Angwin y sus colegas de ProPublica, que cuestionaban la exactitud de los puntajes de Compas y su aplicación en la práctica.

Dicha investigación analizó los puntajes de 7.000 personas arrestadas en el estado de Florida durante dos años y la conclusión fue impactante.

"Si comparas a una persona negra y una blanca que tienen el mismo historial, la misma edad, el mismo género, el mismo pasado judicial y el mismo 'futuro criminal' (las posibilidades de cometer un crimen, dos crímenes o ninguno), el acusado negro tiene un 45% más de posibilidades de obtener un puntaje de riesgo que un acusado blanco", explicó Angwin. (Negritas propias del texto) (Simon, 17/10/2016)

Como se puede observar, los algoritmos aquí estudiados y otros similares generan, indiscutiblemente un círculo de retroalimentación discriminatorio que intensifica la vigilancia de comunidades ya estigmatizadas o asociadas a un mayor número de denuncias, y no realmente a los casos juzgados o procesados, y tampoco toma en cuenta casos no denunciados, por ejemplo, los delitos de agresión sexual, violencia intrafamiliar y trata de seres humanos, son delitos en donde a pesar de todos los esfuerzos existe una infradenuncia, esto implica que quedan casos sin denunciar y por ende sin investigar, tampoco toman en consideración algoritmos claros y transparentes, se basan en sesgos que deducen tendencias de comportamientos en ciertas zonas de una población y no puede aterrizar a casos particulares de forma justa, aunque así se pretenda.

Ejemplo 2: Sistemas de crédito: Usar como variable el historial de impagos penaliza a personas que han sufrido pobreza estructural o discriminación previa en acceso al empleo.

Tal y como lo denuncian en diversos medios periodísticos o de información¹⁵², los sistemas de IA pueden llegar a desencadenar, al usarse inapropiadamente, “una nueva era de desigualdad entre países pobres y ricos” (ONU, 2/12/2025). Advierte la ONU¹⁵³:

Un manejo inadecuado de la inteligencia artificial amenaza con ampliar las brechas económicas, de capacidades y de gobernanza entre países, revirtiendo décadas de progreso en la reducción de desigualdades.

La inteligencia artificial (IA), si no se gestiona adecuadamente, podría aumentar la desigualdad entre países al ampliar las brechas en el rendimiento económico, las capacidades de las personas y los sistemas de gobernanza, ya que el punto de partida es enormemente desigual.

Esta es la conclusión central del nuevo informe del Programa de las Naciones Unidas para el Desarrollo (PNUD), titulado *La próxima gran divergencia: Por qué la IA puede ampliar la desigualdad entre países*.

El informe destaca que, si bien la IA abre nuevas vías cruciales para el desarrollo, los países inician esta transición desde posiciones extremadamente dispares para captar sus beneficios y gestionar sus riesgos. Sin una acción política contundente, estas brechas podrían crecer, revirtiendo la tendencia de largo plazo hacia la reducción de las desigualdades en el desarrollo. (ONU, 2/12/2025).

¹⁵² Entre otros: <https://insurgenciamagisterial.com/utilizando-la-inteligencia-artificial-gobiernos-y-empresas-condenan-a-la-poblacion-a-la-pobreza/> <https://www.es.amnesty.org/en-que-estamos/blog/historia/articulo/pueden-la-inteligencia-artificial-y-el-uso-de-algoritmos-perpetuar-la-desigualdad-tres-casos-en-los-que-la-realidad-supera-la-ficcion/> y <https://news.un.org/es/story/2025/12/1540837>

¹⁵³ ONU. “*La IA puede desencadenar una nueva era de desigualdad entre países pobres y ricos*”. 02/12/2025 <https://news.un.org/es/story/2025/12/1540837>

Karen Hao¹⁵⁴ investigadora del MIT, concluye en su artículo que:

Las calificaciones crediticias se han utilizado durante décadas para evaluar la solvencia crediticia del consumidor, pero en la actualidad su alcance es mucho mayor, ya que funcionan mediante algoritmos: no solo cuentan con muchos más datos, tanto en volumen como en tipo, sino que influyen cada vez más en la posibilidad de comprar un coche, alquilar un piso o conseguir un trabajo a tiempo completo. Su poder integral significa que, si la calificación se reduce, podría resultar casi imposible recuperarla.

Y lo que es peor, los algoritmos pertenecen a empresas privadas que no divulgan cómo toman sus decisiones. Las víctimas pueden acabar en una espiral descendente que a veces termina en la pérdida de la vivienda o en el regreso a su abusador.

Los algoritmos de calificación crediticia no son los únicos que interfieren en el bienestar económico de las personas y en su acceso a servicios básicos. En EE. UU., los algoritmos ya son responsables de decidir qué niños pueden entrar en el programa de familias de acogida, qué pacientes reciben atención médica y qué familias tienen acceso a una vivienda estable. Los que tenemos medios podemos pasar toda la vida sin darnos cuenta de nada. Pero para las personas de bajos ingresos, el rápido crecimiento y la adopción de sistemas automatizados de toma de decisiones ha creado una red oculta de trampas entrelazadas. (Hao, Karen, 13/02/2021)

Para la profesora de Derecho en la Universidad de Baltimore (EE.UU.) Michelle Gilman:

Básicamente, todos los abogados especializados en derecho civil estamos empezando a lidiar con estos problemas, porque, de una forma u otra, todos nuestros clientes están siendo afectados por estos algoritmos. Tenemos que despertar y formarnos. Si queremos ser realmente buenos abogados holísticos, debemos ser conscientes de eso. ¿Tendré que interrogar un algoritmo?

(...) Cada vez más, la batalla por los derechos de un cliente implica algún tipo de algoritmo.

(...) Esto les pasa a nuestros clientes en todos los ámbitos. Están enredados en tantos algoritmos diferentes que los excluyen de los servicios básicos. Y es posible que ni si quiera den cuenta, porque muchos de estos sistemas son invisibles. (Hao, Karen, 13/02/2021)

¹⁵⁴ Hao, Karen. “*Utilizando la Inteligencia Artificial gobiernos y empresas condenan a la población a la pobreza*”. Insurgencia Magisterial. 13/02/2021 <https://insurgenciamagisterial.com/utilizando-la-inteligencia-artificial-gobiernos-y-empresas-condenan-a-la-poblacion-a-la-pobreza/>

2.1.3. Sesgo de etiquetado (*Label Bias*): Se produce cuando los seres humanos que etiquetan los datos introducen sus prejuicios, estereotipos o interpretaciones subjetivas. Entonces, debido a que los modelos aprenden de tales etiquetas, transfieren los sesgos humanos al sistema o al modelo IA.

Ejemplo 1: Etiquetado de emociones: Datasets donde personas caucásicas etiquetan expresiones de enfado en imágenes de personas afrodescendientes suelen inflar la categoría agresivo o peligroso. Resultado expuesto por un estudio de la Universidad de Princeton donde demuestran un sobre-etiquetado de emociones negativas en hombres negros. Situación confirmada por el estudio internacional en que participa la Universidad de Granada (España)¹⁵⁵ en donde se “que demuestra que a las personas blancas, al no mirar a los ojos de las personas de piel negra por considerarlas de un grupo distinto al suyo, les resulta difícil distinguir entre sonrisas genuinas y falsas en las caras negras”, lo que no ocurre a la inversa, con lo cual si estas personas etiquetan expresiones faciales para entrenar a un Sistema de IA, transferirán sus prejuicios o sesgos al modelo computacional.

Los Investigadores de la Universidad de Granada señalan también que “La identificación precisa de la emoción es importante para la interacción social en general, pero especialmente en entornos interraciales, que son propensos a malinterpretaciones y malentendidos” y expone que:

(...) Investigaciones anteriores ya habían demostrado que percibimos de manera más pormenorizada y precisa a los miembros de los grupos a los que pertenecemos (endogrupos) en comparación con los miembros de los grupos a los que no nos consideramos pertenecientes (exogrupos). De hecho, varios trabajos han sugerido que los blancos en los Estados Unidos tienden a percibir las caras negras como más enfadadas que las caras blancas. Ahora, los/as investigadores/as de este estudio querían ver si se producen sesgos similares en la identificación de emociones para emociones positivas, como la felicidad o la alegría.

Los/as científicos/as llevaron a cabo una serie de seis experimentos con más de 425 participantes. En algunos de estos experimentos, los/las participantes eran todos/as blancos/as; en otros, todos eran negros/as; y en algunos, participaron una mezcla de ellos que también incluyó a minorías no negras. En todos ellos a los/as participantes se les mostraba rostros de personas blancas o negras sonrientes y se les pedía que calificaran el nivel de felicidad que percibían en ellas. Algunas caras mostraban una sonrisa genuina, mientras que otras eran sonrisas forzadas o falsas.

Mediante un equipo de «eye tracking», demostramos que prestar atención a los ojos de las personas también nos facilita la detección de características

¹⁵⁵ UGRDivulga. “Un estudio revela que los blancos perciben peor las emociones en las caras de los negros que a la inversa”. 23/04/2019. <https://canal.ugr.es/noticia/un-estudio-revela-blancos-perciben-peor-las-emociones-en-las-caras-de-los-negros/>

importantes en la percepción y en la formación de impresiones sobre las mismas, como por ejemplo la distinción entre emociones genuinas o «falsas» e impostadas, explica Rosa Rodríguez Bailón, catedrática del departamento de Psicología Social de la UGR y una de las autoras de este estudio. (UGRDivulga, 23/04/2019)

Ejemplo 2: Moderación de contenido: Cuando los seres humanos moderadores etiquetan como violento algún contenido cultural o religioso específico, usualmente por desconocimiento, el algoritmo del modelo IA luego censura de forma discriminatoria al grupo etiquetado.

2.1.4. Sesgo por objetivo (*Objective Bias*): Ocurre cuando el objetivo matemático que se optimiza, esto es, la función de pérdida (*loss function*), favorece resultados que aumentan desigualdades, porque la métrica elegida no captura valores éticos o sociales. En este caso, el sesgo no está en los datos, sino en lo que el sistema considera éxito.

Ejemplo 1: Sistemas de contratación: Si el objetivo de un sistema de contratación es predecir quién permanecerá más tiempo en la empresa, el algoritmo favorecerá perfiles mayoritarios porque históricamente esos son los que permanecieron más tiempo en la empresa, esto es, no tomará en cuenta las situaciones actuales, por ello, discriminará a mujeres jóvenes, al hacer una asociación a la licencia de maternidad y las responsabilidades posteriores o a personas migrantes, por sus historias laborales discontinuas.

Ejemplo 2: Plataformas de contenido: Si la función objetivo del modelo IA es maximizar clics, el algoritmo prioriza contenido emocional o polarizante.

Esto favorece publicaciones amarillistas o discursos incendiarios o de odio, afectando derechos de minorías.

2.1.5. Sesgo Emergente en Sistemas Complejos (*Emergent Bias*) (Comportamiento no previsto): Este sesgo surge en modelos IA complejos, especialmente de propósito general (GPT, Gemini, Llama, etc.), cuando el sistema desarrolla comportamientos inesperados, no previstos durante el entrenamiento. No es un sesgo que estuviera en los datos o en el código, sino que emerge por la dinámica interna del modelo y sus interacciones con el entorno.

Ejemplo 1: Modelos de lenguaje que refuerzan estereotipos: Un modelo puede no estar diseñado para discriminar, pero al combinar millones de patrones lingüísticos puede: asociar profesiones técnicas a hombres, sexualizar más a mujeres, reproducir patrones raciales nocivos en traducciones (doctor → él, enfermera → ella).

Estos efectos no fueron programados, sino que emergen del comportamiento global del modelo.

Ejemplo 2: Sesgos en sistemas de recomendación: Un algoritmo de YouTube o TikTok puede amplificar contenido: extremista, misógino, racista, pseudocientífico, sin que los programadores lo hayan previsto. Esto ocurre por la interacción entre: el algoritmo, las preferencias del usuario, la dinámica de red, y la búsqueda de maximización de retención.

Cuestión especialmente relevante en modelos de propósito general cuyos efectos sólo se detectan en despliegue.

En la siguiente tabla se esquematiza lo analizado sobre los tipos de sesgo con fines de una mejor comprensión del lector:

Tipo de sesgo	¿De dónde surge?	Ejemplo típico
Representación	Datos poco diversos	Reconocimiento facial en mujeres con piel oscura
Medida	Métricas contaminadas por injusticias	Arrestos previos como indicador de riesgo
Etiquetado	Prejuicios de personas que etiquetan	Emociones mal etiquetadas según raza
Objetivo	Función de pérdida mal diseñada	Maximizar clics → contenidos discriminatorios
Emergente	Conducta del sistema en despliegue	Modelos de lenguaje que generan estereotipos

Tabla 1. Tipos de Sesgo resumen. Autor: Angélica María Castro Acosta

Un meta-análisis de todo el acápite de sesgos algorítmicos, reforzados por la casuística real aportada como ejemplo y contexto, y soportada en la información contenida en numerosos estudios académicos y reportes institucionales, se observa la presencia de sesgos algorítmicos en sistemas de IA de toda índole, en el sector sanitario (riesgo de diagnóstico), en el sector crédito y seguros (score crediticio), en el sector justicia (cálculo de reincidencia, sesgo de raza y género) y en el publicidad (publicidad dirigida), entre muchos otros. Toda la literatura científica analizada, suelen coincidir en que los sesgos algorítmicos se encuentran potencialmente presentes en toda la vida del Sistema de IA, emergiendo tanto por problemas de datos como por falta de diseño orientado a la equidad.

3. DERECHOS HUMANOS EN LA ERA DIGITAL

En este panorama es indiscutible la gran posibilidad que existe de que los Sistemas de IA con sesgos algorítmicos pongan en peligro o vulneren Derechos Humanos,

por ello, se estudiaran los Derechos Humanos, pretendiendo vislumbrar aquellos que puedan estar en esta situación y proponer una forma de prevenir su vulneración.

El mundo actual en medio de la era de la globalización y la revolución metamórfica de las Tecnologías de la Información y las Comunicaciones (en adelante TIC), dentro de las cuales se eleva sobre las otras, por su gran campo de utilización, los sistemas computacionales de Inteligencia Artificial o la Inteligencia Artificial de forma sintética (en adelante IA), presenta una serie de problemáticas de carácter global, pero cuyo desarrollo legal y jurisprudencial se está desarrollando a nivel local, pero los demás Estados siguen de cerca tales soluciones legales y jurisprudenciales buscando luces que guíen su camino, las cuales a veces sólo son fuegos fatuos, ligeros y nada perdurables, en razón a que como la casuística se está desarrollando y cambiando día a día, algunas decisiones jurisprudenciales ponderan situaciones aún no consolidadas, como cuando de IA se trata.

Pero estos constantes avances están generando que se olvide poner el foco en los seres humanos, y habría que preguntarse ¿Cuál será el papel que esperamos que desempeñen los Derechos Humanos en esta carrera desordenada e inconexa de avances en una multiplicidad de escenarios de los sistemas de IA?.

Tal cuestión ha motivado una inspiración constante en estudiar y analizar el estado académico, legal y jurisprudencial de la IA en España y en Europa, y en esta oportunidad ha conducido al análisis del riesgo que representan los sistemas de IA, mal gobernados, a los Derechos Humanos.

La Unión Europea como institución se inspira y pregonas como valor máximo la protección de los Derechos Humanos, y esta inspiración permea a sus instituciones, en tal línea el Consejo de Europa en 2019 luego de la Conferencia “Governing the Game Changer – Impacts of artificial intelligence development on human rights, democracy and the state of law” llevada a cabo en Helsinki, en su conclusión número 12 pregonas:

Todas las partes interesadas pertinentes deberían explorar e investigar a fondo los impactos de la IA en los derechos humanos, la democracia y el Estado de derecho. El Consejo de Europa, con un enfoque multisectorial, debería seguir elaborando recomendaciones, directrices y códigos de conducta sectoriales para promover los derechos humanos y la viabilidad de las instituciones y los procesos democráticos. Debería supervisar el impacto de la IA en los cimientos colectivos de las sociedades democráticas, seguir identificando posibles lagunas en la aplicabilidad y el cumplimiento de los marcos regulatorios existentes y, sobre esta base, evaluar la necesidad de adoptar nuevas medidas para garantizar que el desarrollo, el diseño y la implementación de la IA respeten los derechos humanos. (Consejo de Europa, 2019, Conclusión 12)

Por su parte, la conclusión número 11 declara que para garantizar los Derechos Humanos existen diversos instrumentos internacionales, los cuales no han perdido aplicabilidad a pesar de que no hagan referencia expresa a la IA, y deben usarse para garantizar que el progreso tecnológico respete los principios de los Derechos Humanos, apoyándose en directrices éticas y autorregulaciones:

Los instrumentos internacionales emblemáticos existentes, como la Declaración Universal de Derechos Humanos y el Convenio Europeo para la Protección de los Derechos Humanos y de las Libertades Fundamentales, son aplicables independientemente de los cambios contextuales provocados por la IA y deben cumplirse para garantizar que el progreso tecnológico se produzca de conformidad con los principios de los derechos humanos, la democracia y el Estado de derecho. Las directrices éticas y la autorregulación constituyen herramientas adicionales para promover estos valores. (Consejo de Europa, 2019, Conclusión 11)

De forma introductoria, hay que indicar que, la Declaración Universal de Derechos Humanos y los tratados internacionales consagran principios como la igualdad ante la ley y la prohibición de discriminación. Sin embargo, la introducción y aplicación actual de Sistemas de IA en casi todos los ámbitos importantes de la vida humana, ejerce presión sobre las conquistas jurídicas internacionales de declaración y protección de los Derechos Humanos, e introduce nuevos desafíos jurídicos, que es importante, primero detectar, para luego de diagnosticarlos establecer una forma de prevención y tratamiento.

3.1. Marco normativo europeo y español: ante estos posibles riesgos, múltiples organismos internacionales han impulsado marcos regulatorios específicos que, desde la determinación de requisitos o prohibiciones, diseñadas con una óptica ético-legal y con una clara visión humano-céntrica pretenden prevenir efectos indeseados en las vidas humanas. Por ello se proseguirá este análisis a tales instrumentos:

3.1.1. Instrumentos y principios internacionales y multilaterales: dentro de los Instrumentos contentivos de principios internacionales y multilaterales se encuentran:

-UNESCO – *Recommendation on the Ethics of Artificial Intelligence* (2021): establece principios globales como justicia, explicabilidad y responsabilidad, aplicables a todas las fases del desarrollo tecnológico, brindando un marco global de principios centrados en los Derechos Humanos, dignidad humana, equidad, transparencia y rendición de cuentas, así como, unas recomendaciones aplicables a Estados y actores privados sobre gobernanza, evaluación de impacto y educación.

La UNESCO reconoce:

(...) las repercusiones positivas y negativas profundas y dinámicas de la inteligencia artificial (IA) en las sociedades, el medio ambiente, los ecosistemas y las vidas humanas, en particular en la mente humana, debido en parte a las nuevas formas en que su utilización influye en el pensamiento, las interacciones y la adopción de decisiones de los seres humanos y afecta a la educación, las ciencias sociales y humanas, las ciencias exactas y naturales, la cultura y la comunicación y la información (UNESCO, 2021, Preámbulo)

Afirma además que esta Recomendación es un instrumento normativo con un enfoque mundial,

(...) basado en el derecho internacional y centrado en la dignidad humana y los derechos humanos, así como en la igualdad de género, la justicia social y económica y el desarrollo, el bienestar físico y mental, la diversidad, la interconexión, la inclusión y la protección del medio ambiente y de los ecosistemas, puede dar una orientación responsable a las tecnologías de la IA (UNESCO, 2021, Preámbulo)

En esta Recomendación la UNESCO luego de reconocer la gran utilidad para la humanidad que representan los Sistemas de IA plantea preocupaciones éticas

(...) en relación con los sesgos que pueden incorporar y exacerbar, lo que puede llegar a provocar discriminación, desigualdad, brechas digitales y exclusión y suponer una amenaza para la diversidad cultural, social y biológica, así como generar divisiones sociales o económicas; la necesidad de transparencia e inteligibilidad del funcionamiento de los algoritmos y los datos con los que han sido entrenados; y su posible impacto en, entre otros, la dignidad humana, los derechos humanos y las libertades fundamentales, la igualdad de género, la democracia, los procesos sociales, económicos, políticos y culturales, las prácticas científicas y de ingeniería, el bienestar animal y el medio ambiente y los ecosistemas (UNESCO, 2021, Preámbulo)

-OECD AI Principles (2019): La OCDE identifica cinco principios basados en valores para la administración responsable de los Sietmas de IA, con base en los cuales determina cinco recomendaciones, basados en el trabajo desarrollado por el Grupo de trabajo integrado por 50 expertos en IA.

Los cinco principios se pueden resumir, de la mano de la infografía del Gobierno español dispuesta en datos.gob.es así:

1. La IA debe beneficiar a las personas y al planeta impulsando el crecimiento inclusivo, el desarrollo sostenible y el bienestar.
2. Los sistemas de IA deben diseñarse respetando el estado de derecho, los derechos humanos, los valores democráticos y la diversidad, y deben incluir los mecanismos apropiados - como por ejemplo, permitir la

intervención humana cuando sea necesario- para garantizar una sociedad justa y equitativa.

3. Debe existir una transparencia y divulgación responsable en torno a los sistemas de IA, para garantizar que las personas entiendan sus resultados y puedan desafiarlos.

4. Los sistemas de IA deben funcionar de manera sólida y segura a lo largo de su ciclo de vida y los riesgos potenciales deben evaluarse y gestionarse continuamente.

5. Las organizaciones e individuos que desarrollan, despliegan u operan sistemas de IA deben ser responsables de su correcto funcionamiento, en base a los principios anteriormente descritos. (Gobierno de España, 2024)

Continúa el Gobierno español explicando que “Tomando como referencia estos principios, la OCDE también proporciona cinco recomendaciones a los gobiernos” y los sintetiza así:

1. Facilitar la inversión pública y privada en investigación y desarrollo, con el objetivo de estimular la innovación en Inteligencia Artificial de manera segura y fiable.

2. Fomentar los ecosistemas de IA accesibles con infraestructura y tecnologías digitales, y mecanismos para compartir datos y conocimientos.

3. Asegurar un marco de políticas que abra el camino para el despliegue de sistemas de IA fiable.

4. Capacitar a las personas con habilidades necesarias para la IA y apoyar a los trabajadores para una transición justa.

5. Cooperar a través de las fronteras y los sectores para avanzar en la administración responsable de IA fiable. (Gobierno de España, 2024)

Se observa que estos principios y recomendaciones se centran en la persona, la transparencia, la robustez y la responsabilidad, instando a los Gobiernos a evaluar los posibles riesgos y a implementar las salvaguardias requeridas para evitar la vulneración de Derechos.

3.1.2. Unión Europea:

- RGPD (Reglamento (UE) 2016/679): Tal y como explica Castro-Acosta y Andrés Molina (2025) “El uso de la IA en los sitios web, en las apps y en los aparatos informáticos personales ya es una realidad, a la cual el Derecho en general y especialmente el constitucional debe dar respuesta”, por ello la Unión Europea ha regulado la protección de esos datos personales en el Reglamento en mención, partiendo de la Consideración de que:

4. El tratamiento de datos personales debe estar concebido para servir a la humanidad. El derecho a la protección de los datos personales no es un derecho absoluto sino

que debe considerarse en relación con su función en la sociedad y mantener el equilibrio con otros derechos fundamentales, con arreglo al principio de proporcionalidad. El presente Reglamento respeta todos los derechos fundamentales y observa las libertades y los principios reconocidos en la Carta conforme se consagran en los Tratados, en particular el respeto de la vida privada y familiar, del domicilio y de las comunicaciones, la protección de los datos de carácter personal, la libertad de pensamiento, de conciencia y de religión, la libertad de expresión y de información, la libertad de empresa, el derecho a la tutela judicial efectiva y a un juicio justo, y la diversidad cultural, religiosa y lingüística. (Reglamento (UE) 2016/679, Considerando 4)

En razón a tal consideración, establece obligaciones de protección de datos, los cuales deben ser “tratados de manera lícita, leal y transparente en relación con el interesado” (Reglamento (UE) 2016/679, Art. 1.a), establece derechos de los interesados y límites a la toma de decisiones automatizada ex. art. 22, así como obligaciones de transparencia ex. arts. 12 al 15 y la evaluación del impacto si el tratamiento supone riesgos elevados para los derechos y libertades de los titulares de esos datos personales.

- AI Act (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) nº 300/2008, (UE) nº 167/2013, (UE) nº 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)).

Este reciente e importantísimo Reglamento, en su considerando 6 pregonaba que:

Dadas las importantes repercusiones que la IA puede tener en la sociedad y la necesidad de generar confianza, es fundamental que la IA y su marco reglamentario se desarrollen de conformidad con los valores de la Unión consagrados en el artículo 2 del Tratado de la Unión Europea (TUE), los derechos y libertades fundamentales consagrados en los Tratados y, de conformidad con el artículo 6 del TUE, la Carta. Como requisito previo, la IA debe ser una tecnología centrada en el ser humano. Además, debe ser una herramienta para las personas y tener por objetivo último aumentar el bienestar humano. (AI Act, 2024)

Este instrumento representa el primer marco vinculante a nivel mundial que regula la IA con un enfoque basado en el riesgo clasifica a los sistemas de IA por nivel de riesgo, clasificándolos en: inaceptable, alto, limitado, mínimo y estableciendo obligaciones para garantizar la transparencia y la supervisión humana de los Sistemas de IA, así como obligaciones muy estrictas para sistemas de alto riesgo, tales como: evaluación de riesgos, documentación técnica, gestión de datos,

transparencia, supervisión humana, además de medidas específicas sobre modelos de propósito general.

- Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho (2024):

Este nuevo Convenio, que aún no es una realidad jurídica porque aunque el tratado fue adoptado por el Comité de Ministros del Consejo de Europa en mayo de 2024 y se abrió a la firma en septiembre de 2024, su entrada en vigor depende de un proceso de ratificación, el cual implica que se cumpla lo que el propio tratado estipula y es que entrará en vigor el primer día del mes siguiente a la expiración de un período de tres meses después de que cinco signatarios, incluyendo al menos tres Estados miembros del Consejo de Europa, lo hayan ratificado, pero al ser un gran avance, no puede dejar de mencionarse.

La Propuesta de Decisión del Consejo COM(2024) 264 Final, de 26 de junio, relativa a la firma, en nombre de la Unión Europea, del Convenio Marco del Consejo de Europa sobre Inteligencia Artificial, Derechos humanos, Democracia y Estado de derecho, en el acápite 3 indica:

El Convenio tiene por objeto abordar los posibles riesgos y perjuicios para los derechos humanos garantizando que las actividades durante el ciclo de vida de los sistemas de IA se ajusten a los principios de respeto de los derechos humanos, la democracia y el Estado de Derecho, reconociendo al mismo tiempo el potencial de la IA para proteger y facilitar el ejercicio de esos derechos en el entorno digital y mejorar el bienestar social y medioambiental y el progreso tecnológico.

Los principios y obligaciones concretos previstos en el Convenio tienen por objeto proteger y respetar los derechos humanos, consagrados en múltiples instrumentos internacionales y regionales¹⁵⁶, aplicables a las Partes, incluida la Carta de los Derechos Fundamentales de la Unión y los instrumentos internacionales sobre derechos humanos que la Unión ha celebrado.

Así pues, el Convenio establece un nivel mínimo común para la protección de los derechos humanos en el contexto de la IA, salvaguardando al mismo tiempo la

¹⁵⁶ Como la Declaración Universal de Derechos Humanos de 1948; el Convenio para la Protección de los Derechos Humanos y de las Libertades Fundamentales de 1950 (ETS n.º 5); el Pacto Internacional de Derechos Civiles y Políticos de 1966; el Pacto Internacional de Derechos Económicos, Sociales y Culturales de 1961; la Carta Social Europea de 1966 (ETS n.º 35), así como sus respectivos protocolos; y la Carta Social Europea de 1996 (revisada) (ETS n.º 163), la Convención de las Naciones Unidas sobre los Derechos del Niño de 1989 y la Convención de las Naciones Unidas sobre los Derechos de las Personas con Discapacidad de 2006.

protección existente de los derechos humanos y permitiendo a las Partes ofrecer una protección más amplia con salvaguardias más estrictas. (Consejo de Europa, 2024, 3)

3.1.3. España: LOPDGDD y medidas nacionales

- Ley Orgánica 3/2018 (LOPDGDD): En España se adapta el RGPD Europeo al ordenamiento español y se regula, entre otras materias, derechos digitales y el régimen de la Agencia Española de Protección de Datos, esta norma combina la protección de datos con las garantías relativas a automatización y tratamiento de datos públicos y privados.
- Desarrollos estratégicos y guías: España, como otros Estados miembros, ha ido publicando estrategias y guías para la adopción responsable de IA (documentos ministeriales y de la AEPD), en línea con la ética de la IA europea, que deben ser tenidos en cuenta a la hora de pensar en la protección de Derechos.

3.2. Impacto sobre derechos humanos y principios afectados

Luego de ese devenir enunciativo (efectuado sin pretensiones de erigirse como una lista taxativa) de instrumentos jurídicos en donde se integra la protección de los Derechos humanos en relación con los Sistemas de IA: bien de forma directa, regulaciones en donde se habla de IA, o bien de forma indirecta, en las regulaciones que se refieren a la protección de datos personales, porque en última, los Sistemas de IA se nutren de datos, muchos de los cuales pueden ser de carácter personal, se pasará a analizar algunos Derechos que se consideran principalmente en riesgo en el ciclo de vida de los Sistemas de IA, especialmente, como se ha indicado con los sesgos algorítmicos.

3.2.1. Derecho a la igualdad y no discriminación

Los sesgos algorítmicos pueden traducirse en discriminación indirecta o directa en la práctica, cuando, como se ha analizado antes, se niega un crédito, se establecen sanciones más severas a reos, o se reproducen estereotipos que afectan la dignidad de los seres humanos y el acceso a oportunidades.

Al analizar marco europeo y español se encuentran prohibiciones explícitas de cualquier tipo de discriminación y se exige medidas positivas para evitar impactos desproporcionados en los seres humano; sin embargo, la tecnologización de decisiones exige herramientas específicas de cumplimiento, que se están integrando en las nuevas regulaciones.

3.2.2. Derecho a la privacidad y protección de datos personales

El procesamiento masivo de datos personales y la inferencia de características sensibles (por ejemplo, etnia o religión por proxy) implican riesgos para los datos personales de los seres humanos. Por ello el RGPD introduce límites como: el principio de minimización, el principio de transparencia, los derechos de acceso, rectificación y supresión, y se regulan las decisiones automatizadas que produzcan efectos jurídicos o "similarmente significativos" (RGPD, Art.22). Los cuales servirán para mitigar o sancionar el sesgo discriminatorio de los Sistemas de IA.

3.2.3. Debido proceso y derecho a una decisión motivada

Cuando decisiones importantes se basan en modelos opacos, lo que se conoce como *Cajas Negras*, se compromete el derecho que tienen los seres humanos a conocer las razones de una decisión, especialmente en contextos administrativos, penales o sancionatorios. Por ello, la jurisprudencia Europea y española ha tendido a proteger la exigencia de motivación y proporcionalidad, y recientemente los tribunales han avanzado hacia la idea de transparencia algorítmica en el sector público.

La Sentencia del Tribunal de Justicia de la Unión Europea (TJUE) de febrero de 2025 (Asunto C-203/22, *Dun & Bradstreet*), es un fallo crucial del TJUE que confirmó la existencia de un "derecho de explicación" y dictaminó que los tribunales y las autoridades nacionales pueden obtener acceso a información protegida por secretos comerciales cuando sea necesario para equilibrar estos intereses con los derechos del interesado (ciudadano). Prioriza así los derechos fundamentales sobre los secretos comerciales en el contexto de la transparencia algorítmica.

En España la Sentencia BOSCO, Sentencia del Tribunal Supremo n.º1119/2025, de 11 de septiembre, ha consolidado la transparencia algorítmica como un derecho fundamental y un instrumento esencial para la rendición de cuentas y la buena administración. Estos fallos dictaminan que el código fuente y la documentación de las aplicaciones informáticas utilizadas por la Administración Pública (como la aplicación "Bosco" para el sorteo de tribunales de oposiciones) constituyen información pública accesible para los ciudadanos. El Tribunal rechazó los argumentos de la administración sobre seguridad informática o propiedad intelectual como límites absolutos al derecho de acceso a la información pública.

3.3. Previniendo los Sesgos Algorítmicos

Partiendo de la aceptación de esta realidad aneja a todo Sistema de IA, esto es, que en todo Sistema de IA se puede presentar en alguna parte de la vida del mismo un sesgo algorítmico, en tal razón hay que indicar que la prevención de estos posibles y esperados sesgos no parte del establecimiento de una única medida, sino que para su

efectividad se debe establecer un conjunto de prácticas técnicas, organizativas, legales y éticas, que combinen buenas prácticas de datos, técnicas de *fairness*, auditorías continuas, supervisión humana y cumplimiento de los marcos legales internos Españoles y de la UE, y que sean previstas de manera integral durante todo el ciclo de vida del sistema.

Debido a que, como se ha indicado, prevenir los sesgos algorítmicos en los sistemas de IA es un desafío crítico que requiere un enfoque holístico en todas las etapas del ciclo de vida del desarrollo del Sistema de IA, desde la conceptualización del mismo hasta el despliegue y monitoreo, esta investigación ha concluido en la estructuración de una estrategia integrada, la cual se divide en una fase 0 o previa, tres fases correlacionadas y una fase transversal que articulará todo el modelo de prevención y actuación ante la detección de sesgos algorítmicos en un Sistema de IA.

Para esto, se invita a pensar en los datos como la educación que recibe el Sistema de IA, y basándonos en ello, se plantea la siguiente analogía: si el colegio es malo, lo que aprenda será mediocre y con prejuicios, para lo evitar este resultado indeseado se propone el siguiente protocolo, que consta de fases:

0. Fase Previa: En esta fase se considera importante que, desde la misma concepción y luego diseño de la idea de un sistema de IA, los diseñadores y programadores sean plenamente conscientes de su inestimable papel para la protección de los Derechos de los Seres Humanos, por ello, tomando en consideración que durante toda la vida de sus diseños deben proteger los Derechos humanos, y con la experiencia actual de situaciones vulneradoras de tales Derechos, los diseñadores y programadores podrán precaver situaciones que sean el germen de un sesgo, o evitar incluir situaciones con potencialidad de vulneración de tales Derechos, Y por supuesto, siendo conscientes, de que la inclusión voluntaria de esos sesgos algorítmicos, sobre todo aquellos con contenido discriminatorio, puede ser susceptible de sanciones legales importantes.

Esta fase previa o preliminar, debe ser tenida en cuenta en forma general por aquellos con poder en los proyectos de creación de sistemas de IA, por eso la misma incluirá necesariamente los siguientes pasos:

0.1. Establecimiento de la Idea, con la idea concebida se debe poder establecer unos riesgos iniciales estimados y con ello ser capaces de prevenirlos.

0.2. Sensibilización a todos los miembros de un proyecto de creación de un sistema de IA, incluyendo necesariamente tanto a diseñadores y programadores, como a todo el personal con nivel decisorio, para que, comprendiendo que es un sesgo algorítmico y sus previsibles consecuencias, lo eviten a toda costa, aplicando criterios éticos desde esta fase.

1. Fase de Prevención y Control de los Datos (Raíz del Problema) (Antes del Entrenamiento)

Esta fase inicial de prevención ocurre antes del entrenamiento del Modelo, ya cuando está diseñado el mismo, los sesgos suelen originarse en los datos que nutren el modelo y que luego lo van a entrenar, ya que pueden reflejar desigualdades históricas o prejuicios sociales, de forma voluntaria o involuntaria.

1.1. Auditoría y Desesgo de Datos (Data Debiasing): Para lograr el desesgo se propone una auditoría de datos, donde luego de pasar los pasos siguientes, se pueda afirmar que se ha hecho lo posible para evitar el sesgo de un Modelo de IA.

- Identificación: Este primer paso, consistiría en realizar un análisis exhaustivo para identificar la distribución de atributos sensibles (género, etnia, edad, etc.) y buscar patrones de subrepresentación o sobrerrepresentación.
- Balanceo (Data Balancing): Este paso requiere aplicar técnicas para igualar las proporciones de las clases o grupos minoritarios en el conjunto de entrenamiento.
- Aumento de Datos (Data Augmentation): En este paso se generan datos sintéticos o se aumentan los conjuntos existentes para compensar la falta de ejemplos de grupos minoritarios.
- Anotación Justa: Este último paso pretende garantizar que los anotadores de datos estén capacitados para evitar introducir sus propios sesgos al etiquetar la información.

1.2. Recolección Contextualizada: Esta fase tiene como fin, el recolectar datos que reflejen la diversidad real del dominio de aplicación y garantizar que la información recolectada sea relevante y representativa.

2. Fase de Modelado (Diseño del Algoritmo)

Debido a que el sesgo puede introducirse o amplificarse también en la forma en que el modelo aprende o en las métricas que se utilizan, la estrategia propuesta tiene que tener en cuenta esta realidad y se propone:

2.1. Diseño de Métricas de Equidad: No basta con la exactitud global (*accuracy*) del modelo, es crucial utilizar métricas de equidad que midan el rendimiento del modelo en grupos específicos (p. ej., la paridad de impacto, igualdad de oportunidades, o igualdad de tasas de error para diferentes subgrupos).

2.2. Paridad de Impacto (Demographic Parity): Pretende asegurar que la tasa de selección o resultado positivo sea similar para todos los grupos.

2.3. Técnicas de Desesgo durante el Entrenamiento (In-Processing): Dentro de estas técnicas se pueden integrar:

- Utilizar métodos de aprendizaje consciente de la equidad (*fairness-aware learning*) que modifican la función de pérdida del modelo para penalizar las desviaciones en la equidad mientras optimizan la precisión.
- Aprendizaje Adversario (Adversarial Debiasing): Entrenar un modelo de IA para ser preciso en su tarea principal, mientras se entrena simultáneamente a un "adversario" para que no pueda predecir el atributo sensible (ej., género) basándose en las predicciones del modelo principal.

2.4. Modelos Explicables (XAI - Explainable AI): También se deben implementar técnicas de interpretabilidad que permitan a los desarrolladores y usuarios entender qué variables o características influyen en las decisiones del modelo, facilitando la detección de sesgos ocultos.

3. Fase de Despliegue y Monitoreo (El Entorno Real): Una vez en producción, el sesgo puede reaparecer o evolucionar con el uso real (*drift*), para prevenir esto, es importante aplicar estas estrategias:

3.1. Pruebas Rigurosas (Testeo de Sesgo): Realizar pruebas de caja negra y caja blanca en entornos de simulación y con datos reales (sometiendo el modelo a estrés) para detectar sesgos antes de su lanzamiento.

3.2. Monitoreo Continuo: Implementar *dashboards* de monitoreo en tiempo real para rastrear las métricas de equidad a lo largo del tiempo.

Se requieren establecer umbrales de alerta que se activen si la disparidad de resultados entre los grupos supera un límite predefinido, indicando un posible sesgo en evolución (*model drift*).

3.3. Mecanismos de Recurso y Transparencia: Establecer un proceso claro para que los usuarios o las personas afectadas puedan impugnar las decisiones algorítmicas.

Con lo cual se proporcione a los usuarios información accesible y comprensible sobre cómo se tomó una decisión, fomentando la transparencia de todo el Sistema.

4. Fase Transversal: Gobernanza Ética y Regulatoria (Legal)

Siendo todas las fases del protocolo establecidas por su importancia directa en la prevención y mitigación de los efectos de los sesgos algoritmos ubicando al ser

humano en el centro protector del análisis, esta investigación concluye que la prevención efectiva requiere una estructura organizacional que respalde la equidad y la ética en sus decisiones, para ello, se proponen las siguientes actividades:

4.1. Guías Éticas Internas – Estructuración de Principios éticos: En los equipos de trabajo que requieren todo Sistema de IA, se deben desarrollar códigos de conducta ética y guías de buenas prácticas específicas para el desarrollo de IA, que prioricen principios éticos importantes, como por ejemplo: Equidad, Transparencia, Responsabilidad, Justicia, No discriminación y Proporcionalidad

4.2. Cumplimiento regulatorio: En todo sistema de IA, es de suma importancia que se cumplan con todos los requisitos legales establecidos para ellos, en especial aquellas que regulan los tipos de riesgo, como aquellos de riesgo alto que exigen evaluaciones de impacto de derechos fundamentales y sistemas de gestión de riesgos, incluyendo el sesgo, según los mandatos del Reglamento europeo de IA (AI Act) y el Reglamento General de Protección de Datos (RGPD)

4.3. Equipos Interdisciplinarios: En todo Sistema de IA, es necesario involucrar a expertos en ciencias sociales, ética y Derecho, que incluyan especialistas en protección de datos y no discriminación, Éticos y Expertos en impacto social, junto con los diseñadores, programadores y desarrolladores de *software* para que con sus conocimientos variados, puedan contextualizar el sesgo y diseñar soluciones efectivas.

Con todo este protocolo debería ser posible de garantizar un control adecuado del sesgo algorítmico negativo y una prevención del mismo, pero no se puede olvidar que los Sistemas de IA tienen una prospección de crecimiento exponencial y por ello Castro-Acosta y Andrés Molina (2025) indican:

El uso de la IA en los sitios web, en las apps y en los aparatos informáticos personales ya es una realidad, a la cual el Derecho en general y especialmente el constitucional intenta dar respuesta; prospectivamente se puede augurar que los Estados tenderán a incrementar los controles y las previsiones legales con el fin de adaptar sus regulaciones internas a los nuevos riesgos derivados de tales tecnologías, para poder brindar a sus ciudadanos una protección más adecuada de sus Derechos. Tal vez el mayor riesgo que trae consigo la descentralización absoluta de los datos es el dificultar el seguimiento y la gestión de los sitios web, lo cual podría generar o propiciar un aumento de delitos cibernéticos o el abuso de Internet con fines espurios, que no solo pongan en riesgo a los ciudadanos sino a las estructuras propias de los Estados, a su Seguridad Nacional o a la estabilidad internacional.

Corolario de esta prognosis, es previsible que estos avances tengan que ser acompañados de forma indelible, bien de forma previa, bien de forma posterior, de una evolución legal correlativa, que permita la actualización y/o adaptación de las obligaciones legales y sus consecuentes sanciones por incumplimiento, en temas tan

importantes como los Derechos Fundamentales y los datos personales, para lo cual, también es previsible que se deba generar una tecnología que pueda romper uno o más bloques ya validados por los mineros e integrados en la *blockchain* y rehacerla eliminando los datos personales de el o los bloques específicos en donde se integren los datos que se consideren ilegalmente usados; Para ello se requiere el apoyo de los expertos en *blockchain* para que puedan generar el adelanto informático requerido. (Castro-Acosta y Andrés Molina, 2025)

Con esta investigación se llega a concluir que además de estas regulaciones, las cuales deben ser redactadas desde una visión humano-céntrica y de las sensibilizaciones de los programadores y de expertos en estas tecnologías, se debe contar con un protocolo claro que prevenga, desde el diseño del Sistema de IA la aparición de los sesgos algorítmicos que ponen en jaque a muchos Derechos humanos, con el fin, de que los adelantos de IA ayuden a la humanidad, estén a su servicio.

4. CONCLUSIONES

Luego de todo este, complejo pero muy entretenido e inspirador, proceso de investigación, donde se ha logrado incorporar conceptos que se desconocen en el desarrollo normal de la vida de los profesionales del Derecho, pero que son tan requeridos si se quiere profundizar o especializarse en nuevas tecnologías y en especial en Sistemas de IA, ha llegado el momento de hacer las conclusiones de este artículo, para ello, se hará una primera macro conclusión, la cual luego se desagregará en conclusiones específicas y se concluirá con una conclusión general reflexiva:

Los sistemas de inteligencia artificial (IA) han transformado múltiples esferas de la vida social —desde la justicia penal y la selección de personal hasta la asignación de prestaciones públicas—, pero también han mostrado, en numerosos estudios empíricos y casos reales, la capacidad de reproducir y amplificar sesgos existentes en la sociedad o auto adquiridos de los datos de su diseño o entrenamiento. Estos sesgos algorítmicos deben ser analizados, por los Gobiernos, empresas, usuarios de la información y por los programadores y diseñadores de tales Sistemas, desde una perspectiva de Derechos Humanos, integrando la evidencia del cumplimiento de todo el marco normativo europeo y español aplicable a tales sistemas, y para ello se ha propuesto un protocolo integrado por cuatro fases, en donde se obliga el análisis profundo de tales sesgos, tendiente a evitar y/o mitigar estos sesgos, partiendo de una mirada victimo-céntrica de los Sistemas de IA, que aseguren la protección de los Derechos humanos durante todo el ciclo de vida de los Sistemas de IA.

Los Sistemas de Inteligencia Artificial son una realidad tecnología global, caracterizada por la transversalidad al impactar múltiples dimensiones humanas: sociales, culturales, políticas, jurídicas, económicas, lúdicas, recreativas, entre otras, debido a que al procesar, en poco tiempo, grandes volúmenes de datos y permitir la toma de decisiones automatizadas, permite grandes avances para la humanidad, pero no se

puede desconocer los grandes riesgos que traen aparejados, y en donde, si el centro de su análisis y de su gestión no se retorna a los Seres Humanos, se estarán creando graves caldos de cultivo para la vulneración sistemática e incontenible de los Derechos Humanos, especialmente en relación con Derechos tan importantes como la igualdad y la no discriminación, e incluso la Vida y la Libertad.

Corolario de lo anterior, se debe indicar que con esta investigación se ha podido establecer un marco conceptual que ha incluido Sistemas de IA y sesgos algorítmicos; se ha sintetizado diversas evidencias empíricas y casuísticas sobre los sesgos algorítmicos; se ha conseguido mapear el marco jurídico europeo y español aplicable y relevante; y se ha propuesto un protocolo con lineamientos técnicos que desde una visión humano céntrica alineen los Sistemas de IA con los estándares internacionales de Derechos Humanos.

La investigación ha trabajado con un ejercicio práctico donde a la IA Gemini de Google se le incluían una serie de *prompts* que se indican en el cuerpo del artículo y se dejan referenciada sus respuestas, las cuales sirven para estructurar un entendimiento más profundo de los Sistemas de IA y de los sesgos.

Se ha logrado efectuar una taxonomía de los sesgos algorítmicos que impactan los Sistemas de IA, entre los cuales se conceptualizaron y ejemplificaron el Sesgo de Representación (*Representation Bias*), el Sesgo de Medida (*Measurement Bias*), el Sesgo de etiquetado (*Label Bias*), el Sesgo por objetivo (*Objective Bias*) y el Sesgo Emergente en Sistemas Complejos (*Emergent Bias*) (Comportamiento no previsto)

Se introducen las bases conceptuales de los Derechos Humanos en la Era Digital, y para ello se analiza el marco normativo europeo y español y los Impacto sobre Derechos humanos y los principios afectados, analizando el impacto en los Derechos: a la igualdad y no discriminación, a la privacidad y protección de datos personales y al Debido proceso y derecho a una decisión motivada

Todo lo investigado condujo esta investigación a la necesidad de estructurar un protocolo de una estrategia integrada en fases para prevenir los Sesgos Algorítmicos en los Sistemas de IA, la cual se ha dividido en una fase 0 o previa, tres fases correlacionadas y una fase transversal que articulará todo el modelo de prevención y actuación ante la detección de sesgos algorítmicos en un Sistema de IA.

En la fase previa, se puede resaltar, que partiendo de reconocer la gran importancia de los diseñadores y programadores y demás miembros del equipo de trabajo de un Sistema de IA, se debe trabajar en la sensibilización temprana, para que sean plenamente conscientes de su inestimable papel para la protección de los Derechos de los Seres Humanos durante toda la vida de los Sistemas de IA.

En la fase de prevención y control de datos antes del entrenamiento se pretende detectar la raíz del problema mediante auditorias y posteriores desesgo de los datos, identificando la distribución de los atributos sensibles, aplicando técnicas de balanceo para igualar proporciones, luego de ser requeridos aumentando los datos sintéticos o los conjuntos existentes para lograr una anotación justa del dato y una recolección de datos que reflejen la diversidad real.

Continuando a la fase de Modelado, en donde se diseñen métricas de equidad, se asegure la paridad de impacto, se apliquen técnicas de desesgo y Modelos explicables para detectar sesgos ocultos. Y posteriormente a la fase de despliegue y monitoreo, al establecer el modelo en un entorno real y comprobar con pruebas rigurosas su funcionamiento y específicamente que no reaparezcan sesgos, a través del monitoreo continuo y aplicando políticas de transparencia del modelo.

Este protocolo incluye una fase transversal que la investigación concluye como un aporte científico importante, consistente en la potenciación de la Gobernanza ética y regulatoria, en donde se estructuren Guías éticas internas, técnicas de *compliance* o cumplimiento regulatorio, y la utilización adecuada de equipos interdisciplinarios de expertos que aunado sus conocimientos variados, puedan contextualizar el sesgo y diseñar soluciones efectivas y aplicables, que utilicen como centro al ser humano.

Sirva como conclusión final la siguiente reflexión: Los sistemas de IA son realidades que indiscutiblemente la humanidad ya ha incluido en su forma de vida, abarcan todas las situaciones y actividades importantes para los Seres Humanos, desconocer su gran aporte y ayuda, sería negar a la humanidad este importante paso, pero, por grande que sea, o por el futuro que se le vea, éste paso no puede darse a cualquier costo, la humanidad ha logrado el reconocimiento e implementación casi global de los Derechos Humanos y de la Democracia, a unos costos bastantes altos en términos de tiempo y pérdida de vidas humanas, y ahora, so pretexto de avances científicos o tecnológicos, no se puede permitir que, por no ubicar en el centro de tales avances al propio ser humano, subyugando todos los Sistemas o tecnologías al bienestar y preservación de la especie humana como tal, y de forma indesligable a ella, a los Derechos y Garantías adquiridas, implicaría un riesgo tan elevado, que sería impensable asumirlo desde un punto de vista prudente.

Pero todo esto sirva para decir, que unos Sistemas de IA, con la debida sensibilización y aplicación transversal durante toda la vida de los mismos, de gestión del riesgo y de principios y criterios éticos de gobernanza y *compliance*, llevará a que la Humanidad se beneficie de las grandes virtudes y posibilidades que representan los Sistemas de IA pero que evite los graves riesgos, que en esta investigación se ha confirmado, que son anejos a cualquier sistema de IA, máxime si el mismo es creado con fines espurios, justamente dirigido a vulnerar los Derechos humanos de forma intencional y dolosa.

No pretende este artículo ser un augurador de malos resultados futuros, pero si, que reconociendo como prognosis una realidad adversa causada por una exigua o inadecuada regulación de estos sistemas de IA, el real llamamiento de este artículo, es a que, reconociendo esta posible realidad adversa, se persiga fervorosamente su evitación o su cambio y convertir el futuro en una realidad positiva para el desarrollo de la humanidad, y para la consolidación y protección de los Derechos humanos de todos las personas sin distinguir alguno entre ellas, basado o apoyándose en los Sistemas de IA.

APÉNDICE BIBLIOGRÁFICO

- Armero, Ramón y Melton, Mónica. “El imperdonable error de la IA: científicos y empresarios intentan solucionarlo”. *Business Insider*. 24/07/2024 <https://www.businessinsider.es/tecnologia/imperdonable-error-ia-cientificos-empresarios-intentan-solucionarlo-1381267>
- Barrio Andrés, Moisés. *Reglamento UE de inteligencia artificial (Incluye los actos de desarrollo y ejecución de la AI Act)*. Claves Prácticas. Madrid: Francis Lefebvre, 2025.
- Buolamwini, J. & Gebru, T. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. *Proceedings FAT/ML*.2018.
- Castro Acosta, Angélica María y Andrés Molina, Oscar. *Estudio jurídico sociológico de los retos que emanan de los avances en Tecnologías de la Información y las Comunicaciones (TIC) e Inteligencia Artificial (IA) en el Constitucionalismo Español y Europeo*. En Revista de Derecho Digital e Innovación. El n.º 26 (SEP-DIC). Madrid: Aranzadi La Ley, 2025.
- Consejo de Europa. *Algorithms and Human Rights - Study on the human rights dimensions of automated data processing techniques and posible regulatory implications*, Council of Europe study, DGI(2017)12;
- *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*. 2019. <https://rm.coe.int/unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco>
- *Conclusiones de la Conferencia Governing the Game Changer – Impacts of artificial intelligence development on human rights, democracy and the state of law (Gobernar el cambio de juego: Impactos del desarrollo de la inteligencia artificial en los derechos humanos, la democracia y el estado de derecho)*. Helsinki. 2019. <https://rm.coe.int/conclusions-from-the-conference/168093368c>
<https://www.coe.int/en/web/portal/-/artificial-intelligence-helsinki-conference-conclusions#:~:text=News%202019&text=The%20key%20message%20of%20the,well%20as%20multi%2Dstakeholder%20cooperation>.

- *El Consejo de Europa y la Inteligencia Artificial. Visión global de las actividades del Consejo de Europa en el ámbito de la inteligencia artificial*. Estrasburgo. 2023. <https://rm.coe.int/brochure-artificial-intelligence-es-march-2023-print/1680aab8e8>
- *Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho*. 2024. <https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX:52024PC0264>
- Comisión Europea, “*White Paper: On Artificial Intelligence - A European approach to excellence and trust*”, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
- Dave, Paresh. “*Los algoritmos de IA tienen un sesgo contra la piel con tonos amarillos*”. Wired. 6/11/2023. <https://es.wired.com/articulos/los-algoritmos-de-ia-tienen-sesgo-contra-piel-de-tonos-amarillos>
- Goodman, Amy y Moynihan, Denis. “*Sesgo racial y abuso de la tecnología, desde Henrietta Lacks hasta Porcha Woodruff*”. 11/08/2023. Democracy Now. <https://www.democracynow.org/es/2023/8/11/sesgo-racial-y-abuso-de-la>
- ISO/IEC 8183. Estándar de la International Electrotechnical Commission (IEC) para la gestión del ciclo de vida de los datos de IA.
- Jonker, Alexandra y Rogers, Julie. “*¿Qué es el sesgo algorítmico?*”. IBM. (s/f). <https://www.ibm.com/es-es/think/topics/algorithmic-bias>
- Martínez, Lucía. *Peligrosidad, algoritmos y due process: el caso state vs loomis*. Revista de Derecho Penal y Criminología (UNED), n. 20, 2018.
- Miró, Fernando. *Inteligencia artificial y justicia penal: más allá de los resultados causados por robot*. Revista de Derecho Penal y Criminología (UNED), n. 20, 2018.
- Parlamento Europeo, “*¿Qué es la Inteligencia Artificial y Cómo se usa?*”. 2021. <https://www.europarl.europa.eu/topics/es/article/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa>
- Pérez y Madrid, Aniceto. “*El Reconocimiento Facial es un Superpoder. Cómo te afectará y por qué deberías conocerlo*”. Madrid: Dykinson, 2021. <https://e-archivo.uc3m.es/rest/api/core/bitstreams/72be4448-72cf-4638-a6bc-cd3c5541c7f8/content>
- Pejcinovic Buric, Marija. *Artificial intelligence and human rights*. s/f. <https://www.coe.int/en/web/artificial-intelligence/secretary-general-marija-pejcinovic-buric>
- Recomendación sobre Inteligencia Artificial de la OCDE, <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) (Texto pertinente a efectos del EEE). 2016. <https://www.boe.es/doue/2016/119/L00001-00088.pdf>

- Reglamento (EU) 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) nº 300/2008, (UE) nº 167/2013, (UE) nº 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). <https://www.boe.es/buscar/doc.php?id=DOUE-L-2024-81079>
- Roa Avella, Marcela del Pilar. Sanabria-Moyano, Jesús Eduardo. y Dinás-Hurtado, Katerin. “Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos”. Fundamentos de Direito Processual Penal. (8) 1. Jan-Apr 2022. <https://doi.org/10.22197/rbdpp.v8i1.615>
- Romeo, Carlos. “Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad.” En *Revista de Derecho, Empresa y Sociedad (REDS)*, n. 13, 2018.
- OCDE. *AI Principles*. 2019. <https://www.oecd.org/en/topics/ai-principles.html>
- *Recomendación sobre Inteligencia Artificial*. OECD/LEGAL/0449. Adoptada el 22/05/2019. Modificada el 03/05/2024. 2019/2024 <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>
- UNESCO – Recommendation on the Ethics of Artificial Intelligence (2021) <https://unesdoc.unesco.org/ark:/48223/pf0000381133>
- Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. (LOPDGDD). 2018.
<https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673>
- Villegas Delgado. César. Coordinador. *El Derecho de la Encrucijada Tecnológica. Estudios sobre Derechos Fundamentales, nuevas tecnologías e inteligencia artificial*. Tirant Lo Blanch. 2022.
- Zuiderveen Borgesius, F.J., *Discrimination, artificial intelligence and algorithmic decision-making*. studio del Consejo de Europa, Dirección General de Democracia, DG of Democracy. 2018



CAPÍTULO 5.

ABOGACÍA E INTELIGENCIA ARTIFICIAL: ALGUNAS NOTAS SOBRE LAS OPORTUNIDADES Y DESAFÍOS PARA LA PROFESIÓN¹⁵⁷

IRENE BLANCA CERRILLO TORQUEMADA¹⁵⁸

Abogada en el Área de Estrategia, Seguimiento y Transformación de los Servicios Jurídicos de Banco Bilbao Vizcaya Argentaria, SA.

Resumen

El mundo jurídico no es ajeno a la incorporación de los modelos de lenguaje grandes (LLM por sus siglas en inglés). Estos sistemas son capaces de generar texto para documentos de uso ordinario e, incluso, extraordinario en el tráfico jurídico, respuestas a consultas jurídicas y análisis de normas y resoluciones. Además, han demostrado ser alumnos aventajados a la hora de superar exámenes de acceso a la profesión de abogado¹⁵⁹. En este trabajo se reflexiona sobre las ventajas que estas herramientas pueden suponer para el acceso de los ciudadanos a la justicia. Asimismo, se pone el foco sobre la amenaza que puede representar un uso negligente o malintencionado de las mismas y la forma en la que el gremio de los abogados está tratando la cuestión.

Son las luces y las sombras que rodean a una novedad de reciente incorporación al universo jurídico y que, como todas las herramientas tecnológicas emergentes, puede suponer un cambio de paradigma en el mundo del Derecho según se ha venido aplicando en el panorama de la cultura occidental desde hace varios siglos.

Palabras clave

Inteligencia Artificial, abogacía, directrices, riesgos.

¹⁵⁷ El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

¹⁵⁸ Licenciada en Derecho por la Universidad Complutense de Madrid. Máster en Mercados Financieros por el Instituto de Estudios Bursátiles. Diploma de Estudios Avanzados en la Universidad Complutense de Madrid. Especialista en Derecho Financiero y Bancario.

¹⁵⁹ La Inteligencia Artificial aprueba el examen de acceso a la abogacía:

<https://legamy.com/noticia/inteligencia-artificial-aprueba-examen-para-ser-abogado>

(accedido el 20-08-2025) GPT-4, de OpenAI, superó la prueba Uniform Bar Exam en (UBE) en Estados Unidos colocándose en el percentil 90.

Abstract

The legal world is no stranger to the incorporation of large language models (LLMs). These systems are capable of generating text for documents of ordinary and even extraordinary use in legal transactions, responses to legal inquiries, and analysis of regulations and resolutions. In addition, they have proven to be high achievers when it comes to passing bar exams.

This pages reflect on the advantages that these tools can bring to citizens' access to justice. It also focuses on the threat that negligent or malicious use of these tools can pose and how the legal profession is addressing the issue. Those are the pros and cons surrounding a recent addition to the legal universe which, like all emerging technological tools, could represent a paradigm shift in the world of law as it has been applied in Western culture for several centuries.

Key words

Artificial Intelligence, bar, guidelines, risks.

Sumario: INTRODUCCIÓN 1. UTILIZACIÓN POR LOS ABOGADOS: LEGAL PROMPTING 2. RECOMENDACIONES OFICIALES PARA EL ÁMBITO DE LA ABOGACÍA 3. CONCLUSIONES APÉNDICE BIBLIOGRÁFICO FUENTES JURÍDICAS Y WEBGRAFÍA

INTRODUCCIÓN

La irrupción de la Inteligencia Artificial está suponiendo un importante cambio de paradigma en el mundo. Pocos cuestionan el enorme valor intrínseco de esta nueva herramienta tecnológica y el gran paso adelante que supone en la historia de la humanidad. Sin embargo, la rapidez con que esta novedosa tecnología está posicionándose en el mundo jurídico como sustitutivo, muchas veces, de tareas que antes debían ser desempeñadas exclusivamente por personas específicamente entrenadas para ello, sin el apoyo de algoritmos preelaborados e ideados a fin de generar soluciones rápidas, aconseja detenerse a reflexionar acerca de la conveniencia de estructurar adecuadamente estas nuevas tecnologías y las condiciones que deben cumplir los abogados para utilizarlas sin menoscabo del elemento humano que debe siempre predominar en el mundo del Derecho.

Algunas organizaciones relacionadas con el ejercicio de la abogacía, a nivel nacional e internacional, conscientes de los riesgos que supone la incorporación de la Inteligencia Artificial a la solución de conflictos jurídicos, han salido al paso del problema publicando una serie de guías que permiten a los profesionales del Derecho posicionarse para hacer el mejor uso posible de esas herramientas desde la perspectiva de la evidente economía procedimental, pero también desde unos sólidos fundamentos éticos para que en ningún caso tal innovación se vuelva en contra de los legítimos intereses de los usuarios.

A lo largo de estas páginas se analizan algunos aspectos relacionados con el *legal prompting*, la Inteligencia Artificial generativa, el nuevo papel de los profesionales y la nueva ética en el ejercicio de la abogacía cuando interviene la utilización de herramientas tecnológicas, los riesgos que deben superarse para que el acceso a la justicia no se resienta por el uso indiscriminado e irreflexivo de las nuevas tecnologías, etc.

1. UTILIZACIÓN POR LOS ABOGADOS: *LEGAL PROMPTING*

La utilización de la Inteligencia Artificial generativa por los profesionales del Derecho, y en concreto por los de la abogacía, promete grandes avances en eficiencia. Mediante instrucciones proporcionadas con lenguaje natural se pueden obtener resultados para los que antes hubiera sido necesario la asistencia de especialistas en programación, así como la dedicación de gran cantidad de tiempo. Cuestiones administrativas como la cumplimentación de los datos de un contrato pueden ser resueltas en cuestión de segundos. Asimismo, la IA es capaz de acometer tareas más complejas, como proponer ideas sobre las implicaciones de una nueva normativa o cambio jurisprudencial, además de ofrecer distintas perspectivas sobre una cuestión jurídica. Con el transcurso del tiempo los resultados son cada vez más precisos y las herramientas más avanzadas.

Funciones como la investigación profunda (*Deep Research*) replican en pocos minutos los pasos que seguiría una persona en horas, incluso días, en tareas de búsqueda, comparación, análisis de contradicciones o vacíos legales, construcciones de líneas argumentales, etc. El llamado *modo agéntico* permite dar instrucciones en lenguaje natural a la máquina para que tome el control del navegador y ejecute acciones en otras plataformas. También es posible programar tareas y recordatorios. A todo ello hemos de añadir que la máquina se va entrenando con las interacciones que recibe de los humanos, por lo que el potencial de respuesta es, cada vez, de mayor calidad. Se habla de que, hoy en día, el uso de ChatGPT puede ser equivalente al apoyo que da un pasante.

Además, cada vez existe una mayor oferta en el mercado de herramientas de IA especializada en el ámbito jurídico. Por ejemplo, las tradicionales bases de datos de jurisprudencia están acometiendo desarrollos para permitir búsquedas y generación de escritos a través del lenguaje natural, o comparaciones normativas. De la misma manera, se están ofreciendo gestores de expedientes e información para grandes despachos, instrumentos de ayuda para la toma de decisiones (cálculo de la probabilidad de éxito en un litigio), simuladores de escenarios de negociación, búsqueda de información relevante en grandes volúmenes de documentos.

Todas estas tareas se pueden gestionar a través de herramientas que podríamos llamar *generalistas*, como ChatGPT, o especializadas y entrenadas para determinadas funciones.

En el caso del uso de herramientas generalistas, el mero uso del lenguaje natural puede dar lugar a unos resultados mediocres o incluso equivocados. A fin de obtener

un resultado óptimo es necesario conocer las mejores fórmulas para darle instrucciones a la máquina. Es lo que se conoce como *prompting*, o en el caso de usos relacionados con el mundo jurídico, *legal prompting*. Un estudio de 2022 analiza y compara distintas técnicas experimentadas para obtener resultados de razonamiento jurídico, como por ejemplo IRAC (*Issue, Rule, Application, Conclusion*), ofreciendo también una reflexión sobre si es realmente posible enseñar a pensar como un abogado¹⁶⁰.

De forma general, resulta útil tener en cuenta los siguientes consejos si se quiere optimizar la respuesta del modelo:

- Definición del contexto: Es importante instruir al modelo sobre cuestiones como el contexto geográfico (pues la legislación aplicable no será la misma) o profesional.
- Asignación de un rol: Resulta de gran utilidad indicar a la máquina que debe actuar como un abogado especialista en determinada materia. También conviene indicar quién será el destinatario del *output* (el cliente, el abogado contrario, el contrario...).
- Identificar el tono (profesional, académico, ...).
- División del proceso en tareas sencillas, explicación de los pasos que se han de seguir, etc.
- Ejemplos: proporcionar un caso resuelto resulta muy efectivo a la hora de conseguir mejores resultados.

Además, se pueden añadir guardarraíles que acoten la respuesta obtenida: “*no utilizar la normativa catalana*”, “*tener en cuenta el Código Deontológico de la Abogacía española*” El usuario puede dar instrucciones adicionales hasta que el resultado se ajuste a sus necesidades.

Es fundamental tener siempre presente que la Inteligencia Artificial generativa es un modelo estadístico de lenguaje. El modelo no consulta las fuentes en tiempo real, sino que predice, según patrones estadísticos, la palabra más probable que sigue en la cadena. Por ello, se pueden generar las llamadas alucinaciones, que son resultados que no tienen que ver con las fuentes. Por ello, no se debe asumir la certeza jurídica de la

¹⁶⁰ Fangyi Yu, Lee Quartey, Frank Shilder, “Legal Prompting: Teaching a Language Model to Think Like a Lawyer”, *ArXiv.org* (2022): “...while our analysis shows significant promise in prompt engineering for high-order LLM-based reasoning tasks, it is questionable whether prompting actually teaches a LM to ‘think like a lawyer’ and further exploration is warranted to determine the likelihood of success of fine-tuning with explanation and (legal) reasoning prompting on the legal or other domains and data sets.” Traducción de ChatGPT: “...aunque nuestro análisis muestra un gran potencial en la ingeniería de prompts para tareas de razonamiento de alto nivel basadas en LLM, es cuestionable, si el prompting realmente enseña a un modelo de lenguaje a ‘pensar como un abogado’, y se requiere una exploración más profunda para determinar la probabilidad de éxito del fine-tuning con explicaciones y del prompting basado en razonamiento (legal) en el ámbito jurídico u otros dominios y conjuntos de datos.”

respuesta obtenida y siempre se ha de contrastar la fuente original¹⁶¹. A tal fin, resulta de utilidad solicitar a la herramienta que proporcione un enlace a la fuente de la que ha obtenido la información. Esto significa, como ya anticipábamos, que la IA generativa no razona, y mucho menos como un abogado.

La ingeniería de *prompts* en el mundo de la abogacía está llamada a convertirse en una disciplina en sí misma, y es previsible que en el futuro los perfiles profesionales que destaquen en el mundo jurídico sean aquellos que la dominen. Es por ello por lo que conviene que los abogados sean celosos de regalar¹⁶² su forma de dar instrucciones a la máquina, tengan en cuenta que la máquina se *auto entrena* con ellas y es prácticamente imposible recuperar el conocimiento que el modelo ya adquirió. Una forma de cuidar que ese conocimiento no sea compartido con terceros es solicitar a los proveedores de software de Inteligencia Artificial que creen compartimentos estancos dentro del modelo de forma que el entrenamiento solo aproveche a los usuarios de dicho entorno y la información, conocimiento o técnica introducidos no puedan ser *regurgitados* en entornos no deseados.

Además de tener una inmediata utilidad en la redacción de documentos o para el análisis de la normativa, el *legal prompting* puede plantear cuestiones relacionadas con la metodología, cual es la algorítmica como herramienta auxiliar a la hora de construir nuevas interpretaciones jurídicas. Los modelos lingüísticos, además de procesar datos van aprendiendo paulatinamente las estructuras argumentales y de razonamiento que pueden determinar la manera en que los abogados estructuran e interpretan cualquier ordenamiento jurídico. Pero esa imbricación entre los conocimientos del abogado y el aprendizaje que conlleva el *legal prompting* puede conllevar un riesgo que, hasta ahora, no se ha tenido en cuenta: la generación de un razonamiento jurídico estandarizado. Ello sería inevitable si los profesionales del Derecho buscan respuestas a problemas jurídicos recurriendo a *prompts* suministrados a modelos similares, dando como resultado una indeseada homogeneidad en lo que se presentaría como un falso pensamiento

¹⁶¹ José Manuel Chaves, “Convencer al juez citando jurisprudencia con lealtad”, *Abogacía Española*, (2024) El autor refiere que se han dado casos de procedimientos judiciales en los que el abogado ha citado sentencias que no existen, fruto de alucinaciones de la IA. De momento, en España, únicamente han sido sancionados con apercibimiento del tribunal. <https://www.abogacia.es/publicaciones/blogs/blog-comunicacion-y-marketing-juridicos/convencer-al-juez-citando-jurisprudencia-con-lealtad/> (accedido el 08-08-2025) Vid. también: *AI Hallucination Cases Database – Damien Charlotin* (accedido el 22-10-25) Damien Charlotin: *AI Hallucination Cases*. Este investigador alimenta una base de datos de decisiones de los tribunales en casos en los que se han detectado alucinaciones, normalmente citas de resoluciones que no existen, así como la decisión del tribunal (advertencia, apercibimiento, sanción monetaria, orden de notificar al cliente, información al Colegio de Abogados, etc.). A fecha 22-10-2025 se han registrado 457 casos.

¹⁶² En la *Newsletter* práctica del ICAM sobre innovación y tendencias Descubre, Impulsa y Transforma se hace una reflexión sobre la cuestión: <https://www.linkedin.com/pulse/notificaciones-certificadas-la-desconexi%C3%B3n-uxxrf/> (accedido el 08-08- 2025)

jurídico que abocaría a la profesión a la pérdida irremisible de la anterior pluralidad doctrinal. Los *prompts*, pues, no deben entenderse como un producto que ha de manejarse a partir de la adquisición de una mera pericia técnica, sino como una herramienta de la propia actividad intelectual responsable y consciente de la diversidad de interpretaciones en las que se apoya la evolución del Derecho en su conjunto.

También el ámbito de la formación académica de los juristas se puede ver afectado por el *legal prompting*. Los centros de enseñanza de las disciplinas jurídicas han venido basando tradicionalmente su docencia en los textos normativos y jurisprudenciales de manera analítica y casuística, siempre con el fin de intentar desarrollar la destreza argumentativa de los alumnos. Pero la inmediatez del acceso a la información generada en el ámbito de la inteligencia artificial podría impulsar a los alumnos a la dejación del esfuerzo que supone el análisis de las normas, dando como resultado una mera comprensión superficial de las mismas. Ese esfuerzo añadido también compete al profesorado de las facultades de Derecho y de otros centros de enseñanza para el acceso a la profesión de abogado, siendo preciso que los docentes replanteen su tradicional pedagogía haciendo un especial hincapié en el sentido crítico de los alumnos cuando aborden el uso de la Inteligencia Artificial como nueva herramienta profesional. Ello podría evitar el desembarco en la profesión de generaciones de “autómatas” del Derecho incapaces de aproximarse a los modelos generativos con sentido crítico y riguroso, sin perder de vista las bases éticas del ordenamiento jurídico que deben aplicar a cada supuesto de hecho.

Se vienen desarrollando también herramientas especializadas, como bases de datos de jurisprudencia que permiten las consultas con lenguaje natural pero que acotan la búsqueda a su base de datos, de forma que proporcionan la seguridad de que la búsqueda no va a ofrecer resultados inventados.

También para la confección de documentos o la gestión de contratos o expedientes en los grandes despachos se comercializan distintos programas que funcionan con Inteligencia Artificial y garantizan precisión en el funcionamiento. Permiten, por ejemplo, comparar si el contrato de un proveedor se ajusta a unos estándares contractuales determinados, o acometer una modificación masiva de una cláusula.

La implantación de estas herramientas en los despachos de abogados aconseja la creación de ciertas gobernanzas para que se utilicen de forma uniforme y conforme con unos patrones éticos y de calidad, aconsejándose que solo determinados usuarios puedan gestionar y validar las etiquetas que se le dan a distintas cláusulas o modelos.

Hay otro importante aspecto que no debe ser descuidado al hacer alusión al *legal prompting*. Nos referimos al aspecto económico de la transformación del mercado jurídico y de las tradicionales relaciones de competencia entre los diferentes despachos. Es bien sabido que los costes y la facturación de los distintos servicios ofrecidos por los profesionales del Derecho se basaban, generalmente, en el número de horas que se

precisaban para la obtención de los resultados deseados. Sin embargo, el uso de las herramientas generativas está reduciendo notablemente los tiempos que se necesitan para alcanzar dichos resultados. De este modo, los clientes, cada vez más informados de la supuesta facilidad que ofrecen estas herramientas para evitar un trabajo antes tedioso y penoso, van a exigir la reducción de los honorarios a los abogados bajo la excusa de que su carga laboral en horas se ha reducido notablemente, solicitando que se les cargue solamente por aquellos servicios que suponen el trabajo específico y original aportado por el profesional, independientemente del tiempo medido en horas. De esta manera, es previsible que comenzará una carrera entre los diferentes despachos para demostrar su mejor calidad y, en resumidas cuentas, la tasación económica de sus servicios, sobre la base de la aportación verdaderamente humana que, desde luego, la Inteligencia Artificial no puede proporcionar en la resolución de cualquier asunto. Se puede crear así un abismo entre las firmas de abogados con capacidad para invertir en la generación de su propia Inteligencia Artificial privativa y los pequeños despachos que habrán de ofrecer servicios especializados de calidad contrastada para sobrevivir. El tradicional mantenimiento de la competencia leal entre firmas puede verse así seriamente afectado por la brecha tecnológica, con la consiguiente desigualdad de los ciudadanos a la hora de acceder a la administración de justicia.

3. RECOMENDACIONES OFICIALES PARA EL ÁMBITO DE LA ABOGACÍA

Nadie es ajeno a la velocidad a la que avanzan los desarrollos en materia de Inteligencia Artificial ni a los desafíos que el uso de estas herramientas supone.

En el Estatuto de la Abogacía Española¹⁶³ no se regula el uso de esta herramienta en concreto, si bien a continuación desgranamos algunas de las cuestiones que sí se regulan en el mismo que podrían verse afectadas por la irrupción del uso de las herramientas.

En primer lugar, nos referiremos a los requisitos para el ejercicio de la función de abogado.

Comenzamos reproduciendo el artículo 4 del Estatuto de la Abogacía:

Artículo 4. Los profesionales de la Abogacía.

1. Son profesionales de la Abogacía quienes, estando en posesión del título oficial que habilita para el ejercicio de esta profesión, se encuentran incorporados a un Colegio de la Abogacía en calidad de ejercientes y se dedican de forma profesional al asesoramiento jurídico, a la solución de disputas y a la defensa de derechos e intereses ajenos, tanto públicos como privados, en la vía extrajudicial, judicial o arbitral.

¹⁶³Estatuto de la Abogacía Española. Se puede consultar la versión consolidada en <https://www.boe.es/buscar/act.php?id=BOE-A-2021-4568>.

(Accedido el 20-08-2025)

2. Corresponde en exclusiva la denominación de abogada y abogado a quienes se encuentren incorporados a un Colegio de la Abogacía como ejercientes.

Por lo tanto, con la regulación española actual, para poder ejercer la profesión es obligatorio haber obtenido un título oficial habilitante (grado en Derecho). A ello se ha de añadir el requisito de colegiación, regulado también en el artículo 436 de la Ley Orgánica del Poder Judicial¹⁶⁴.

Es decir, que en España solo los graduados en Derecho colegiados pueden ejercer la profesión de abogado, que consiste, también según el Estatuto General de la Abogacía Española¹⁶⁵, en el *asesoramiento jurídico, a la solución de disputas y a la defensa de derechos e intereses ajenos, tanto públicos como privados, en la vía extrajudicial, judicial o arbitral*.

No hemos de olvidar que el abogado desempeña una función esencial, sin que se pueda desviar del fin supremo de Justicia y efectividad de los derechos y libertades fundamentales a que la profesión se halla vinculada¹⁶⁶.

Lo anterior es relevante porque ya se han dado casos en los que se ha planteado la posibilidad de que los *chatbots* jurídicos sustituyan el consejo y la defensa de los intereses de los particulares, lo que podría interferir en la obtención del fin supremo de Justicia y dejar desamparados a sus usuarios en el caso en el que a través de estos se puedan cometer delitos y faltas o su uso resulte en una lesión económica o de cualquier tipo que no esté debidamente cubierta.

Se plantea si el ofrecimiento de estos servicios podría constituir una forma de intrusismo profesional y si, en caso afirmativo, sería deber de los abogados denunciarlo

¹⁶⁴ BOE-A-1985-12666 Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial.

(accedido el 20-08-2025)

Art. 436 LOPJ: *Para el ejercicio de la abogacía es obligatoria la colegiación en un Colegio de Abogados.*

¹⁶⁵ BOE-A-2021-4568 Real Decreto 135/2021, de 2 de marzo, por el que se aprueba el Estatuto General de la Abogacía Española.

(accedido el 21-08-2025)

Art. 4.1 Estatuto General de la Abogacía Española:

Son profesionales de la Abogacía quienes, estando en posesión del título oficial que habilita para el ejercicio de esta profesión, se encuentran incorporados a un Colegio de la Abogacía en calidad de ejercientes y se dedican de forma profesional al asesoramiento jurídico, a la solución de disputas y a la defensa de derechos e intereses ajenos, tanto públicos como privados, en la vía extrajudicial, judicial o arbitral.

¹⁶⁶ *Ibidem*, Art. 1 La Abogacía y sus principios rectores. (...) 5. En el Estado social y democrático de Derecho, los profesionales de la Abogacía desempeñan una función esencial y sirven los intereses de la Justicia, mediante el asesoramiento jurídico y la defensa de los derechos y libertades públicas. 6. La Abogacía española proclama su especial compromiso con el reconocimiento y la defensa de los derechos humanos.

al Colegio¹⁶⁷. (art 34 del Estatuto de la Abogacía Española: *es deber de los abogados denunciar al Colegio todo acto de intrusismo que llegue a su conocimiento*).

En otros gremios, ya se están poniendo en marcha acciones de concienciación a la ciudadanía sobre los riesgos del uso de estos instrumentos por no profesionales (por ejemplo, para el tratamiento psicológico o diagnóstico médico). Es recomendable que también se implemente este tipo de campañas en el ámbito de la abogacía, pues se trata de cuestiones que afectan a algo tan delicado como la tutela judicial efectiva de los ciudadanos. Si bien se podría afirmar que el ciudadano medio es capaz de discernir que un consejo jurídico no es suficientemente cualificado y no cuenta con todas las garantías cuando se obtiene de los medios de la sociedad de la información (Internet), hoy en día, la sofisticación de los *chatbots* jurídicos, puede llegar a confundirle con una falsa sensación de seguridad jurídica, por lo que es aconsejable protegerle.

El Código Deontológico de la Abogacía Española¹⁶⁸, publicado en 2019, tampoco aborda de forma específica la cuestión del uso de sistemas de Inteligencia Artificial, si bien en su artículo 21¹⁶⁹ se regula el empleo de las tecnologías de la información y la comunicación.

Se atisba en el Código que el uso de herramientas tecnológicas puede tener implicaciones en el ámbito deontológico, así como en el del secreto profesional, lo que no es ajeno a las herramientas de las que trata este trabajo.

En cuanto a algunos de los principios concretos a los que se refiere el Código deontológico, han de ser tenidos en cuenta por los profesionales a la hora de hacer uso

¹⁶⁷ *Ibidem*, Art. 87 del Estatuto General de la Abogacía Española: *Obligaciones de los colegiados*:

Son obligaciones de los colegiados: (...)d) Denunciar ante el Colegio todo acto de intrusismo o ejercicio ilegal que llegue a su conocimiento, ya sea debido a falta de colegiación, por suspensión o inhabilitación del denunciado, o por concurrir en supuestos de incompatibilidad o prohibición.

¹⁶⁸ *Código Deontológico de la Abogacía española*

<https://www.abogacia.es/wp-content/uploads/2019/05/Codigo-Deontologico-2019.pdf>

(accedido el 10-09-2025)

¹⁶⁹ *Ibidem*, Art. 21:

1. El uso de las tecnologías de la información y la comunicación no exime de cumplir las normas deontológicas que regulan la profesión ni las obligaciones que imponen las reguladoras de la sociedad de la información.

2. Se debe hacer uso responsable y diligente de la tecnología de la información y la comunicación, debiendo extremar el cuidado en la preservación de la confidencialidad y del secreto profesional.

3. En especial, en las comunicaciones, aplicaciones, webs y servicios profesionales prestados por medios electrónicos deberá:

a. Identificarse con su nombre y, en su caso, el de la sociedad profesional titular del servicio, Colegio de adscripción y número de colegiación. b. Asegurarse de la recepción de las comunicaciones privadas por la persona destinataria y sólo por ella. c. Abstenerse de reenviar correos electrónicos, mensajes o notas remitidos por otros profesionales de la Abogacía sin su expreso consentimiento.

de la Inteligencia Artificial algunas cuestiones relativas a la independencia, la confianza y la integridad. A ello se refiere el título IV del Código Deontológico¹⁷⁰.

Si bien el uso de la IA *per se* no compromete la independencia, confianza e integridad del abogado, sí que se han descrito sesgos en los modelos de lenguaje que pueden comprometer el *output* o resultado que recibe el usuario. Así, la máquina tiende a un exceso de complacencia, y puede llegar a actuar por ¿propios? objetivos (se han descrito casos de inteligencias artificiales que han llegado a tener comportamiento autónomo, siendo capaces de mentir, manipular, y amenazar a sus creadores¹⁷¹). A todo ello se ha de añadir que las herramientas de IA más exitosas podrían afectar a la independencia del abogado interfiriendo en su relación con los clientes o forzando al abogado a aceptar términos que puedan comprometer su independencia.

No menos importantes son las prescripciones relativas al secreto profesional y las obligaciones con los contrarios¹⁷².

¹⁷⁰ *Ibidem*, Tit. IV: Relaciones entre profesionales de la Abogacía y clientes:

Art. 47. *Independencia y libertad del profesional de la Abogacía.*

1. *La independencia y libertad son principios rectores de la profesión que deben orientar en todo momento la actuación del profesional de la Abogacía, cualquiera que sea la forma en que ejerza la profesión. El profesional de la Abogacía deberá rechazar la realización de actuaciones que puedan comprometer su independencia y libertad.*
2. *La relación del profesional de la Abogacía con el cliente debe fundarse en la recíproca confianza.*
3. *En todo caso, deberá cumplir con la máxima diligencia la misión de asesoramiento o defensa que le haya sido encomendada, procurando de modo prioritario la satisfacción de los intereses de su cliente.*
4. *El profesional de la Abogacía realizará, con plena libertad e independencia y bajo su responsabilidad, las actividades profesionales que le imponga la defensa del asunto que le haya sido encomendado, ateniéndose a las exigencias técnicas y deontológicas adecuadas a la tutela jurídica del asunto.*

¹⁷¹ https://www.unotv.com/tecnologia/la-ia-aprende-a-mentir-manipular-y-amenazar-a-sus-creadores-estos-casos-alertan-a-las-personas/?utm_source=chatgpt.com

(accedido el 12-09-2025)

¹⁷² *Ibidem*, CAPÍTULO IV

Secreto profesional

Artículo 21. Secreto profesional.

1. *La confianza y confidencialidad en las relaciones con el cliente imponen al profesional de la Abogacía, de conformidad con lo establecido por la Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial, el deber y el derecho de guardar secreto de todos los hechos o noticias que conozca por razón de cualquiera de las modalidades de su actuación profesional, no pudiendo ser obligado a declarar sobre ellos.*
2. *Lo previsto en el presente capítulo se entenderá sin perjuicio de lo previsto en cada caso por la Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial, por las leyes procesales y demás legislación aplicable.*

Artículo 22. Ámbito del secreto profesional.

1. *El deber y derecho de secreto profesional del profesional de la Abogacía comprende todos los hechos, comunicaciones, datos, informaciones, documentos y propuestas que, como profesional de la Abogacía, haya conocido, emitido o recibido en su ejercicio profesional.*

En este aspecto es en el que se está haciendo especial hincapié por parte de los colegios profesionales a la hora de dar recomendaciones sobre la cuestión. Hay que considerar que, para obtener el mejor *output*, se ha de proporcionar el más completo *input*. Esto implica compartir con la herramienta la información del cliente. Otro uso podría ser la grabación y transcripción de conversaciones con el cliente o con la parte contraria. Esto se debe llevar a cabo con la mayor de las cautelas, puesto que la información suele pasar a formar parte del modelo de lenguaje. Por lo tanto, la primera precaución que debe tener un abogado a la hora de contratar una de estas herramientas es informarse sobre qué es lo que va a hacer el modelo de lenguaje con la información que se le comparta, y valorar si el secreto profesional queda comprometido. Así, el hecho de que la máquina utilice las entradas de información que se le proporcionan para entrenarse, significa que la misma, si se da el caso, puede aparecer a otro usuario. Esto va más allá de la protección de datos personales (cuya regulación excede el objetivo del presente trabajo pero que también es de la mayor relevancia).

En otro orden de cosas, la recomendación también es que el abogado sea transparente con el titular de la información sobre el tratamiento de los datos que va a someter a la Inteligencia Artificial y el uso de estos que va a hacer la herramienta, sin que esto le pueda eximir de responsabilidad sobre el secreto profesional. De la misma manera, en el caso de que se usen asistentes de IA que puedan grabar y transcribir

2. El secreto profesional no ampara las actuaciones del profesional de la Abogacía distintas de las que son propias de su ejercicio profesional y, en especial, las comunicaciones, escritos y documentos en que intervenga con mandato representativo de su cliente y así lo haga constar expresamente.

3. Las conversaciones mantenidas por los profesionales de la Abogacía con sus clientes, los contrarios o sus profesionales de la Abogacía, de presencia o por cualquier medio telefónico o telemático, solo podrán ser grabadas con la previa advertencia y conformidad de todos los intervinientes, quedando en todo caso amparadas por el secreto profesional. Están igualmente amparadas por el secreto profesional, las grabaciones realizadas por el cliente, no conocidas por su profesional de la Abogacía, incluso si éste no lo era o no intervino en dicho momento, de conversaciones en que intervenga el profesional de la Abogacía de la otra parte.

4. El profesional de la Abogacía deberá hacer respetar el secreto profesional a sus colaboradores y asociados, así como al personal correspondiente y demás personas que cooperen con él en su actividad profesional.

5. El deber de secreto profesional permanece incluso después de haber cesado en la prestación de los servicios al cliente, sin que se encuentre limitado en el tiempo.

6. El Abogado quedará relevado de este deber sobre aquello que solo afecte o se refiera a su cliente, siempre que éste le haya autorizado expresamente.

Artículo 23. Confidencialidad de las comunicaciones entre los profesionales de la Abogacía.

El profesional de la Abogacía no podrá aportar a los Tribunales, ni facilitar a su cliente, las cartas, documentos y notas que, como comunicación entre profesionales de la Abogacía, mantenga con el profesional de la Abogacía de la otra parte, salvo que este lo autorice expresamente. Esta prohibición no alcanzará a las cartas, documentos y notas en que intervenga con mandato representativo de su cliente y así lo haga constar expresamente.

conversaciones, resultará necesario obtener el consentimiento de todos los participantes en las mismas.

Pero se entiende que el profesional debe mantenerse firme a la hora de ejercer su cometido, de manera que también se establece la obligatoriedad de abstenerse de seguir las indicaciones del cliente si al hacerlo pudiera comprometer la observancia de los principios que rigen la profesión¹⁷³.

Si bien estas cuestiones se podrían introducir como guardarraíles, resulta de especial trascendencia la necesidad de que siempre haya un ser humano que valide que el resultado obtenido es acorde con los mismos.

En cuanto al aseguramiento de la responsabilidad civil, ya se ha señalado este es uno de los puntos por los que un usuario de *chatbots* con fines de asesoramiento jurídico debería cuestionarse la falsa sensación de seguridad jurídica que pueden producir¹⁷⁴.

El código también trata la forma de que los abogados se publicitan, lo que a priori no parece que pueda suponer riesgos para el letrado si se usa la herramienta para crear piezas publicitarias y siempre que el abogado tenga presentes sus obligaciones como profesional a la hora de diseñarlas. Sobre el particular caso de que un abogado diseñe un *chatbot* legal y pretenda ponerlo a disposición del público, se pueden tener en cuenta las obligaciones generales de transparencia reguladas en el Código¹⁷⁵. Todas estas cuestiones deberían ser transparentes para el usuario del *chatbot*.

Se espera que se publique próximamente por el Consejo General de la Abogacía Española el Libro Blanco sobre Inteligencia Artificial y Abogacía. Ese organismo, junto con el Colegio de Abogados de Valencia, ha lanzado una encuesta¹⁷⁶ de 10

¹⁷³ *Ibidem*, art. 12.a.4: *La libertad de defensa comprende la de aceptar o rechazar el asunto en que se solicita la intervención, sin necesidad de justificar su decisión. Será obligatorio, pues, abstenerse de seguir las indicaciones del cliente si al hacerlo pudiera comprometer la observancia de los principios que rigen la profesión.*

¹⁷⁴ *Ibidem*, art. 20: *Cobertura de la responsabilidad civil 1. Se deberá tener cubierta la responsabilidad profesional en cuantía adecuada a los riesgos que implique. 2. La contratación de un seguro es obligatoria para las sociedades profesionales y en los demás casos que prevea la ley.*

¹⁷⁵ *Ibidem*, apartado B): *Deberes de identificación e información 1. Es obligación de quien ejerce la Abogacía identificarse ante la persona a la que asesora o defiende, incluso cuando lo hiciere por cuenta de un tercero a fin de asumir las responsabilidades civiles y deontológicas que correspondan. También lo es en el supuesto de consulta telefónica o por red informática con un despacho o asesoría cuyos integrantes sean desconocidos para el comunicante. Esta identificación, así como la del Colegio al que se pertenece, es su primera e inmediata obligación antes de la prestación de servicios y, en todo caso, antes de solicitar el abono de contraprestación alguna.*

¹⁷⁶ Esta acción se ha realizado al amparo del Convenio suscrito entre Red.es del Ministerio para la Transformación Digital y de la Función Pública y el Consejo General de la Abogacía Española para el impulso a la implementación de la Carta de Derechos Digitales, en el ámbito de los derechos en los nuevos entornos digitales.

preguntas que le permitirá analizar los beneficios, oportunidades y riesgos que plantea la IA en el ámbito jurídico y cómo afecta a la realidad profesional de la abogacía.

Por su parte, el Ilustre Colegio de la Abogacía de Madrid (ICAM) ha publicado el 7 de octubre de 2025 su *Guía de Buenas Prácticas para el Uso de la IA en la Abogacía*¹⁷⁷. En ella, tras explicar cómo funcionan los sistemas de IA generativa y advertir de sus riesgos¹⁷⁸ (carece de capacidad de comprensión jurídica real, propensión a los errores y sesgos, confidencialidad, protección de datos y seguridad, falta de explicabilidad y “caja negra”, obsolescencia de datos, ...) hace una recomendación de análisis de evaluación de las necesidades del despacho y las posibles soluciones tecnológicas¹⁷⁹. Se aconseja, por ejemplo, hacer ensayos en entornos controlados antes de su implementación definitiva, seleccionar herramientas que cumplan con estándares internacionales reconocidos, como la norma ISO/IEC 42001 sobre sistemas de gestión de IA, la ISO/IEC 27001 sobre seguridad de la información o la ISO/IEC 27701 sobre privacidad y documentar la versión de la herramienta utilizada y el fin concreto para el que se aplica, para justificar su empleo ante clientes, tribunales o autoridades en caso necesario. El abogado debe estar en condiciones de justificar profesionalmente por qué recurrió a una herramienta de Inteligencia Artificial, explicando el beneficio que aportó al cliente y la forma en que se integró en el encargo¹⁸⁰. Asimismo, se recomienda en la Guía que

<https://www.abogacia.es/actualidad/noticias/participa-en-la-encuesta-sobre-el-uso-de-la-inteligencia-artificial-en-la-abogacia/>

(accedido el 22-09-2025)

¹⁷⁷ *Guía ICAM de Buenas Prácticas para el Uso de la Inteligencia Artificial en la Abogacía*. Ilustre Colegio de la Abogacía de Madrid.

https://web.icam.es/wp-content/uploads/2025/10/Guia-ICAM_IA_2025_2.pdf

(accedido el 8/10/2025)

¹⁷⁸ *Ibidem*, 2.1.3), p. 14.

¹⁷⁹ *Ibidem*, Cap. 2.3, *Buenas prácticas*, p. 16:

Es fundamental que los profesionales de la abogacía y despachos comprendan en profundidad el funcionamiento de la IA —y en particular de la IA generativa— antes de incorporarla en sus procesos, teniendo presente sus limitaciones (falta de comprensión jurídica real, sesgos, errores o alucinaciones). Por ello, debe promoverse siempre una diligencia proactiva reforzada en la selección, integración y uso efectivo de estas herramientas en el ejercicio profesional.

¹⁸⁰ *Ibidem*:

El uso de la inteligencia artificial en la abogacía debe estar siempre guiado por criterios de necesidad y proporcionalidad. Su incorporación no puede responder a la comodidad o a la inercia tecnológica, sino que debe emplearse únicamente cuando aporte un valor real y concreto al servicio jurídico. La adecuación al caso constituye un principio esencial. El nivel de automatización debe ajustarse a la naturaleza y a la complejidad del asunto, evitando un uso excesivo o desproporcionado que pueda afectar a la calidad del servicio o generar riesgos innecesarios. Asimismo, el abogado/a debe estar en condiciones de justificar profesionalmente por qué recurrió a una herramienta de inteligencia artificial, explicando el beneficio que aportó al cliente y la forma en que se integró en el encargo (...) Realizar una evaluación exhaustiva permite seleccionar la tecnología adecuada, optimizando la eficiencia de la función legal y asegurando el cumplimiento normativo y ético. Este enfoque estratégico es el primer paso para una integración exitosa y responsable de la IA en la práctica legal. Las garantías

los profesionales de la abogacía cuenten a su vez con asesoramiento jurídico especializado sobre IA, sin que esto signifique que deba desconocer la regulación sobre la materia.

Los profesionales deberán asegurar que el equipo cuenta con el nivel de competencia tecnológico necesario para comprender, supervisar y evaluar el funcionamiento de las herramientas de IA. A tal fin, se recomienda invertir en formación y realizar evaluaciones internas. Advierte también de la necesidad de colaborar con tecnólogos y expertos en seguridad de la información, gobierno del dato, deontología, ética y cumplimiento normativo y asignar claramente los roles y responsabilidades sobre la materia en el despacho.

También se debe evitar, según se indica en la Guía, el uso de herramientas que puedan generar asesoramiento jurídico sin supervisión humana, o mermar otro tipo de garantías básicas conforme la normativa vigente, para no vulnerar la misma ni las reglas de práctica autorizada. Por último, la misma tiene un interesante glosario¹⁸¹ de términos

de esta fase inicial deberán mantenerse en el tiempo, aplicando un enfoque continuo de riesgo para garantizar tal cumplimiento en todo momento...

¹⁸¹ *Ibidem*, Cap. 3. GLOSARIO, pp. 31-34. *Alucinaciones de la IA: Situaciones en las que un sistema de IA genera información incorrecta, engañosa o sin base en los datos reales, lo que puede llevar a resultados inesperados o problemáticos. Análisis Predictivo: Uso de la IA para predecir eventos futuros basándose en datos históricos. Aprendizaje Automático (Machine Learning): Subcampo de la IA que permite a los sistemas aprender patrones a partir de datos y generalizar a casos nuevos, logrando de este modo mejorar su rendimiento sin necesidad de programación explícita para cada tarea. Automatización: Uso de sistemas de IA para realizar tareas repetitivas sin intervención humana. Bias (Sesgo): Inclinação o prejuicio no deseado en los resultados de la IA, a menudo debido a datos de entrenamiento no representativos. Big Data: Conjuntos de datos masivos que se utilizan para entrenar modelos de IA. Chatbots: Sistemas de IA que simulan conversaciones humanas. Dataset o Datos de Entrenamiento: Conjunto de datos utilizados para entrenar o evaluar un modelo de IA. | 31Capítulo 3. Glosario. Discriminación Algorítmica: Resultado de sesgos en algoritmos que perpetúan desigualdades sociales. Ética Algorítmica: Conjunto de principios que guían el diseño y uso de sistemas de IA para garantizar que respeten valores como la justicia, la privacidad y la igualdad. Explicabilidad: Capacidad de un sistema de IA para ofrecer explicaciones comprensibles sobre los procesos y resultados que genera. Grandes Modelos de Lenguaje (LLM): Algoritmos de aprendizaje profundo diseñados para procesar y generar texto de manera similar al lenguaje humano, como GPT, DeepSeek, PaLM o Cohere, entrenados con vastas cantidades de datos textuales. IA Agent (Agente de Inteligencia Artificial): Sistema basado en inteligencia artificial que percibe su entorno, procesa información y toma decisiones para alcanzar un objetivo específico. Puede actuar de manera autónoma o semiautónoma en diversas tareas, desde chatbots hasta robots inteligentes. Ejemplos: asistentes como Siri o Alexa, sistemas de recomendación de las plataformas audiovisuales, coches autónomos, domótica avanzada, procesos de seguimiento de envíos en e-commerce, etc. IA General: Tipo de IA capaz de realizar cualquier tarea cognitiva que un ser humano puede realizar. IA Generativa: Subcampo de la IA centrado en crear contenido nuevo, como texto, imágenes o música, basándose en datos previos. Inteligencia Artificial (IA): Campo de la informática que desarrolla sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, la resolución de problemas y la toma de decisiones. Procesamiento de Lenguaje Natural (PLN): Área de la IA enfocada en la interacción entre computadoras y el lenguaje humano. Prompt: Conjunto de instrucciones o texto que se proporciona a un modelo de lenguaje para guiar la generación de una respuesta.*

relacionados con las nuevas tecnologías en general y la IA Generativa en particular. En él se contienen todas las novedades terminológicas al respecto, sirviendo de útil *vademecum* para los profesionales de la abogacía.

El 2 de diciembre de 2025 el ICAM ha publicado en el espacio digital Descubre. Impulsa. Transforma una recomendación de uso de las herramientas de IA lingüística gratuita de la Comisión Europea, indicando que son gratuitas, seguras y especialmente útiles para quienes trabajan con documentación jurídica¹⁸².

El Consejo de la Abogacía Europea (CCBE) también ha abordado de manera integral el uso de la Inteligencia Artificial (IA) en la profesión jurídica mediante la publicación de la *Guía sobre uso de herramientas basadas en Inteligencia Artificial por abogados y despachos jurídicos en la Unión Europea*¹⁸³, presentada en marzo de 2022 como resultado del proyecto AI4Lawyers, financiado por la Unión Europea en colaboración con la *European Lawyers Foundation* (ELF)¹⁸⁴.

En dicha Guía se destaca la importancia para la sociedad del uso de las herramientas de IA por los abogados, sobre todo para que los despachos pequeños, con menor presupuesto dedicado a innovación tecnológica, puedan competir con las grandes firmas. La guía incluye una sección en la que se ofrecen distintos escenarios de cómo los

En el contexto legal, un prompt bien formulado puede ayudar a obtener borradores o análisis específicos de documentos jurídicos. Redes Neuronales Artificiales: Modelos computacionales inspirados en la estructura del cerebro humano que se utilizan en tareas como el reconocimiento de imágenes, el procesamiento de lenguaje natural y otros problemas complejos. Responsabilidad Algorítmica: Obligación de las entidades que desarrollan o usan sistemas de IA de rendir cuentas sobre los resultados y decisiones generados por estos sistemas. Riesgo Alto (High-Risk AI): Según el Reglamento de IA de la UE, se trata de sistemas de IA cuya aplicación puede impactar significativamente en los derechos fundamentales o la seguridad de las personas, como los utilizados en la justicia, la educación o la salud. Sandbox (o entorno de pruebas): Espacio aislado y controlado donde se puede ejecutar, probar o experimentar con nuevas aplicaciones o modelos de IA sin afectar los sistemas productivos. Esto permite evaluar la funcionalidad y seguridad de una herramienta antes de su implementación en un entorno real. Sesgo Algorítmico: Desviación o inclinación en los resultados de un | 33 Capítulo 3. Glosario. sistema de IA debido a errores en los datos de entrenamiento o en el diseño del algoritmo. Sistemas de Caja Negra: Tecnologías de IA cuyos procesos internos no son comprensibles para los humanos, dificultando la explicación de cómo se obtuvieron ciertos resultados. Token: Unidad básica de procesamiento en un modelo de lenguaje (puede ser una palabra, parte de ella o incluso un carácter), utilizada para analizar y generar textos. Trazabilidad: Capacidad de seguir y registrar todas las etapas del procesamiento de datos en sistemas de IA, crucial para garantizar responsabilidad y transparencia.

¹⁸² <https://www.linkedin.com/pulse/herramientas-de-ia-segura-la-ce-gu%C3%ADa-jdjyf/>
(accedido el 3.12-2025)

¹⁸³ [Guía sobre el uso de herramientas basadas en IA por abogados y despachos de abogados en la UE](#)
(accedido el 11.10-2025)

¹⁸⁴ European Lawyers Foundation.

ai4lawyers.eu+2ccbe.eu+2fbc.org+2
(accedido el 12-10-2025)

despachos pueden usar esta tecnología para hacer más eficientes sus operaciones¹⁸⁵. También hace una relación de riesgos¹⁸⁶ para los principios de la profesión, especialmente la independencia del abogado, la obligación de evitar conflictos de interés y el deber de secreto profesional. En muchos de los casos, los riesgos son similares a aquellos con los que ya convivíamos, derivados de la utilización de plataformas online (por ejemplo, uso de *cloud computing* y recuperación de los datos, riesgos para la privacidad). Nos centraremos en los riesgos relacionados específicamente con la aparición de las herramientas de IA:

- Dificultad, o incluso imposibilidad de explicar o interpretar la forma de funcionar de los modelos. Por ejemplo, pueden aparecer sesgos difíciles de interpretar¹⁸⁷.

¹⁸⁵ *Ibidem*, 2.2:

AI tools are a great opportunity to empower law firms of all sizes to be able to respond to changing client requirements (including changing preferred channels of communication) and the increase in the amount of data generated at the level of society, including the ever-increasing amount and diversity of digital evidence which is created and which requires to be processed. From one perspective, this is an opportunity for a number of smaller firms to compete successfully with larger firms in sectors that they previously could simply not have served due to technical bottlenecks and lack of capacity. Based on more extensive automation, IT tools could provide a new chance for lawyers to improve further their workflows, to make their deliverables more consistent, and to increase the added value of their work. A human expert with a human understanding of the full context of a client is able to provide a much higher quality of service to clients, compared with automated assistance in a restricted set of piecemeal activities that can be fully automated by software, but where clients have to rely on their own judgement and self-help.

¹⁸⁶ *Ibidem*, 7.1

The amount of information to be processed by lawyers continues to increase every year: a more digitised society creates more and more data. A surge in regulatory instruments created or court cases published boosts the demand for the processing capacities of lawyers. All this intensifies competition among legal service providers on the basis of the capabilities of their IT tools. We have to remember that greater processing capacity for a lawyer does not necessarily mean either higher quality of work for the client, greater legal certainty or a more complete adherence to the rule of law. Similarly, it does not follow that lawyers will necessarily provide better services just because they have access to more regulatory instruments and cases. When lawyers try to meet expectations with regard to the increase of their data processing capabilities, they need to ensure that they act in compliance with the core principles of the European legal profession. These principles are not only enshrined in deontological rules, but also in the jurisprudence of the European Courts and the European Court of Human Rights. These principles include especially the independence of the lawyer, the obligation to avoid any conflicts of interest and the duty to respect professional secrecy. In this part, we highlight some risks to lawyers' compliance with their professional obligations when using AI tools. Some of these risks are more closely connected to the nature of the tools used or their delivery method (such as risks of using tools provided by third parties, risks of bias and lack of transparency in machine learning methods), and some risks arise from the perspective of the professional obligations of lawyers, such as the obligations of confidentiality or of competence.

¹⁸⁷ *Ibidem*, 7.2.1.

However, one has to be mindful that a predictive justice AI tool comes with its own considerable risks and special dangers compared to the other two levels. In informative and analytical justice, interpretation is undertaken by the user (the lawyer), and there is no risk of introducing further bias into the service provided (other than the bias already

- El riesgo específico para la privacidad es el de la violación de las obligaciones de secreto profesional. Hay que tener en cuenta asimismo que la normativa del país donde se aloje el proveedor no tiene por qué coincidir, en términos de protección del secreto abogado-cliente, con la del país en el que se presta el servicio. Esto podría implicar que una autoridad o un juez oficie al proveedor para que le dé acceso a la información introducida en la herramienta, lo que podría poner en riesgo el derecho fundamental a un juicio justo.
- Compromiso de la obligación profesional de competencia a costa de probar tecnologías novedosas que todavía no han sido suficientemente testadas, lo que se considera imprudente por parte del Consejo¹⁸⁸.
- Riesgo de asumir más compromisos profesionales de los que se pueden prestar de forma apropiada.
- Riesgo para la independencia del abogado, al que ya nos hemos referido.

included in past cases, but that is rarely the responsibility of the lawyer). Today, lawyers are not expected to understand how text retrieval actually works and why certain results might have been excluded. If that missed result negatively affects the service provided by the lawyer and the outcome of the case, the lawyer will nonetheless bear full responsibility to the client according to current rules of deontology. But at the third level of legal analysis, due to its new functionalities, the lack of explainability and introduction of new bias could cause a problem, and therefore lawyers using such tools should be aware of the increase in such risks and take appropriate steps if necessary. Both the ethics guidelines of the European Commission⁵³ and the draft AI Act⁵⁴ define transparency requirements for such tools (even though AI tools are not currently considered as high risk so long as they are used by lawyers and not by judicial authorities⁵⁵)

¹⁸⁸ *Ibidem*, 7.3.1:

Risks to professional competence: dangers of trying out new technology Changes in technology may affect what questions are asked of lawyers and what answers lawyers should give (the knowledge base a lawyer works with), but these may also alter how lawyers operate their legal practices. Such changes may create a conflict between the expectations of some clients and the professional expectation of a competent approach by the lawyer. Clients' wishes and business development reasons could both create an incentive for lawyers to appear as a pioneer in trying out (and relying on) new technologies and also to «boldly give advice» in areas where no lawyer has done so before. As for the first aspect, lawyers are often expected to guess how society and courts will react to the impacts of a technical tool. Clients can exhibit overconfidence and shallow knowledge of a technology, regardless of whether they are inventors, professional investors or just private persons in the know. However, the professional expectation of competence for lawyers requires that they should not "take on a case which he or she is not competent to deal with".¹⁰⁹ Lawyers must act in accordance with this principle at all times, even if a client suggests relying instead on a technological tool. Continuing professional development for lawyers is based on the appropriate training of the lawyer, which in turn implies some previous accumulation of experience that is used for that training. The proper focus of a lawyer's field of work is the social impact of technology, and not on understanding how technology works and, based on that, to make predictions on what future impacts a piece of technology will have on society. Of course, there are many ways lawyers can provide appropriate advice to their clients regarding the results of innovative technologies, but sometimes the best advice a lawyer can give is to communicate clearly any uncertainty and to demonstrate how weak predictions may be, and how misleading past forecasts are. With regard to the internal working of a law firm, like in any company, the fear of missing out (FOMO) in the field of AI may increase risks related to insufficient professional competence...

La Federación Europea de Colegios de Abogados ha emitido siete *DIRECTRICES SOBRE CÓMO DEBE APROVECHAR LA ABOGACÍA LAS OPORTUNIDADES QUE OFRECEN LOS GRANDES MODELOS LINGÜÍSTICOS Y LA IA GENERATIVA* en su informe *Abogacía Europea en la era de ChatGPT*¹⁸⁹.

La primera directriz se refiere a la necesidad de que los abogados tengan un profundo conocimiento de la tecnología en que se basa la Inteligencia Artificial. Para ello, los profesionales deben dedicar tiempo al estudio de la misma. Aunque inicialmente ello suponga una inversión en tiempo y recursos de otro tipo, a la larga esos abogados se beneficiarán de las ventajas que ofrece esta herramienta de forma consciente y eficiente¹⁹⁰.

La segunda directriz se refiere a la adopción de decisiones realistas que reconozcan las limitaciones de la herramienta, pues su uso indiscriminado puede conducir a soluciones imprecisas o desactualizadas¹⁹¹.

La tercera directriz se refiere a la necesidad de que el abogado, a la hora de utilizarla, respete la normativa vigente sobre Inteligencia Artificial¹⁹².

¹⁸⁹ Abogacía europea en la Era de CHAT GPT. Junio 2023.

<https://www.fbe.org/wp-content/uploads/2023/07/Spanish-version-Abogacia-Europea-en-la-Era-de-ChatGPT.pdf>

(accedido el 05-10-2025)

¹⁹⁰ *Ibidem*, Directriz 1: *Entender la tecnología de IA generativa.*

“Asegúrese de que conoce a fondo la tecnología de inteligencia artificial generativa utilizada en su práctica jurídica. Dedique tiempo a familiarizarse con sus capacidades, funcionalidades y posibles implicaciones jurídicas. Este conocimiento le permitirá tomar decisiones informadas sobre el uso adecuado y responsable de la tecnología, determinando cuándo y cómo aprovecharla eficazmente en su profesión jurídica.”

La FBE añade que es esencial comprender tanto los riesgos asociados al aprendizaje automático —como el sesgo y la falta de interpretabilidad—, como las amenazas específicas de los grandes modelos lingüísticos y la IA generativa, incluidas las “alucinaciones”, es decir, la invención de información que puede no ajustarse a la realidad o a la exactitud de los hechos.

¹⁹¹ *Ibidem*, Directriz 2: *Reconocer las limitaciones y el contexto.*

“Comprender cómo los LLM y la GenAI pueden mejorar los procesos jurídicos y ofrecer soluciones más eficaces es crucial para mantenerse a la vanguardia del sector. Sin embargo, es importante tener en cuenta las limitaciones de esta tecnología. A pesar de sus avances, tiene restricciones inherentes. Hay que reconocer que los contenidos generados por IA no siempre son del todo precisos, completos o actualizados. Tenga siempre presente el objetivo principal de los grandes modelos lingüísticos, los riesgos de alucinaciones y los sesgos. Recuerde que los sistemas GenAI que sirven para conversar, por ejemplo, no son necesariamente adecuados para otras aplicaciones, incluidas las jurídicas. Verifique siempre el resultado generado por el sistema.”

¹⁹² *Ibidem*, Directriz 3: *Respetar las normas vigentes sobre el uso de la IA.*

“Es esencial que la abogacía europea conozca y respete las normas vigentes que regulan el uso de la IA. Esto incluye mantenerse informado sobre la Ley de Inteligencia Artificial de la UE, actualmente en proceso legislativo. Además,

La cuarta directriz alude a la necesaria conservación de los conocimientos y las destrezas que eran propios de los profesionales del Derecho antes de la eclosión de la Inteligencia Artificial. Esta nunca debe utilizarse como un elemento sustitutivo, sino complementario, de las decisiones típicamente humanas. Hay que actuar de manera que se puedan verificar los resultados de la herramienta mediante el análisis, la diligencia y el buen juicio de que debe hacer gala el buen profesional¹⁹³.

Otro elemento esencial en la práctica de la abogacía consiste en la relación de confidencialidad entre el abogado y el cliente. La Inteligencia Artificial puede ser un ámbito en el cual se transfieren datos confidenciales de los clientes hacia destinos o destinatarios no deseados. A ello se refiere la quinta Directriz¹⁹⁴.

La sexta Directriz está en relación con la anterior, pues alude a la garantía de la privacidad y la protección de datos de los clientes¹⁹⁵.

es crucial estar al tanto de cualquier posible normativa nacional o de las normas establecidas por los Colegios de la Abogacía, aunque no sean jurídicamente vinculantes. Los abogados y abogadas también deben revisar a fondo los términos de servicio proporcionados por los proveedores de GenAI para garantizar el cumplimiento de las normas éticas y legales. Es importante señalar que las normas sobre el uso de la LA están sujetas a evolución y cambios constantes (...)."

¹⁹³ *Ibidem*, Directriz 4: Complementar los conocimientos jurídicos.

"GenAI debe complementar sus competencias jurídicas en lugar de sustituirlas. Utilícela como una herramienta para agilizar los procesos en su despacho y mejorar la eficiencia de los servicios jurídicos. Sin embargo, recuerde que GenAI no debe sustituir el ejercicio de su juicio profesional, análisis jurídico y experiencia. Evite depender demasiado de los resultados de GenAI. Recuerde que, aunque puede ayudar a generar documentos jurídicos, no debe sustituir a la toma de decisiones humana. Actúe con cautela y verifique de forma independiente los resultados mediante análisis, investigación, comprobación de hechos y juicio profesional."

Además, la FBE advierte que "existe el riesgo de que algunas personas opten por no buscar asesoramiento jurídico porque consideren suficiente una 'consulta' con un chatbot basado en LLM", por lo que es fundamental informar a los clientes sobre los riesgos de confiar exclusivamente en tales herramientas.

¹⁹⁴ *Ibidem*, Directriz 5: Mantener la confidencialidad entre abogado y cliente.

"Salvaguarde la confidencialidad abogado-cliente al utilizar herramientas o plataformas de LA generativa. Tome las precauciones necesarias para proteger la confidencialidad de la información y las comunicaciones de los clientes. Asegúrese de que los sistemas de LA generativa que utiliza cuentan con las medidas de seguridad adecuadas. Aplique sus propias medidas de seguridad sólidas para salvaguardar los datos confidenciales y cumplir las obligaciones éticas y legales relacionadas con la confidencialidad de los clientes."

La FBE subraya que "la introducción directa o indirecta de datos de clientes como parte de una consulta en los sistemas GenAI puede no estar permitida debido al potencial tratamiento posterior de la información contenida en la consulta".

¹⁹⁵ *Ibidem*, Directriz 6: Garantizar la protección de datos y la privacidad.

Hace hincapié en la necesidad de una evaluación detallada de la conformidad de GenAI con el Reglamento General de Protección de Datos (RGPD), destacando la importancia de actuar con cautela, aplicar las medidas adecuadas y abordar los problemas de privacidad y los riesgos emergentes (...). Los abogados deben ser conscientes de que la introducción de datos personales en los sistemas GenAI requiere una base jurídica adecuada (...). Para garantizar la privacidad y la protección de los datos, es esencial aplicar medidas de seguridad sólidas que abarquen tanto la tecnología como los procesos.

Todos los anteriores principios están en línea con los riesgos a los que nos hemos referido y concuerdan con las recomendaciones de organismos similares de otros países.¹⁹⁶

Por ejemplo, la *Formal Opinion* 512 de la American Bar Association¹⁹⁷, que pone el foco en el deber de competencia (los letrados deben mantenerse actualizados sobre los beneficios y riesgos de la IA), la verificación independiente (no depender ciegamente del contenido generado), la protección de la confidencialidad y consentimiento informado del cliente si la herramienta aprende de sí misma, obligación de comunicación al cliente del uso de la IA si es relevante para los objetivos del cliente, el costo de los servicios, o si el cliente lo solicita, de manera que no se socave la relación de confianza (no cabe consentimiento genérico), establecimiento de políticas claras sobre el uso de estas herramientas en el despacho y capacitación del personal no legal y supervisión de proveedores externos que la utilicen.

Nos resulta llamativa la indicación de que los abogados cobren por el tiempo real dedicado y no por la velocidad aumentada por GAI, de forma que la facturación no resulte inflada. Tampoco se debe, a entender de la ABA, facturar al cliente por el propio aprendizaje del abogado sobre la tecnología. El coste del uso de las herramientas de Inteligencia Artificial puede considerarse como gastos razonables diferenciados de los gastos generales, resultando adecuado imputar al cliente, por ejemplo, los servicios de terceros por uso específico de la IA. Ahora bien, la ABA considera la evolución constante del estándar tecnológico, de manera que las herramientas se vuelvan más fiables y comunes, de modo que pueda esperarse que los abogados las usen para cumplir con su deber de competencia, de forma similar a como se normalizó el uso del correo electrónico o la investigación legal en Internet.

¹⁹⁶ *Ibidem*, Directriz 7: *Informar a los clientes y asumir responsabilidades*.

“Mantenga una comunicación transparente con sus clientes sobre el uso de GenAI en su práctica legal. Explique claramente el hecho de que la utiliza, así como la finalidad de dicho uso, beneficios, limitaciones y garantías, asegurándose de que los clientes entienden el papel de esta tecnología en sus asuntos legales. Especifique si se utiliza para la investigación jurídica o para tareas adicionales de oficina. Asuma la responsabilidad de los resultados y las consecuencias de la utilización de GenAI, teniendo en cuenta al mismo tiempo sus obligaciones profesionales y su posible responsabilidad.”

La FBE aconseja además *“esperar a que estén disponibles sistemas especializados diseñados para ayudar a la abogacía”* y *“evitar utilizar sistemas genéricos de IA generativa conversacional que puedan no ser adecuados para aplicaciones estrictamente jurídicas”*. Finalmente, exhorta a *“adoptar un enfoque responsable dentro del despacho, informar a los compañeros de las normas establecidas y velar por su cumplimiento”*.

¹⁹⁷ Stefan Nigam, Makhani Al-Karim, Reuben Miller, *“Is IA finally going to take our jobs? Meeting client”* IBA Association, (2024).

https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf

(accedido el 05-10-2025)

La Asociación Internacional de Colegios de Abogados (IBA por sus siglas en inglés) publicó en septiembre de 2024 un informe sobre Inteligencia Artificial y la profesión legal¹⁹⁸. Uno de los desafíos que menciona es el del riesgo de futura brecha de perfiles senior que se puede llegar a producir si las firmas se acomodan en el uso de la tecnología para suplir a los *juniors*, haciendo hincapié en que se siga formando a abogados jóvenes en el trabajo legal, aunque este pueda ser llevado a cabo por la IA¹⁹⁹, eso sí, debiendo contar a partir de ahora los nuevos perfiles legales con conocimientos tecnológicos sólidos complementarios a los jurídicos.

Por otra parte, esa misma cuestión puede enfocarse desde el lado opuesto. Existe también la posibilidad de que las firmas desarrollen paulatinamente o de forma rápida una exacerbada preferencia por los perfiles jóvenes y, hasta cierto punto, inexpertos en la praxis jurídica general, pero dotados de conocimientos notables en el campo de la Inteligencia Artificial. En ese caso, muchos abogados senior con una sólida base en el foro se verían preteridos en los despachos, con lo que ello supondría una pérdida de experiencia y profundidad a la hora de abordar los distintos asuntos.

Otro aspecto digno de ser tenido en cuenta a la hora de abordar estos nuevos retos consiste en la necesidad de una regulación en materia de lenguaje tecnológico para que este sea común entre los juristas, los legisladores y los ingenieros informáticos. Muchos de los términos que han saltado recientemente a las normas o las guías sobre Inteligencia Artificial no representan idealmente los mismos conceptos. Palabras como *riesgo*, *explicabilidad*, *transparencia algorítmica*, etc., ostentan semánticamente distintos significados dependiendo de los textos en los que se insertan. Ese es uno de los mayores obstáculos que se plantean a la hora de implementar los principios que se recogen en el *AI Act* y en los códigos deontológicos redactados en diferentes idiomas. Así pues, una de las principales tareas que deben acometer en adelante los colegios profesionales consiste en la elaboración de glosarios técnico-jurídicos y de métodos que permitan

¹⁹⁸ <https://www.ibanet.org/document?id=The-future-is-now-artificial-intelligence-legal-profession> (accedido el 06-10-2025)

The correlation between the use and development of different lawyer categories suggests that the adoption of AI will necessitate the creation of new roles or the evolution of existing ones. Also, training young lawyers beyond their use of AI represents a new challenge, with the risk of losing the expertise needed at more senior roles or directly losing talent if work were limited to the use of AI and the review of its outcome. Respondents reported various challenges in adopting AI, with larger firms showing more robust training and adoption programmes. These programmes emphasise data and AI literacy, ensuring that employees are well-equipped to use AI tools effectively.

¹⁹⁹ En el mencionado trabajo *Is LA finally going to take our jobs?...* la IBA sugiere que los jóvenes abogados se centren en tareas de mayor valor, como la ingeniería de *prompts* y la integración de la tecnología en los procesos tradicionales. Así, los abogados jóvenes pueden convertirse en el “*human-in-the-loop*” que asegure que los resultados obtenidos son correctos y salvaguarde las obligaciones profesionales y éticas del abogado.

https://www.ibanet.org/is-AI-finally-going-to-take-our-jobs?utm_source=chatgpt.com (accedido el 07-10-2025)

evaluar y comprobar efectivamente el cumplimiento de los fundamentos éticos estandarizados sin que los abogados deban llevar a cabo un profundo aprendizaje técnico previo al uso de las nuevas herramientas.

Otro papel estelar de los órganos colegiales consistirá en promover la creación de una normativa que ayude a rellenar el vacío que existe en materia de responsabilidad civil y penal que se pueda derivar del uso de la Inteligencia Artificial en el ejercicio de la abogacía, pues resulta todavía difícil distinguir entre los errores achacables a los abogados y los que se derivan de un defectuoso funcionamiento de los algoritmos en que basa sus soluciones la nueva herramienta.

Todavía se considera que la responsabilidad recae totalmente sobre el abogado debido al incipiente estado de la mencionada herramienta jurídica. Pero cuando se avance en esta tecnología y los sistemas puedan actuar con mayor autonomía, los proveedores de esas nuevas soluciones tecnológicas deberán ser corresponsables necesarios.

En caso de derivarse un daño importante para el cliente, incluso para sus derechos fundamentales, podrían darse supuestos de responsabilidad penal por el uso de estas herramientas sin el debido control. Una vez más, correspondería a los órganos colegiales adelantarse a esas posibles quiebras del sistema implementando ciertas cláusulas contractuales estandarizadas que puedan proteger simultáneamente al usuario y al profesional de la abogacía.

Aun reconociendo el alto valor de las nuevas guías de buenas prácticas, estas no deben considerarse solo instrumentos de obligado cumplimiento con carácter normativo, sino también como permanentes recordatorios de una ética que promueve en todo momento la responsabilidad de los profesionales en el uso de las tecnologías. Los valores de la prudencia, la transparencia o la reflexión que se reflejan en estas guías como característica necesaria del profesional del Derecho todavía son solamente intrínsecas al ser humano, de modo que los abogados reciben por medio de ellas una valiosa orientación a la hora de evitar conflictos. Por eso, a pesar de la rápida evolución de la Inteligencia Artificial, las guías se constituyen en una especie de normas-marco que trascienden la rapidez cotidiana y sientan unos principios orientadores que siempre deben iluminar a los profesionales en los momentos de duda. Para conseguir semejante efecto, las guías deben ser textos versátiles y revisables cada cierto tiempo, susceptibles, además, de propiciar la constante relación entre todos los actores que intervienen en el uso de la Inteligencia Artificial dentro del campo jurídico, ya sean las facultades de Derecho, los colegios profesionales o las empresas tecnológicas. Solo así el mundo del Derecho podrá instaurar unos nuevos fundamentos éticos dentro del emergente universo jurídico que se abre camino en la sociedad actual.

4. CONCLUSIONES

Nos encontramos ante un momento disruptivo para la profesión, con numerosas oportunidades para el tratamiento de información que en otros tiempos no habrían estado al alcance ni siquiera de los grandes despachos. La abrumadora oferta de herramientas y lo novedoso de las mismas implicará una enorme dedicación inicial de tiempo a la reingeniería del proceso jurídico, selección de proveedores, aprendizaje y, por supuesto, a la adaptación de los despachos.

Resulta innegable que aquel abogado que disponga de mejores herramientas y mejor domine su uso tendrá una ventaja diferencial. No obstante lo anterior, no parece cercano el día en el que se pueda delegar en un robot la tarea del asesoramiento o defensa jurídica, pues la diferencia, en definitiva, a igualdad de herramientas, seguirá radicando en el profesional humano, con una mayor base de experiencias vitales y mejor capacidad de raciocinio jurídico.

Como sucede en el ámbito de otras profesiones cualificadas, la sociedad, para bien o para peor, va cambiando la percepción del profesional del Derecho. Antes de la aparición de la informática como fuente ilimitada de información para el público en general, se le atribuía comúnmente al abogado el exclusivo dominio de las normas y también de su interpretación. Sin embargo, hoy en día cualquiera puede acceder a la información jurídica que se precisa y solicitar a la Inteligencia Artificial que genere informes o escritos jurídicos aparentemente rigurosos y satisfactorios. Por ello, es previsible que en el futuro el valor del profesional se fundamentará en la fiabilidad contrastada de su trabajo y en la aportación del elemento humano en lo que se refiere a la valoración de elementos como la equidad y el rigor dentro del contexto que le planteen sus clientes. Ello alude a la especificidad de la casuística. Los usuarios deben sentirse seguros y únicos en su pretensión de obtener una solución jurídica ajustada a Derecho y satisfactoria para sus intereses, fuera de la sensación de adocenamiento que puede derivarse de saberse sometidos a los dictámenes de la “máquina”. Elementos tales como la capacidad para diferenciar lo justo de lo injusto, lo admisible éticamente, lo técnicamente correcto y ajustado a Derecho, etc. se convertirán en la verdadera esencia del jurista en ejercicio. No será, como antes, valorado por la cantidad de conocimientos jurídicos almacenados en su cerebro, sino por la demostrada percepción correcta, ética, eficiente y responsable de la justicia, así como por su capacidad para interpretar adecuadamente cada supuesto de hecho. Eso tiene que ver con la idea de que no hay dos supuestos de hecho idénticos cuando se examina el mundo jurídico desde el punto de vista humano. Esa identidad única del caso no se garantiza desde la herramienta de la Inteligencia Artificial.

En la sociedad actual, donde el materialismo y la celeridad en la consecución de resultados rentables se valoran al alza, el uso de la Inteligencia Artificial corre el peligro de considerarse estrictamente desde un punto de vista económico. Ahí es donde el abogado debe intervenir en el ámbito del medio social y cultural en que le ha

correspondido desarrollar su profesión. En primer lugar, el profesional del Derecho debería examinar cuál es la percepción que tiene de sí mismo. Si antes se erigía en el único intérprete cualificado del Derecho, la eclosión de sistemas generadores de posibles soluciones jurídicas lo ha desbancado de ese puesto privilegiado y exclusivo que ostentaba en la sociedad tradicional. Su actividad profesional solo tendrá verdadera razón de ser en tanto en cuanto sea capaz de ofrecer garantías sobre la legitimidad (que no legalidad), la calidad de sus razonamientos y la veracidad de sus resultados. Se trata, en suma, de un profundo cambio en el modelo tradicional que puede resultar nocivo si no se ponen en marcha mecanismos que garanticen la fiabilidad de las instituciones profesionales, la ética humanizadora de los procesos y la estandarización de unos límites en el uso de las nuevas tecnologías en lo tocante al asesoramiento y defensa de los usuarios.

La abogacía conservará su centralidad si integra las nuevas herramientas con un férreo control de la deontología profesional. El ser humano, desnudo de cualquier herramienta tecnológica, debe estar presente para garantizar que, en todo caso, prevalecen el mejor servicio al cliente y la tutela judicial efectiva.

BIBLIOGRAFÍA

Chaves, José Manuel, “Convencer al juez citando jurisprudencia con lealtad”, *Blog de Comunicación, Marketing y Coaching del Consejo General de la Abogacía Española*, (2024).

Nigam, Stefan; Al-Karim, Makhani; Miller, Reuben, “Is IA finally going to take our jobs? Meeting client” *IBA Association*, (2024).

Yu, Fangyi; Quartey, Lee; Shilder, Frank, “ Legal Prompting: Teaching a Language Model to Think Like a Lawyer”, *ArXiv.org* (2022).

FUENTES JURÍDICAS Y WEBGRAFÍA

Real Decreto 135/2021, de 2 de marzo, por el que se aprueba el Estatuto General de la Abogacía Española. Boletín Oficial del Estado.

Ley Orgánica 6/1985, de 1 de julio, del Poder Judicial. Boletín Oficial del Estado.

Código Deontológico de la Abogacía Española aprobado por el Pleno del Consejo general de la Abogacía española el 6 de marzo de 2019. Boletín Oficial del Estado.

Guía ICAM de Buenas Prácticas para el Uso de la Inteligencia Artificial en la Abogacía. Web del Ilustre Colegio de la Abogacía de Madrid. web.icam.es

Guía sobre el uso de herramientas basadas en IA por abogados y despachos de abogados en la UE. Web del CCBE-ELF. www.ccbe.eu

Abogacía Europea en la era de ChatGPT. Web de la FBE.

Formal Opinion 512 de la American Bar Association. Web de la ABA. www.americanbar.org

The Future Is Now: Artificial Intelligence And The Legal Profession. www.ibanet.org

<https://legamy.com/noticia/inteligencia-artificial-aprueba-examen-para-ser-abogado>

[AI Hallucination Cases Database – Damien Charlotin](#)

<https://www.abogacia.es/publicaciones/blogs/blog-comunicacion...uridicos/convencer-al-juez-citando-jurisprudencia-con-lealtad/>

<https://www.linkedin.com/pulse/notificaciones-certificadas-la-desconexi%C3%B3n-uxxrf/>

<https://www.boe.es/buscar/act.php?id=BOE-A-2021-4568>

<https://www.abogacia.es/wp-content/uploads/2019/05/Codigo-Deontologico-2019.pdf>

https://www.unotv.com/tecnologia/la-ia-aprende-a-mentir-manip...res-estos-casos-alertan-a-las-personas/?utm_source=chatgpt.com

<https://www.abogacia.es/actualidad/noticias/participa-en-la-e...sta-sobre-el-uso-de-la-inteligencia-artificial-en-la-abogacia/>

https://web.icam.es/wp-content/uploads/2025/10/Guia-ICAM_IA_2025_2.pdf

https://www.ccbe.eu/fr/actions/evenements/ai4lawyers/?utm_source=chatgpt.com

<https://www.fbe.org/wp-content/uploads/2023/07/Spanish-version-Abogacia-Europea-en-la-Era-de-ChatGPT.pdf>

https://www.americanbar.org/content/dam/aba/administrative/pr...onal_responsibility/ethics-opinions/aba-formal-opinion-512.pdf

<https://www.ibanet.org/document?id=The-future-is-now-artificial-intelligence-legal-profession>

https://www.ibanet.org/is-AI-finally-going-to-take-our-jobs?utm_source=chatgpt.com

<https://www.abogacia.es/wp-content/uploads/2012/06/codigodeontologico.pdf>

<https://www.linkedin.com/pulse/herramientas-de-ia-segura-la-ce-gu%C3%ADa-idjyf/>



CAPÍTULO 6.

IDENTIFICACIÓN BIOMÉTRICA REMOTA EN EL ÁMBITO POLICIAL, ¿CÓMO AFECTA A LA PROTECCIÓN DE DATOS DE CARÁCTER PERSONAL? ²⁰⁰

CRISTINA DOMINGO JARAMILLO

Investigadora postdoctoral en Derecho Penal²⁰¹

Universidad de Granada

cdomingo@ugr.es

Resumen

El Reglamento Europeo de Inteligencia Artificial regula los sistemas de identificación biométrica remota en espacios públicos con fines de cumplimiento del Derecho, tanto los que actúan en tiempo real como en diferido. Los primeros están prohibidos porque el impacto que tienen en los derechos y libertades se considera intolerable, salvo en tres circunstancias excepcionales. Por su parte, los segundos se clasifican de alto riesgo, motivo por el que pueden utilizarse cuando se cumplan los requisitos generales establecidos en el Reglamento para ese tipo de sistemas. A pesar de la distinción, ambos interfieren de forma considerable en los derechos, sobre todo en el de protección de datos de carácter personal, en tanto la información que utilizan los mecanismos es de esta tipología y especialmente sensibles. El presente capítulo aborda algunas de las cuestiones más controvertidas que pivotan en torno a la afectación de este derecho por los sistemas de identificación biométrica remota, concluyendo que el impacto es muy elevado, tal vez excesivo. Por este motivo, hubiese sido más conveniente que no se incluyesen en el Reglamento.

Palabras clave

Biometría, datos personales, seguridad, prevención y persecución de la delincuencia, Reglamento de Inteligencia Artificial.

²⁰⁰ El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

²⁰¹ Investigadora postdoctoral en el Departamento de Derecho Penal de la Universidad de Granada. Contrato DGP_POST_2024_00899, financiado por la Junta de Andalucía/CUII y por el FSE+. Doctora en Criminología por la Universidad de Granada y Doctora en el Programa de Doctorado Internacional *Law, Science and Technology* por la Universidad de Bolonia. Entre las principales líneas de investigación destaca la relativa a la Inteligencia Artificial en materia de prevención y comisión delictiva.

Abstract

The European Artificial Intelligence Act now regulates remote biometric identification systems in public places for law enforcement purposes, both real-time and delayed ones. The former are banned because their impact on rights and freedoms is seen as unacceptable, except in three exceptions. The latter are classified as high risk, which is why they can be used when the general requirements established in the Regulation for this type of system are met. Despite the distinction, both interfere considerably in rights, especially in the protection of personal data, as the information used by the mechanisms is of this type and particularly sensitive. This chapter addresses some of the most controversial issues surrounding the impact of remote biometric identification systems on this right, concluding that the impact is very high, perhaps excessive. For this reason, it would have been more appropriate not to include them in the AI Act.

Key words

Biometrics, personal data, security, crime prevention and prosecution, AI Act.

Sumario: INTRODUCCIÓN. 1. DATOS BIOMÉTRICOS COMO DATOS PERSONALES ESPECIALMENTE SENSIBLES 1.1. Distinción de los datos biométricos en función de la intrusividad en su recolección. 1.2. Disponibilidad o visibilidad del rostro en el espacio público. 2. IMPACTO DE LA BIOMETRÍA EN LA PROTECCIÓN DE DATOS PERSONALES: ESPECIALMENTE EL RECONOCIMIENTO FACIAL. 2.1. La protección de datos personales en los distintos momentos en los que opera el sistema. 2.2. Fuentes de las que se extrae la información: ¿Qué ocurre con las imágenes publicadas en redes sociales? 3. CONCLUSIONES. APÉNDICE BIBLIOGRÁFICO

INTRODUCCIÓN

El Reglamento Europeo de Inteligencia Artificial²⁰² (a partir de ahora, RIA) se ha planteado con un enfoque antropocéntrico, ético, sostenible y respetuoso con los

²⁰² 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° 168/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE y (UE) 2020/1828 (Reglamento de Inteligencia Artificial).

El RIA entró en vigor a los 20 días de su publicación en el Diario Oficial de la Unión Europea pero su contenido se hace aplicable de forma progresiva. Por lo general, a partir del 2 de agosto de 2026, salvo los Capítulos I y II que lo son desde el 2 de febrero de 2025; el Capítulo III, la sección 4 del Capítulo V, el Capítulo VII y el Capítulo XII y el art. 78 lo son a partir del 2 de agosto de 2025, a excepción del art. 101; y, el art. 6, apartado 1, y las obligaciones del RIA lo serán a partir del 2 agosto de 2027.

De este modo, a los efectos de nuestro estudio, señalar que la parte relativa a las prácticas prohibidas (entre las que se encuentran los SIBR en tiempo real en espacios públicos con fines de cumplimiento del Derecho) entró en vigor en febrero de 2025, mientras que todo el capítulo relativo a las prácticas de

valores y derechos fundamentales. De este modo, sus principales objetivos se orientan a garantizar el desarrollo de la IA salvaguardando unas condiciones equitativas entre los operadores del mercado interno y garantizar un estándar elevado de tutela de los derechos. Por ello, pretende asegurar la puesta en marcha de sistemas de IA fiables y que operen teniendo como base el respeto a los susodichos derechos²⁰³, intentando proteger a las personas de los peligros derivados del uso de tales mecanismos²⁰⁴.

En el RIA se introducen normas específicas para proteger a las personas en lo relativo al tratamiento de datos de carácter personal que restringen el uso de la IA cuando se utilice con fines específicos como el de aplicación del Derecho, la realización de evaluaciones de riesgo de personas físicas con objetivos, nuevamente, de aplicación del Derecho y para la categorización biométrica. En este punto, es de destacar que el enfoque basado en el respeto a los derechos fundamentales es esencial porque los instrumentos inteligentes, aunque presentan muchos beneficios en ámbitos muy diversos, entrañan al mismo tiempo peligros en los intereses públicos y derechos de la ciudadanía²⁰⁵.

Por ese motivo, realiza una clasificación de los sistemas de IA basada en el riesgo que entrañan en los derechos y libertades. Así, prohíbe los instrumentos considerados de riesgo inaceptable y clasifica otras prácticas de alto riesgo. Para estas últimas establece una disciplina estricta fijando requisitos específicos de uso y las obligaciones que los proveedores y operadores de los sistemas deben tener en consideración, así como observaciones de transparencia. En un nivel inferior, los sistemas de riesgo limitado únicamente tienen que seguir obligaciones de señalización para advertir a los usuarios que están interactuando con un mecanismo inteligente, llegando a los sistemas de riesgo mínimo que no presentan restricción alguna, como los servicios de traducción.

Entre las herramientas prohibidas²⁰⁶ se encuentran los sistemas de identificación biométrica remota (en adelante, SIBR) en tiempo real utilizados en espacios

alto riesgo, entre las que se encuentran todos los sistemas de identificación biométrica y, más concretamente los SIBR en diferido, no lo hizo hasta el 2 de agosto del mismo año.

²⁰³ En relación a lo expuesto, hemos de concretar que el RIA se centra en regular los sistemas y aplicaciones específicas de IA y no la IA en su conjunto. Finocchiaro, Giusella. *Intelligenza artificiale. Quali regole?*, Il Mulino, Bologna, 2023, p. 43.

²⁰⁴ Donatti, Filippo. “Diritti fondamentali e algoritmi nella proposta di Regolamento sull’Intelligenza Artificiale”, en Pajno, A. et al. (Dirs.), *Intelligenza Artificiale e diritto: una rivoluzione?*, il Mulino, Bologna, 2022, p. 123.

²⁰⁵ Entre ellos, la no discriminación y equidad, protección de datos y el derecho a la vida privada y familiar, las libertades de expresión e información, de reunión y asociación, así como de pensamiento, conciencia y religión, el derecho a recibir un remedio efectivo y un juicio justo, y la presunción de inocencia y el derecho de defensa.

²⁰⁶ La Comisión Europea, en un intento por aportar claridad legal y ofrecer información sobre estas prácticas prohibidas, aprobó el día 4 de febrero de 2025 unas Directrices. Con ello pretende asegurar consistencia, efectividad y aplicación uniforme. Comisión Europea. *Communication to the Commission*.

públicos con fines de cumplimiento del Derecho, mientras que los que operan en diferido y los demás mecanismos basados en la identificación biométrica son simplemente considerados de alto riesgo.

Antes de nada, es necesario concretar qué se entiende por datos biométricos, en tanto son el elemento fundamental con el que actúan estos sistemas. Siguiendo la definición establecida en el artículo 3.34 RIA son “los datos personales obtenidos a partir de un tratamiento técnico específico, relativos a las características físicas, fisiológicas o conductuales de una persona física, como imágenes faciales o datos dactiloscópicos”²⁰⁷. Por tanto, la biometría es la técnica de procesamiento mecánico de las características de los seres humanos que permite identificar o verificar la identidad de un individuo.

De esto último se infiere que la biometría puede utilizarse tanto en labores de identificación como de verificación de la identidad. El RIA excluye la segunda del término identificación biométrica y, por ende, del de sistemas de identificación biométrica remota²⁰⁸. Ello porque la verificación se realiza de uno-a-uno y se basa en la comparación de los datos biométricos de una persona con los datos biométricos que ella misma ha facilitado previamente (artículo 3.36 RIA). El objetivo es determinar si una persona es quien dice ser y, con ello permitirle, por ejemplo, el acceso a determinados espacios, ya sean físicos o virtuales. Asimismo, se inserta en los dispositivos telefónicos que pueden desbloquearse con los datos faciales, del iris o de las huellas dactilares del titular.

Por el contrario, la identificación es uno-a-muchos y, tal como establece el artículo 3.35 RIA, supone el reconocimiento automatizado de características físicas, fisiológicas, conductuales o psicológicas de los seres humanos para determinar la identidad de una persona física, a través de la comparación de sus datos biométricos con los de personas que se encuentran almacenados en una base de datos. A la vista de las diferencias entre ambas funciones, es adecuada la distinción realizada en el Reglamento, dado que, además, la verificación no conlleva el mismo nivel de riesgo en los derechos que la identificación.

El uso de la biometría y, más específicamente, la dedicada a la identificación, se está extendiendo en el ámbito policial con el objetivo de mantener la seguridad. Así, por ejemplo, se recogen las huellas dactilares en bases de datos y se incluye la fotografía del rostro en los documentos nacionales de identidad y los pasaportes.

Approval of the content of the draft Communication from the Commission – Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act), Bruselas, 4 de febrero de 2025. C(2025) 888 final.

²⁰⁷ Definición que reproduce de forma literal la contenida en los artículos 4.14 del Reglamento General de Protección de Datos y 3.13 de la Directiva de Protección de Datos en el ámbito penal.

²⁰⁸ Véanse los Considerandos 15, 17 y 54 del RIA.

No solo eso, sino que, con el avance de la tecnología, se están empezando a poner en marcha mecanismos que permiten reconocer a distancia y sin participación activa, a personas físicas por sus datos biométricos, los cuales se comparan con todos los contenidos en una base de datos de referencia. Estos son los denominados sistemas de identificación biométrica remota, tal como recoge el art. 4.41 del RIA. Entre las distintas técnicas, destaca la tecnología de reconocimiento facial.

El despliegue de esta última plantea una serie de controversias, pues se inserta en los circuitos cerrados de televisión (CCTV), en las cámaras de videovigilancia o se utiliza en grabaciones previas en el transcurso de una investigación policial. Permite la identificación unívoca de individuos en el espacio público a distancia y en muy poco tiempo (a veces, segundos en el caso de los que actúan en tiempo real), incluso entre una gran multitud, utilizando la información del rostro. Este último es el dato personal por antonomasia, especialmente visible y difícil de esconder, lo cual convierte al reconocimiento facial en una poderosa herramienta en materia de investigación policial y, mal utilizada -en casos extremos- puede derivar en una situación de vigilancia indiscriminada al permitir la monitorización de los movimientos de determinados sujetos.

En vista de las características de estos sistemas, la privacidad y la protección de datos de carácter personal son solo dos de los derechos más afectados por su uso. En un intento por contenerlo, el RIA incluye los SIBR en tiempo real entre las prácticas prohibidas en la letra h) del artículo 5.1.

No obstante, establece al mismo tiempo tres situaciones en las que su despliegue sí está permitido cuando se considere estrictamente necesario y se observen una serie de condiciones. Dichas situaciones son: i) búsqueda selectiva de víctimas de secuestro, trata de seres humanos o explotación sexual y personas desaparecidas; ii) prevención de una amenaza específica, importante e inminente para la vida o la seguridad física de las personas o de una amenaza real y actual o real y previsible de atentado terrorista; y iii) localización o identificación de un sospechoso de haber cometido un delito para llevar a cabo una investigación o enjuiciamiento penal o de ejecutar una sanción penal por alguno de los delitos listados en el anexo II (entre los que se encuentran el terrorismo, la trata, la explotación sexual, etc.) y que esté castigado en el Estado miembro en cuestión con una pena o medida de seguridad privativa de libertad que tenga una duración máxima de al menos cuatro años.

La preocupación por contener la expansión de los SIBR en tiempo real no alcanza a los sistemas en diferido. Estos últimos no están sujetos a las restricciones de los anteriores y puede recurrirse a ellos para investigar cualquier delito. En dicha línea, el artículo 26.10 RIA concreta en su párrafo primero que la finalidad será la búsqueda selectiva de una persona sospechosa o condenada por haber cometido un delito y que cada uso del sistema deberá limitarse a lo estrictamente necesario para la investigación de un concreto ilícito penal, sin especificar cuál. Continúa el mismo precepto en su

párrafo tercero señalando que no podrá utilizarse de forma indiscriminada sin relación con un delito, un proceso penal, o amenaza real y actual o real y previsible de delito, o con la búsqueda de una persona desaparecida determinada.

Dejando a un lado el estudio de las previsiones del RIA para cada uno de los instrumentos señalados, pues ello escapa en mucho de nuestro objeto de estudio, a continuación se expondrán los principales riesgos o cuestiones controvertidas que surgen en relación con el derecho a la protección de datos de carácter personal por el uso policial de los SIBR en el espacio público, especialmente con los mecanismos de reconocimiento facial, dado que, como señalé en líneas superiores son los que más problemas plantean²⁰⁹.

²⁰⁹ De hecho, la inclusión de estos mecanismos en la forma de SIBR en el RIA ha sido muy controvertida. Por ese motivo, la propuesta inicial del mes de abril de 2021 fue muy criticada. Véase, Christakis, Theodore. et al. *Facial Recognition in the Draft European AI Regulation: Final Report on the High-Level Workshop Held on April 26, 2021*, AI-Regulation.com, May 27th, 2021. Recurso electrónico obtenido a través de la Web: <https://ai-regulation.com/facial-recognition-in-the-draft-european-ai-regulation-final-report-on-the-high-level-workshop-held-on-april-26-2021/>, [última consulta el día 12 de noviembre de 2025], pp. 1-16.

En este sentido, distintas instituciones y organismos internacionales consideraron necesaria una moratoria en su uso. Entre otros, el Alto Comisionado de Naciones Unidas para los Derechos Humanos en *Impact of new technologies on the promotion and protection of human rights in the contexts of assemblies, including peaceful protests*, 2020. A/HRC/44/24, p. 14. Es más, solicita a las fuerzas del orden que se abstengan de usar la tecnología de reconocimiento facial para identificar a los asistentes a una reunión pacífica; y el Supervisor Europeo de Protección de Datos en *Artificial Intelligence Act: a welcomed initiative, but ban on remote biometric identification in public space is necessary*. Recurso electrónico obtenido a través de la Web: https://www.edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative_en, [última consulta el día 12 de noviembre de 2025].

No solo eso, sino que este mismo, junto con el Comité Europeo de Protección de Datos fueron más allá y demandaron una prohibición general de cualquier uso de la IA con estos fines: CEPD-SEPD. *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, 18 June 2021. Recurso electrónico obtenido a través de la Web: https://www.edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf, [última consulta el día 11 de noviembre de 2025], pp. 11-12. También recomiendan que su uso se prohíba en espacios virtuales, así como el de los sistemas de categorización de individuos a través de la biometría.

Además, en el ámbito del Consejo de Europa, el *Ad hoc Committee on Artificial Intelligence* (CAHAI) recomendó la moratoria parcial o prohibición de los sistemas de IA calificados de riesgo inaceptable, entre los que se encontraban los mecanismos basados en la identificación biométrica, sobre todo si con ellos se corre el riesgo de vigilancia masiva. Véase el documento *Possible elements of a legal framework on Artificial Intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law*, Strasbourg, December 3, 2021. CAHAI(2021)09rev, p. 5.

1. DATOS BIOMÉTRICOS COMO DATOS PERSONALES ESPECIALMENTE SENSIBLES

Los datos biométricos, desde el momento en el que permiten distinguir a un individuo de los demás²¹⁰ y, por ende, identificarlo de forma unívoca, son una categoría especial de datos personales, especialmente sensibles. En dicho sentido, el Supervisor Europeo de Protección de Datos (SEPD) en algunas Opiniones los califica como un tipo de información de naturaleza inherentemente sensible, motivo por el cual han de tener salvaguardas concretas²¹¹. Además, hace hincapié en la calificación del rostro y las huellas dactilares como una categoría de dato sensible, debiéndose por ello limitar su procesamiento a lo estrictamente necesario para alcanzar los objetivos propuestos²¹². Siguiendo al anterior, el Abogado General Mengozzi, en el *Caso C-291/12*, sostuvo el “carácter sensible por definición de los datos biométricos”²¹³.

Esta tendencia interpretativa no es estable y así, el Grupo de Trabajo del Artículo 29 (en adelante, A29 WP) en el Dictamen 3/2012 sobre la evolución de las tecnologías biométricas, plantea que algunos datos biométricos, como la imagen del rostro, pueden revelar información sensible, como la raza o la etnia y permiten el seguimiento, rastreo o elaboración de perfiles, por lo que el impacto a la intimidad y la protección de datos de carácter personal es elevado²¹⁴. En tal línea, el A29 WP no reconoce que los datos biométricos sean *per se* un tipo de información sensible, sino que son susceptibles de revelarla.

Idéntica interpretación realiza YUE LUI quien, basándose en una decisión de la Autoridad de Protección de Datos Noruega sobre el uso de las imágenes de CCTV en autobuses, explica que las imágenes del rostro tienen que procesarse para que sean capaces de identificar a alguien. De este modo, sostiene que una imagen por sí misma no puede considerarse “identificable”, no es información personal por sí misma, sino portadora de la misma. Concluye aseverando que los datos biométricos no pueden

²¹⁰ Lynch, Nessa. “Facial Recognition Technology in Policing and Security-Case Studies in Regulation”, *Laws*, vol. 13, n.º. 35, 2024, p. 2.

²¹¹ Así, por ejemplo, *Opinions of 19 October 2005 on three Proposals regarding the Second Generation Schengen Information System (SIS II) (Com(2005) 236 final y COM(2005) 237 final)*, OJ C 91, 19.4.2006, p. 43; of 18 July 2013, *on the Proposals for a Regulation establishing an Entry/Exit System (EES) and a Regulation establishing a Registered Traveller Programme (RTP)*, p. 7; 3/2016 of 13 April 2016, *on the Exchange of information on third country nationals as regards the European Criminal Records Information System (ECRIS)*, p. 8.

²¹² *Opinion of 07/2016 on the First reform package on the Common European Asylum System (Eurodac, EASO and Dublin Regulations)*, p. 9; *Opinion 7/2018 of 10 August 2018, on the Proposal for a Regulation strengthening the security of identity cards of Union citizens and other documents*, p. 9.

²¹³ *Opinion of Advocate General Mengozzi of 13 June 2013*, Case C-291/12. Michael Schwarz v. Stadt Bochum. ECLI:EU:C.2013:401, p. 14.

²¹⁴ GRUPO DE TRABAJO DEL ARTÍCULO 29. *Dictamen 3/2012 sobre la evolución de las tecnologías biométricas*, 00720/12/EN WP103, 27 de abril 2012, p. 3.

considerarse datos sensibles a menos que se adopten algunas medidas técnicas para extraer información. Por ende, para este autor los datos biométricos son solamente personales, adquiriendo la condición de sensibles cuando son procesados técnicamente²¹⁵.

La definición ofrecida por el RGPD en su artículo 9.1 y la Directiva de Protección de Datos en el ámbito penal en el 10, al entender como dato biométrico aquel dirigido a identificar de manera unívoca a una persona, parece que excluyen las plantillas biométricas y las fotografías insertas en bases de datos de tal definición, hasta que no sean procesadas de forma que permitan esa identificación. En tal sentido, parece que distinguen entre los datos personales relativos a las características biológicas o conductuales (p.ej. las imágenes) por un lado y, datos biométricos que requieren de un procesamiento técnico específico, por otro²¹⁶. Más aún cuando el Considerando 51 del RGPD excluye explícitamente las fotografías del término datos biométricos hasta que no se haya producido el susodicho procesamiento.

No obstante, tal como afirma SUMER, si acudimos a las definiciones técnicas que entienden todo procesamiento -incluso el inicial- de características físicas o conductuales que permiten identificar individuos como datos biométricos, las imágenes incluidas en bases de datos adquieren tal consideración, dado que ya su inclusión en dichas bases de datos es la primera fase del procedimiento para tratar técnicamente esta información²¹⁷. Igualmente, hemos de tener en cuenta que una vez que esa información ha sido registrada, el titular pierde el control de su uso futuro, desconociendo cuál será, a lo que se añade que la naturaleza del dato, personal, permite la identificación al ser susceptible de comparación biométrica²¹⁸.

Si ya algunos datos comunes como los hábitos de alimentación que, a priori, podrían considerarse inocuos, desde el momento en el que pueden revelar cuestiones

²¹⁵ Yue Liu, Nancy. *Bio-Privacy. Privacy Regulations and the Challenge of Biometrics*, Routledge, Oxon, Abingdon, 2012, p. 120. De forma similar, Jasserand no entiende que los datos biométricos tengan naturaleza sensible *per se*, pues para ello tienen que ser capaces de identificar de forma única a una persona y esto dependerá de otras muchas circunstancias ajenas a la característica que se pretenden medir. Por ejemplo, en el caso del rostro, las imágenes son aptas para la identificación si algunos elementos como la luz, localización, exposición o resolución de la cámara permiten que la imagen sea reconocible. Jasserand, Catherine. “Legal Nature of Biometric Data: From ‘Generic’ Personal Data to Sensitive Data”, *European Data Protection Law Review*, vol. 2, 2016, pp. 307-308.

²¹⁶ En la misma línea, Kindt, Els.J. “Having yes, using no? About the new legal regime for biometric data”, *Computer Law & Security Review*, vol. 34, 2018, pp. 531-532.

²¹⁷ Sumer, Bilgesu. “When do the images of biometric characteristics qualify as special categories of data under the GDPR?: A systematic approach to biometric data processing”, *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2022, p. 2.

²¹⁸ Kindt, Els.J. “Having yes, using no?...”, cit., p. 532.

personales como las confesiones religiosas de los individuos, son datos sensibles²¹⁹, más todavía ha de adjudicarse a la cara, al iris o a las huellas dactilares la cualidad de dato especialmente sensible porque, como decimos, permite la identificación unívoca de individuos, aun cuando no han sido procesados técnicamente. Por tanto, deben estar sometidos a especiales garantías para salvaguardar adecuadamente los derechos y libertades ciudadanas, puestos en peligro por el uso policial de los mismos en labores de investigación.

1.1. Distinción de los datos biométricos en función de la intrusividad en su recolección

La recogida de datos biométricos puede ser más o menos intrusiva, dado que en algunos casos se hace a distancia, como ocurre con el rostro, frecuentemente sin el conocimiento ni consentimiento del titular²²⁰; mientras que otros, como sucede con las huellas dactilares, requieren mayor participación del titular. La intrusividad no solo depende de la forma de recolección, sino también de la sensibilidad de la información²²¹.

Como los SIBR actúan normalmente a distancia y sin la participación activa del individuo, se entiende que son menos intrusivos, sobre todo en el caso del rostro. En este último, de forma significativa, por el hecho de que su identificación automatizada se realiza de la misma forma que lo hacen los seres humanos, es decir, observando los rasgos faciales²²², no siendo necesario contacto físico para capturar la imagen²²³.

²¹⁹ Stitilis, Darius. & Laurinaitis, Marius. "Treatment of biometrically processed personal data: Problem of uniform practice under EU personal data protection law", *Computer Law & Security Review*, vol. 33, 2017, p. 624.

²²⁰ Smith, Marcus. & Miller, Seumas. *Biometric identification, Law and Ethics*, Springer, 2021, p. 22.

²²¹ En dicho sentido, el TEDH en el *Caso P.N. v. Alemania*, de 11 de junio de 2020, considera que la recolección y conservación de datos de identificación en causas criminales (fotografías, huellas digitales y palmares) son menos intrusivos en el derecho al respeto a la vida privada que la conservación de perfiles de ADN, pues estos últimos contienen información más sensible.

²²² Gates, Kelly.A. *Our biometric future: Facial recognition technology and the culture of surveillance*. Vol. 2, New York University Press, 2011, pp. 45-46. El autor expone que el escáner de retina es una técnica que causa especial desconcierto en los usuarios. Ello porque es difícil usar, requiere cuidado para alinear el ojo con el fin de capturar una imagen adecuada de los vasos sanguíneos del globo ocular.

²²³ Niklas, Pablo. & Barrera, Gabriel. "Biometría aplicada a la seguridad pública", *Simposio Argentino sobre Tecnología y Sociedad (STS)-JAIIO 46*, 2017, p. 151. Además, señalan que la cara es de información pública porque muchos rostros se encuentran en documentos oficiales, la web y redes sociales.

Es más, ya los CCTV en general, sin contar con la tecnología de reconocimiento facial, se consideran una técnica menos invasiva porque capturan de forma discreta situaciones que suceden en la zona en la que actúan, además se observan como parte del fondo, no como el primer plano del control social. Goold, Benjamin. et al. "The banality of security. The curious case of surveillance cameras", *British Journal of Criminology*, vol. 53, n°. 6, 2013, p. 984. Los autores realizaron entrevistas a ciudadanos sobre su opinión acerca de los CCTV, de las que extrajeron lo que acaba de apuntarse, así como que las

Sin embargo, la *Court of Appeal (Civil Division) on appeal from the High Court of Justice Queen's Bench Division (Administrative Court) Cardiff District Registry* en la Sentencia *R (Bridges) v CCSWP and SSHD* (por un caso de reconocimiento facial que involucraba a la policía del Sur de Gales) observa dos similitudes entre la biometría del rostro y otros parámetros como las huellas dactilares, pues entiende que ambos pueden ser capturados sin necesidad de ningún tipo de muestreo íntimo y son una parte del cuerpo generalmente visible al público. Sin embargo, continúa señalando que, por el uso de los sistemas de reconocimiento facial, la biometría del rostro puede obtenerse sin el consentimiento o conocimiento del individuo ni usando la fuerza, procesándose a gran escala²²⁴.

En este caso, la integridad corporal está garantizada porque no se requiere ningún tipo de contacto personal pero sí entraña riesgos con relación al consentimiento informado y la privacidad. PURSHOUSE/CAMPBELL cuando analizan esta sentencia, sostienen que es erróneo considerar solamente como intrusión policial la de tipo físico, desde el momento que las tecnologías biométricas y/o algorítmicas facilitan el uso y almacenamiento de información personal sensible por parte de las autoridades públicas. Continúan exponiendo que la tecnología de reconocimiento facial combina formas físicas e informativas de vigilancia porque recopilan información del cuerpo de las personas, aun sin contacto físico y la descomponen en una estructura informativa que puede procesarse tiempo después. Por dicho motivo, concluyen que el contacto físico es un parámetro deficiente para medir el grado de intrusión de una medida de vigilancia²²⁵.

Por otro lado, consideramos que hay ciertos parámetros como las huellas dactilares que no son tan visibles a los demás como las imágenes del rostro, a no ser que exista un tratamiento específico que involucre, como ocurre con el ADN, al individuo implicado, quien tiene que ofrecer sus huellas dactilares o recogerse de un espacio determinado como la escena de un crimen. Por este motivo, el rostro es mucho más accesible a los demás. Asimismo, entendemos que su uso en labores de identificación es mucho más intrusivo y conlleva un mayor riesgo en los derechos, con especial incidencia en el de protección de datos de carácter personal, no por la intervención del titular sino porque la cara es la parte corporal más expuesta a la vista de los demás, difícil de ocultar, pudiendo ser reconocida a distancia²²⁶. Esto pone en situación de

cámaras de vigilancia no deberían molestar a nadie, salvo que sea un individuo que tenga una conciencia culpable o esté realizando actividades que preferiría mantener ocultas a la vista de los demás.

²²⁴ Raposo, Véra.Lucía. "(Do not) remember my face: uses of facial recognition technology in light of the general data protection regulation", *Information & Communications Technology Law*, vol. 32, n.º. 1, 2023, p. 48.

²²⁵ Purshouse, Joe. & Campbell, Liz. "Automated facial recognition and policing: A Bridge too far?", *Legal Studies*, vol. 42, 2022, pp. 222-223.

²²⁶ Dushi, Desara. "The use of facial recognition technology in EU law enforcement: Fundamental rights implications", *Global Campus of Human Rights. South East Europe*, Policy Briefs, 2020, p. 8.

indefensión al titular de los datos frente a quienes los utilizan en materia de vigilancia e investigación policial.

El reconocimiento facial en tiempo real (y también en diferido) es una tecnología que favorece la vigilancia masiva y, a diferencia de otras biometrías como las huellas dactilares, puede utilizarse para una vigilancia pasiva y general que no requiere el conocimiento, consentimiento ni la participación de las personas objeto de monitorización. Si el sistema identifica la imagen de un individuo, éste puede verse incluido en una lista utilizada para la identificación posterior, incluso no estando perseguido por la policía ni los tribunales de justicia.

A la situación que acabamos de plantear hay que añadir las bases de datos utilizadas. Al contrario de lo que ocurre con el ADN, los agentes no necesitan obtener plantillas adecuadas para el reconocimiento del rostro de bases de datos creadas *ad hoc* con este fin. Por el contrario, ya existen repositorios, como las fotografías de los carnés de conducir, documentos nacionales de identidad y pasaportes, que se integran en los dispositivos de reconocimiento facial. Además, las imágenes de delincuentes conocidos y sospechosos están disponibles para la policía, a lo que se une la recolección e inclusión en las bases de datos de imágenes de redes sociales sin la aquiescencia del titular²²⁷. Nuevamente, esto último tiene un gran impacto en la protección de datos de carácter personal, como más adelante tendremos lugar de poner de manifiesto.

Por ende, para concluir, la recolección y procesamiento de todo tipo de datos biométricos es intrusiva y conlleva una invasión en este derecho, ya sean tomados a distancia o de forma personal. Es más, el RIA en el concepto de identificación biométrica establecido en su Considerando 15, no realiza ninguna distinción entre los datos utilizados en dicha tarea y los engloba igualmente en este término. Por tanto, el tratamiento que realiza el Reglamento es idéntico, motivo por el que cabe concluir que

²²⁷ Couchman, Hannah. “Mass Surveillance on Our Streets: Facial Recognition and the Threat to Our Rights”, *European Human Rights Law Review*, n.º. 6, 2019, p. 585; y Smith, Marcus. & Miller, Seumas. *Biometric identification, Law and Ethics...*, cit., pp. 23-24. Los autores destacan el caso de Australia. En este país, desde 2017 se permite la utilización de reconocimiento facial. Su legislación posibilita a las agencias federales buscar y compartir imágenes de carnés de conducir, pasaportes, etc. Por otro lado, desde 2015 puede usarse el reconocimiento facial para la investigación de delitos importantes o ataques terroristas sin orden de búsqueda y captura incluso sin conocimiento de los interesados. Hay autores que consideran que las bases de datos que toman imágenes de redes sociales que el individuo ha publicado no merecen el mismo nivel de protección que otros datos sensibles. Sin embargo, desde el momento en el que el sujeto no haya otorgado su consentimiento e incluso desconozca que se están utilizando con otros fines de los que no es consciente, entendemos que deberían tratarse de forma idéntica a los demás. Jasserand, Catherine. “Massive Facial Database and the GDPR. The New Data Protection Rules Applicable to Research”, in Leenes, R. et al. (Eds.), *Data Protection and Privacy. The Internet of Bodies*, vol. I, Hart Publishing, 2018, p. 183.

todos comparten idénticas características. Asimismo, su recolección, almacenamiento y tratamiento es intrusivo, independientemente del tipo de información de que se trate.

1.2. Consecuencias de la disponibilidad o visibilidad del rostro en el espacio público

Como ya indicáramos en líneas superiores, la cara es el elemento fundamental de la identificación personal. Actualmente, con los sistemas automatizados, es posible tanto vincular la información del rostro con otra como la ideología política²²⁸, como transformar las expresiones faciales en datos, lo cual supera a todas luces la identificación tradicional, alcanzando al núcleo de la identidad personal²²⁹.

De este modo, con la biometría, el reconocimiento que inicialmente era un proceso intuitivo derivado de la interacción humana, se ha automatizado, convirtiendo las partes del cuerpo en meros datos. Por ese motivo, ha pasado a ser un objeto controlado por otro²³⁰ y cuya identidad se determina a través del lenguaje de las matemáticas.

Al ser una representación digital del ser humano, como los demás datos insertos en un ordenador, se pueden reproducir, transmitir, analizar y reutilizar fácilmente, mientras que los titulares tienen poco o ningún control sobre esos datos. Todo ello deriva en un proceso de cosificación o deshumanización de las personas que no ocurre con otros identificadores como los códigos personales insertos en algunos documentos como los pasaportes²³¹. Así las cosas, el titular del dato biométrico pierde el control sobre el mismo, quedando en manos de terceros. De modo que los datos biométricos impactan de forma más intensa en la protección de datos de carácter personal que el tratamiento de datos derivado de los identificadores tradicionales.

²²⁸ Raffiotta, Edoardo. Carlo. & Baroni, Maximiliano. “Inteligencia artificial, herramientas de identificación y protección de la identidad”, in PAJNO, A. et al. (Dirs.), *Intelligenza artificiale e Diritto: Una rivoluzione?*, Bologna: Il Mulino, 2022, pp. 383-384.

²²⁹ Entendida dicha identidad personal como el conjunto de convicciones, pensamientos y sentimientos que califican a cada individuo. *Ibid.*, pp. 384.

²³⁰ Por este motivo, se corre el riesgo de que los individuos pierdan el control sobre sus propios datos. De esta opinión, Woodward, John.D. *Super Bowl surveillance. Facing up to biometrics*, RAND Arroyo Center, 2001, p. 5.

²³¹ Brey, Philip. “Ethical aspects of facial recognition systems in public places”, *Journal of Information, Communication and Ethics in Society*, vol. 2, n°. 2, 2004, p. 107. El autor en cita plantea que la reducción funcional de partes del cuerpo en información, hace que éstas adquieran un nuevo significado o función porque permiten identificar a los individuos cuando se insertan en los dispositivos inteligentes. A esto se añade el uso y control de la información por terceros. En este punto, esos datos codificados ya no pertenecen al titular sino a quienes los procesan. Añade que, incluso aunque los titulares los poseyesen, si no entienden cómo funciona la tecnología no los pueden utilizar. De modo que esa parte del cuerpo ya no es completamente suya porque ha adquirido un nuevo significado incompresible para el individuo.

Como apuntamos en el apartado anterior, no todos los datos biométricos son igualmente visibles o accesibles. El rostro es uno de los más expuestos (por no decir el que más) a la identificación. En tal sentido, cuando una persona está en el espacio público, hace su cara disponible a la vista de los demás. No solo a las personas físicas, sino también a los sistemas de vigilancia, entre los que se encuentran los CCTV en los que se inserta la tecnología de reconocimiento facial. El principal problema con esta última es su capacidad para procesar masivamente datos en muy poco tiempo y la gravedad de los resultados. Recordemos que en el ámbito policial puede derivar en la privación de libertad.

El tratamiento de los datos biométricos está prohibido tanto en el RGPD como en la Directiva salvo determinadas excepciones. Una de ellas es precisamente que el titular los haga manifiestamente públicos (artículos 9.2 letra e RGPD y 10 letra c de la Directiva). En este caso, se plantea la cuestión de qué sucede cuando la persona está en un lugar público dotado con sistemas de videovigilancia que tienen incorporados mecanismos de reconocimiento facial. En tal sentido, nada impediría que esos sistemas reconociesen la imagen del rostro, en tanto es un dato que el titular está haciendo manifiestamente público y susceptible de ser captado.

El TEDH en el *Caso de P.G. y J.H. v Reino Unido*, se pronuncia de forma similar cuando sostiene que quien camina por la calle será inevitablemente visible para cualquier miembro del público que esté también presente. La vigilancia por medios tecnológicos de esa escena pública tiene carácter similar. Aunque continúa señalando que pueden surgir cuestiones que impacten en la privacidad, por lo que los archivos recuperados por la policía tienen que protegerse en aplicación del artículo 8 del Convenio Europeo de Derechos Humanos.

Sin embargo, siguiendo este pronunciamiento, consideramos que, al igual que ocurre con el ADN o las huellas dactilares, la tecnología de reconocimiento facial permite la extracción de información única sobre un sujeto, lo cual posibilita su identificación en circunstancias distintas. Tomado aisladamente o uniéndolo a otros metadatos, esta tecnología es una fuente importante de información personal que tiene un carácter intrínsecamente privado. Continúa señalando que, el hecho de que los datos se deriven de las características del rostro “manifiestamente público”, no le quita valor a lo anterior.

Por tanto, aun reconociendo que la cara es un tipo de información que se hace manifiestamente pública a la vista de los demás en un espacio común, no es menos cierto que existe una esfera o ámbito privado en el sentido de que debe dejarse al titular la capacidad de decidir si quiere exponer su rostro a los sistemas de vigilancia²³². Por ese motivo, adquiere especial importancia la información y el consentimiento.

²³² La Sentencia *R (Bridges) v CCSWP and SSHD* (ya citada) vino motivada por esto mismo, dado que la policía del sur de Gales había implantado el reconocimiento facial en dos momentos distintos en los

En tal sentido, del mismo modo que existe una expectativa razonable de privacidad en el propio hogar, como espacio privado por excelencia y la misma es violada cuando se instalan sistemas de vigilancia por las autoridades públicas, ocurre algo similar cuando una persona está en el espacio público. En este último, no debería abrirse la puerta a someterse a dicha vigilancia²³³.

2. IMPACTO DE LA BIOMETRÍA EN LA PROTECCIÓN DE DATOS PERSONALES: ESPECIALMENTE EL RECONOCIMIENTO FACIAL

La IA en general interfiere en la vida privada de la ciudadanía, ampliando, intensificando o incentivando las consecuencias en la misma, sobre todo en cuanto a la recopilación y uso de datos personales se refiere²³⁴. La retención y uso de datos por parte de la policía es una de las cuestiones que más interés y controversias ha generado²³⁵, principalmente cuando para ello se utilizan sistemas como el reconocimiento facial. Esto se debe a las características del rostro como dato personal especialmente sensible (y visible), como ya apuntamos en líneas superiores.

Por ese motivo, los SIBR y, más concretamente en este punto, el reconocimiento facial desplegado en lugares públicos con fines de cumplimiento del Derecho, impactan en la protección de datos personales²³⁶, porque el individuo tiene derecho a controlar su imagen y los datos relativos a la misma. Dicho control se pierde por dos cuestiones fundamentalmente. Por un lado, cuando la imagen del individuo se inserta y conserva en una base de datos (ya sea para el entrenamiento del sistema o para incluirse en una lista de vigilancia) sometida a tratamiento automatizado; y, por otro, en el momento en el que los sistemas pueden captar la imagen de una persona en el espacio público, aunque no se produzca la identificación, como a continuación desarrollaremos.

que el demandante podría haber sido identificado por tales mecanismos, pues en ningún momento sintió que ello tuviese lugar y alega además que no fue debidamente informado.

²³³ Raposo, Véra.Lucía. “Health, privacy and liberty: a call for digital governance during (and after) the pandemic”, *The International Journal of Human Rights*, vol. 27, n°. 3, 2023, p. 533.

²³⁴ Así lo entiende el Consejo de Derechos Humanos de Naciones Unidas. Véase, *El derecho a la privacidad en la era digital. Informe de la Alta Comisionada de las Naciones Unidas para los Derechos Humanos*. 48º período de sesiones, 2021, p. 4.

²³⁵ Coghlan, Niall. “*Gaughran* [2015] UKSC 29 on police biometric data retention: Strasbourg unheard, Strasbourg unheeded”, *European Human Rights Law Review*, n°. 21, 2016, p. 61.

²³⁶ Herverg, Jean. & Van Gyseghem, Jean.Marc. “La protection des données à caractère personnel en droit européen- Chronique de jurisprudence (2020)”, *Journal Européen des Droits de l'Homme*, n°. 1, 2021, p. 69.

2.1. La protección de datos personales en los distintos momentos en los que opera el sistema

Los SIBR y, específicamente, la tecnología de reconocimiento facial, actúan en distintos momentos, desde la recogida de la información de los datos biométricos en cuestión, su procesamiento, hasta la identificación última. En todas esas etapas, la protección de datos de carácter personal puede verse dañada de distinta forma.

En primer lugar, cuando se recolecta información del rostro y se transforma en una plantilla de caracteres alfanuméricos capaz de ser procesada por los algoritmos con los que opera el mecanismo. Todo ello, obviamente, sin el conocimiento ni consentimiento del titular, dado que podría estar en su ánimo negarse a dicho tratamiento. Esta situación es mucho más controvertida cuando los datos se han extraído de redes sociales u otros sitios web en los que se comparten fotografías o vídeos. En este último caso, pueden utilizarse tanto para el entrenamiento del mecanismo, el cual necesita una ingente cantidad de datos para actuar con mayor nivel de precisión, como para su inclusión en las listas de vigilancia policial.

Por otro lado, cuando el sistema ya está puesto en marcha y es capaz de identificar a una persona -ya sea el objetivo policial, o no- a través de la comparación de su información biométrica captada por los CCTV u otros dispositivos de vigilancia con las plantillas recogidas en la base de datos en cuestión. En este momento, el problema aparece por el hecho de que la captación de la imagen tiene lugar en el ámbito público, donde la persona, aun viendo limitadas sus expectativas de privacidad y, en este caso concreto, de que sus datos personales pasen inadvertidos, espera razonablemente que los mismos no sean recogidos y utilizados con fines de identificación.

Si ya la mera conservación y almacenamiento de datos personales por parte de la policía, con independencia de cómo se hayan obtenido, impacta directamente en la privacidad de los afectados, sin importar que luego sean utilizados o no²³⁷ ni el sentido del resultado emitido²³⁸; un efecto mucho mayor se deriva de la posibilidad de que el rostro o cualquier otro dato biométrico sea captado y, eventualmente, identificado con los contenidos en la lista de vigilancia. Por tanto, ya desde antes de su despliegue, esta tecnología interfiere considerablemente en el derecho aquí analizado.

En este punto vamos a indicar algunos de los aspectos por los que el reconocimiento facial puede colisionar con la protección de datos. No vamos a desarrollar

²³⁷ Cotino Hueso, Lorenzo. “Reconocimiento facial automatizado y sistemas de identificación biométrica bajo la regulación superpuesta de inteligencia artificial y protección de datos”, en Balaguer Callejón, F. y Cotino Hueso, L. (Coords.), *Derecho Público de la Inteligencia Artificial*, Fundación Manuel Jiménez Abad de Estudios Parlamentarios y del Estado Autonómico, Zaragoza, 2023, p. 360.

²³⁸ SEPD, *Directrices 5/2022 sobre el uso de la tecnología de reconocimiento facial en el ámbito de cumplimiento del Derecho*. Versión 2.0, adoptadas el 26 de abril de 2023, p. 5.

cada uno de ellos, dado que escapa en mucho del objetivo y extensión de un trabajo de estas características y debiera constituir un estudio mucho más amplio. Por tanto, nos limitaremos a señalarlos, teniendo también en consideración que algunas de las cuestiones que afectan a este derecho impactan en otros como la no discriminación, por la posibilidad de emitir resultados erróneos y sesgados.

Así, siguiendo a NERONI REZENDE, estos serían: 1) errores porque, aunque se utilicen técnicas de *Deep learning*, no son infalibles y pueden emitir resultados incorrectos, lo cual pone en entredicho la supuesta imparcialidad y precisión del procesamiento de datos; 2) vinculado con el anterior, la inexactitud estadística, porque los datos utilizados en el entrenamiento pueden no ser muy representativos de la sociedad en la que se vaya a insertar, de lo cual pueden derivarse sesgos y, de ahí, cuestiones discriminatorias que pueden intensificarse si los agentes hacen un mal uso de los sistemas, con el objetivo, por ejemplo, de monitorizar a minorías étnicas o comunidades marginadas y oprimidas; 3) seguridad de los datos por su posible pérdida o fuga, lo cual podría derivar en consecuencias irremediables para los titulares. Además, los instrumentos pueden alterarse, por ejemplo, a través de *Deepfakes*; 4) las autoridades públicas, a menudo, usan esta tecnología en cooperación con entidades privadas, lo cual conlleva el riesgo de que los datos puedan ser recogidos por estas últimas y que los faciliten a aquellas con fines meramente lucrativos; y 5) *function creep*, o que los datos se utilicen con otros fines distintos de los iniciales²³⁹, ya sea en actividades o personas objeto de investigación.

A los señalados, es necesario incluir la recolección masiva de datos que puede derivar en un sentimiento de vigilancia masiva²⁴⁰. Tal práctica es capaz de vulnerar este derecho desde el momento en el que, esa información puede vincularse a otra, también de carácter personal y sensible en algunos casos, como la ideología política, creencias religiosas, estado de salud, hábitos de vida, lugares y personas que se frecuentan, etc. con el objetivo -entre otros- de crear perfiles.

²³⁹ Neroni Rezende, Isadora. *Surveillance Risks in IoT applied to Smart cities*, Tesis doctoral defendida en 2023 en la Universidad de Bolonia, Italia, p. 205.

²⁴⁰ En dicho sentido, el TJUE se ha pronunciado sobre esta cuestión, confirmando que la recogida general e indiscriminada de datos supone una violación del derecho a la protección de datos de carácter personal en la Unión Europea. Véanse los Casos C-203/15 y C-698/15, de diciembre de 2016. ECLI:EU:C:2016:970. En los mismos, sostiene que la conservación de datos de los servicios de comunicaciones electrónicas supone una grave injerencia en este derecho, pudiendo provocar un sentimiento de vigilancia indiscriminada.

Igualmente, en Sentencia de 6 de octubre de 2020, para el caso de *La Quadrature du net y otros*, C-511, 512 y 520/18, ECLI:EU:C:2020:791, ha concretado como contrario al Derecho de la Unión y la Carta, las actividades de recogida y conservación generalizada e indiferenciada de datos personales por motivos de seguridad nacional.

2.2. Fuentes de las que se extrae la información: ¿Qué ocurre con las imágenes publicadas en redes sociales?

Las bases de datos policiales utilizadas para la puesta en marcha de los SIBR, concretamente aquellas que recurren al rostro como parámetro para realizar la identificación, pueden provenir de fuentes oficiales como los carnés de identidad o los pasaportes, así como bases de datos de delincuentes conocidos contra los que existe una orden de búsqueda y captura o un proceso judicial abierto. Como venimos señalando en otras partes de este capítulo, junto a las anteriores fuentes, también pueden extraerse de redes sociales y demás recursos virtuales, no solo recogidas por la policía sino también desde el sector privado²⁴¹.

A modo de ejemplo, el FBI usa desde hace tiempo la base de datos *Next Generation Identification system*, en la que se incluyen plantillas faciales con otros tipos de datos biométricos como las huellas dactilares. En este caso, la tecnología de reconocimiento facial tiene la capacidad de buscar en las bases de datos estatales de licencias de conducir y otros repositorios como *US Visitor and Immigrant Status Indicator Technology*, que recoge la biometría de todos los ciudadanos no nacionales que acceden a Estados Unidos²⁴².

Otro país en el que se recurre a la tecnología de reconocimiento facial aplicada a la identificación forense es Pakistán. Para ello, se vinculan las imágenes extraídas de los CCTV con la base de datos de identidad nacional (NADRA) que se une, a su vez, a otras que contienen otros identificadores biométricos como las huellas dactilares²⁴³. Algunos autores consideran que esta base de datos no es suficiente para realizar la identificación, motivo por el que entienden que debería acudir a redes sociales como Facebook o Instagram, entre otras, y utilizarse datos de acceso abierto²⁴⁴.

²⁴¹ El FBI reconoce que sus bases de datos pueden estar conformadas por imágenes provenientes de fuentes oficiales o bien de Internet o redes sociales y son recogidas desde el sector privado. Prest, Erind.M. *Privacy Impact Assessment for the Next Generation Identification-Interstate Photo System*, Federal Bureau of Investigation, 2019, p. 8.

²⁴² Otras agencias con acceso a la búsqueda con reconocimiento facial son *Customs and Border Protection*, *Coast Guard*, *Citizenship and Immigration Services*, *Department of State*, *Department of Defence* y *Department of Justice*, así como agentes policiales y servicios de inteligencia. Además, el Gobierno de EE.UU. en casos concretos puede ofrecer acceso a los repositorios de imágenes a países socios como Australia, Canadá, Nueva Zelanda y Reino Unido. Smith, Marcus. & Miller, Seumas. *Biometric identification, Law and Ethics...*, cit., p. 27.

²⁴³ Más ampliamente, Kausar, Shaban. et al. "Comparative study of forensic recognition and fingerprint during crime scene investigation and the role of artificial intelligence tools in forensics", *Journal of Development and Social Sciences*, vol. 5, n.º. 1, 2024, pp. 585-597.

²⁴⁴ *Ibid.*, 594.

Bajo nuestro punto de vista, acudir a esta última fuente de información plantea serios problemas con la protección de datos de los usuarios cuyas fotografías han sido utilizadas, bien para entrenar el sistema²⁴⁵, bien en las bases de datos de búsqueda policial.

Un supuesto controvertido que implica a agentes privados en la recolección y creación de bases de datos a través de redes sociales y otros sitios web, es el de *Clearview AI*. Esta empresa privada estadounidense ofrece su servicio de reconocimiento facial a las autoridades públicas con el objetivo de identificar a sospechosos en investigaciones policiales y a compañías privadas de seguridad²⁴⁶. Va mucho más allá de las tradicionales herramientas de reconocimiento facial limitadas a imágenes recogidas por el Gobierno, pues combina esta tecnología con una base de datos que actualmente cuenta con más de 60 billones de imágenes publicadas en Internet, recogidas con un *software* automático que permite la recolección masiva de información²⁴⁷.

La técnica que acaba de exponerse se denomina *data scraping* y normalmente se realiza en el ámbito empresarial. Esta se define como “una plétora de metodologías de recuperación de datos a través de Internet, usadas sin el permiso del propietario de los

²⁴⁵ De esta opinión, entre otros, Colacuri, Marco. “Riconoscimento faciale e rischi per i diritti fondamentali alla luce delle dinamiche di relazione tra poteri pubblici, imprese e cittadini”, en Balbi, G., et al, (Dirs.), *Diritto penale e intelligenza artificiale “nuovi scenari”*, Torino: G. Giappichelli Editore, 2022, pp. 128-129.

²⁴⁶ Smith, Marcus. & Miller, Seumas. *Biometric identification, Law and Ethics...*, cit., pp. 27-28.

En el año 2022 la empresa vendió su tecnología al Gobierno de Ucrania para identificar personas en la guerra, vivas o muertas. Clayton, James. “La controvertida tecnología que usa Ucrania para identificar a los muertos y enemigos en la guerra”, *BBC News*, 3 de abril de 2022. Recurso electrónico obtenido a través de la Web: <https://www.bbc.com/mundo/noticias-internacional-61096583>, [recuperado el día 20 de noviembre de 2025].

Cuestión distinta es que las bases de datos se utilicen para el entrenamiento de algoritmos, es decir, desde el ámbito privado, y no con fines policiales. En este caso, su uso estaría regulado por el RGPD y no por la Directiva de protección de datos en el ámbito penal. En este sentido, traemos a colación la base de datos MegaFace, que tiene almacenadas más de un millón de caras de alrededor de 700.000 individuos. El consentimiento no tiene virtualidad en este caso, dado que este requisito puede ser sustituido por la persecución de otro fin, como sucede con la investigación, en este supuesto concreto, en materia de reconocimiento facial y desarrollo de algoritmos y su entrenamiento para mejorar su eficacia. Para construir esta base de datos se recogieron las imágenes de Yahoo, con el consentimiento de quien había tomado la fotografía o había grabado el vídeo en el que aparecían las imágenes del rostro tomadas para conformar la base de datos, creada con una licencia *Creative Commons* que permite reutilizar las imágenes. Para conocer más sobre la cuestión, véase, Jasserand, Catherine. “Massive Facial Database and the GDPR...”, cit., pp. 169-188.

²⁴⁷ Véase su página web: <https://www.clearview.ai/>. [recuperado el día 25 de noviembre de 2025].

datos”²⁴⁸. Actualmente se realiza con *machine learning*, lo cual le otorga un ingente poder de recopilación de datos. La práctica mencionada lesiona claramente la protección de datos, llegando incluso algunas redes sociales como Twitter a demandar a esta compañía, exigiendo que dejen de recoger imágenes de sus usuarios y se eliminen todos los datos almacenados²⁴⁹.

Como es una empresa estadounidense, la cuestión que surge inmediatamente es si podría utilizarse en la Unión Europea. La policía de Suecia recurrió a la base de datos de *Clearview AI* hace unos años y, por ese motivo, la Autoridad Sueca de Protección de la Privacidad (IMY) inició una investigación en la que concluyó que la policía no cumplió sus obligaciones como responsable del tratamiento de datos porque no implementó medidas organizativas suficientes para garantizar y demostrar que el tratamiento fue conforme a Derecho y tampoco realizó una evaluación de impacto.

Por ello, la IMY impuso una multa administrativa al cuerpo policial, ordenando además que impartiese información y capacitación adicional a sus empleados para evitar un futuro tratamiento de datos personales que infrinja la normativa de protección de datos; también se le impuso el deber de informar a los interesados cuyos datos fueron revelados a la aplicación sobre este punto, cuando las normas de confidencialidad lo permitan. Junto a lo anterior, tienen que eliminar la información personal recogida²⁵⁰. A lo anterior se añade que algunos países europeos como Países Bajos²⁵¹, Grecia o Italia han demandado a la compañía por recoger y almacenar datos de sus

²⁴⁸ Campbell, Fiona. “Data Scraping-Considering the Privacy Issues”, *Fieldfisher*, 2019. Recurso electrónico obtenido a través de la Web: <https://www.fieldfisher.com/en/services/privacy-security-and-information/privacy-security-and-information-law-blog/data-scraping-considering-the-privacy-issues>, [recuperado el día 18 de noviembre de 2025]. Kindt muestra su preocupación por el hecho de que las autoridades públicas tengan cada vez más acceso a bases de datos no creadas inicialmente con fines policiales, ya provengan del sector público o privado. Ello porque esta situación puede provocar que el siguiente paso sea crear una normativa que permita el acceso a esas bases de datos. Kindt, Els.J. Having yes, using no?...”, cit., p. 528.

²⁴⁹ BBC, *Twitter demands AI company stops ‘collecting faces’*, 23 de enero de 2020. Recurso electrónico obtenido a través de la Web: <https://www.bbc.com/news/technology-51220654>, [recuperado el día 18 de noviembre de 2025].

²⁵⁰ SEPD. *Swedish DPA: Police unlawfully used facial recognition app*, 12 de febrero de 2021. Recurso electrónico obtenido a través de la Web: https://www.edpb.europa.eu/news/national-news/2021/swedish-dpa-police-unlawfully-used-facial-recognition-app_en, [recuperado el día 21 de noviembre de 2025].

²⁵¹ Soler, Adrián. “Clearview AI recibe una multa de 30 millones de euros por incumplir el RGPD”, *Paréntesis Media*, 5 de septiembre de 2024. Recurso electrónico obtenido a través de la Web: <https://www.parentesis.media/clearview-ai-recibe-una-multa-de-30-millones-de-euros-por-incumplir-el-rgdp/>, [recuperado el día 20 de noviembre de 2025].

nacionales²⁵², violando el RGPD y la Directiva de protección de datos personales en el ámbito penal.

En dicho sentido, la práctica se ejecuta con un fin puramente económico/comercial, dado que *Clearview AI* no realiza actividades policiales, sino que vende sus productos a la policía y otras empresas privadas de seguridad. Por tanto, no puede considerarse que esté realizando una función delegada en los términos establecidos en el artículo 3.7.b) de la Directiva. Así, entra dentro del ámbito regulatorio del RGPD²⁵³. Además, teniendo en cuenta que el rostro es una categoría especial de dato personal, su tratamiento tiene que realizarse según lo dispuesto en el artículo 9 RGPD, estando prohibido en su apartado primero. El apartado segundo letra e) del mismo precepto establece como excepción a dicha prohibición que el interesado haya hecho manifiestamente públicos esos datos.

En vista de lo que acabo de apuntar, surge inmediatamente la cuestión de si el hecho de que los usuarios de redes sociales publiquen una imagen, le confiere ese carácter de manifiestamente público. Las redes sociales, por lo general, no permiten la técnica del *scraping*, por lo que la compañía estaría vulnerando desde el inicio la protección de datos, al no cumplir con la política de privacidad de la red social en cuestión.

No obstante, en el supuesto de que esté permitido, si las imágenes se han compartido a través de una cuenta pública sin restricción de acceso a los usuarios, podría pensarse que se está consintiendo, aun de forma tácita a que sean utilizadas por terceros con otros fines. En este punto, debe aclararse que el acceso al contenido es lícito y permitido porque el titular de la imagen lo está autorizando. Sin embargo, lo que no está consintiendo es su divulgación o uso por otros, quienes, en tal caso, sí necesitarán consentimiento expreso para ello. En tal línea, cabe destacar la STC (Sala Segunda) 27/2020, de 24 de febrero, donde se señala que la publicación en una red social no es equiparable a la captación de imágenes en lugares públicos. A ello hay que sumar que los espacios virtuales no se consideran públicos en el RIA.

Si la captación y uso de imágenes en perfiles abiertos en redes sociales impacta en la protección de datos, más aún lo hará cuando aquellos son privados, pues en este caso el titular está restringiendo el círculo de quienes pueden acceder al contenido compartido, haciendo de este modo que la imagen sea “privada”²⁵⁴. De este modo,

²⁵² Aguilar, Alberto.R. “Clearview AI, la controvertida compañía proveedora de reconocimiento facial a policías de todo el mundo, denunciada en cinco países europeos”, *Business insider*, 27 de mayo de 2021. Recurso electrónico obtenido a través de la Web: <https://www.businessinsider.es/politica/controvertida-clearview-ai-denunciada-5-paises-europeos-872917>, [recuperado el día 20 de noviembre de 2025].

²⁵³ Neroni Rezende, Isadora. “Facial recognition in police hands: Assessing the ‘Clearview case’ from a European perspective”, *New Journal of European Criminal Law*, vol. 11, n°. 3, 2020, pp. 378-380.

²⁵⁴ De forma similar, *ibid.*, pp. 380-381. Por su parte, el SEPD en las *Directrices 5/2022 sobre el uso de la tecnología de reconocimiento facial en el ámbito de cumplimiento del Derecho...*, cit., p. 6, sostiene que el hecho de

aunque libremente se publique y comparta información sobre uno mismo en dichos espacios virtuales, no quiere decir que se esté desdeñando la privacidad²⁵⁵. Ello porque el usuario tiene derecho a que sus datos permanezcan ocultos a los demás, de forma que, si la red social o sitio web es privado, la información compartida adquiere tal consideración, no pudiendo los datos personales recogerse ni utilizarse por otros, aunque tengan conocimiento técnico para acceder a la imagen. Asimismo, forma parte del ámbito privado de la vida de las personas que no puede violarse.

En relación con lo que acabamos de apuntar, hay que indicar que, incluso cuando esos espacios virtuales sean de acceso público, como ocurre en la vida real, se tienen expectativas de privacidad, no pudiéndose tomar las imágenes con fines de creación de bases de datos, más aún cuando la red social no tiene una política de privacidad que lo permita. Esto se debe a que el titular no ha prestado su consentimiento a ese tratamiento y puede no conocer que sus datos se están utilizando con tales fines, ofreciéndose a las autoridades policiales o privadas de seguridad. A lo anterior hay que añadir que la empresa dedicada al *scraping* recoge imágenes de forma indiscriminada de personas que no están involucradas en un proceso penal, suponiendo una recogida masiva de información personal que atenta, a todas luces, contra la protección de datos.

Que sea útil o conveniente para la policía no justifica su puesta en marcha porque, en caso de aceptar su despliegue, se le estaría otorgando un poder de vigilancia sin precedentes, derivando en serios perjuicios para la ciudadanía. Por tanto, esas bases de datos no deberían utilizarse. En tal línea, el SEPD solicita la prohibición de la técnica del *scraping* para que los agentes no puedan acceder a bases de datos con información recogida a gran escala²⁵⁶.

No solo eso, sino que dicha prohibición ya viene recogida en el RIA, más concretamente en el Considerando 43, bajo el entendimiento de que la creación de bases de datos a través de la extracción no selectiva de imágenes faciales a través de Internet

que el usuario haya hecho pública una imagen no extiende dicha publicitación de los datos biométricos que contiene, los cuales pueden extraerse de la fotografía mediante instrumentos técnicos.

En sentido contrario se pronuncia el A29 WP en la *Dictamen sobre determinadas cuestiones fundamentales de la Directiva sobre protección de datos en el ámbito penal*, 29 de noviembre de 2017, p. 10. Sostiene que en los casos en los que se han publicado datos personales en una red social existe una clara intención de hacerlos públicos y la aceptación de ciertas reglas de protección de datos que prevén que todos los socios del proveedor, incluidas las autoridades policiales, tengan acceso a los mismos.

²⁵⁵ De esta opinión, North-Samardzic, Andrea. “Biometric technology and ethics: Beyond security applications”, *Journal of Business Ethics*, vol. 167, n.º. 3, 2020, p. 440. Como establece la autora, que se comparta información personal en tales espacios no significa que los individuos estén menos preocupados o sean menos conscientes de su privacidad.

²⁵⁶ SEPD, *Directrices 5/2022 sobre el uso de la tecnología de reconocimiento facial en el ámbito de cumplimiento del Derecho...*, cit., p. 7.

o de los CCTV agrava el sentimiento de vigilancia masiva y puede provocar graves violaciones de los derechos fundamentales.

Por tanto, este tipo de prácticas están prohibidas en la Unión Europea, no siendo posible recurrir a las mismas. Sin embargo, las bases de datos faciales policiales se aceptan en caso de delitos serios, siempre que estén sujetas a adecuados mecanismos de rendición de cuentas²⁵⁷. A pesar de todo, hemos de tomar con cautela lo que acabamos de indicar, pues el procesamiento y tratamiento del rostro por la policía, aun en casos de delincuencia grave, puede interferir en la protección de datos de carácter personal.

3. CONCLUSIONES

Como hemos tenido ocasión de poner de manifiesto a lo largo del presente estudio, el uso policial de los SIBR tanto en tiempo real como en diferido, pone en serio peligro (y vulnera) el derecho a la protección de datos de carácter personal por diversos motivos, con especial incidencia la técnica de reconocimiento facial.

En primer lugar, encontramos que los datos utilizados son personales y de naturaleza especialmente sensible, motivo por el que su procesamiento y tratamiento está prohibido en la normativa europea de protección de datos. Sin embargo, el RIA lo permite. A esto hay que añadir, en relación con el sistema de reconocimiento facial, que el rostro es el parámetro biométrico más expuesto. Es visible y fácilmente accesible, tanto en el espacio físico como en el virtual. Por ello, la información del mismo puede recogerse y utilizarse en muchas ocasiones de forma encubierta, sin el conocimiento ni, por supuesto, el consentimiento del titular; aunque bien es cierto que, en el transcurso de una investigación policial, muchas actuaciones deben ser secretas para su adecuado discurrir, lo cual deriva en que no se conozca la puesta en marcha de esta técnica.

En segundo y, vinculado con lo anterior, cualquier persona que esté en el espacio público en el que estos mecanismos actúan, corren el riesgo de que su biometría sea capturada por el sistema y, en consecuencia, puedan ser identificados. En un principio, solamente debería identificar a quienes son objeto de búsqueda. Sin embargo, la técnica no es infalible y puede cometer errores, pudiendo realizar identificaciones incorrectas. A ello se añade que, aunque la identificación no tenga lugar, la imagen del rostro es captada por esta tecnología, aunque sea descartada inmediatamente en caso de no producirse la identificación.

Ya la mera recolección y almacenamiento de datos por parte de la policía impacta en el derecho a la protección de datos personales, más aún cuando se realiza un

²⁵⁷ De esta opinión, SMITH, M. & MILLER, S. *Biometric identification, Law and Ethics...*, cit., p. 32.

tratamiento automatizado de los mismos, que pasan a convertirse en meros dígitos que escapan del control del titular.

En tercero, los SIBR actúan con datos, en este caso, personales. Para una mayor efectividad, el algoritmo tiene que entrenarse, para lo cual requiere también de datos. El problema en este punto surge de las fuentes de las que se extraen. Así, pueden provenir de bases de datos policiales de delincuentes conocidos y sospechosos, así como de las imágenes de las licencias de conducir, documentos de identidad y pasaportes, es decir, fuentes oficiales. No obstante, también cabe la posibilidad de que provengan de bases de datos creadas *ad hoc* por empresas privadas que luego las venden a la policía o compañías privadas de seguridad. En tal supuesto, suelen realizar una recolección masiva e indiscriminada de imágenes faciales a través de redes sociales y otros sitios web. Esta práctica impacta directamente y lesiona la protección de datos personales porque el titular no consiente que la información de su rostro, aun publicada en dichos espacios, sea utilizada con estos fines. A lo que se añade que la mayoría de estos entornos digitales no permiten semejante técnica en su política de privacidad.

Por tanto, cabe concluir que los SIBR, con mayor incidencia el reconocimiento facial, impactan en el derecho aquí analizado (y en otros muchos) de forma considerable, por todo lo expuesto a lo largo del presente estudio, entre otras cuestiones que, por motivos de espacio no hemos desarrollado. El hecho de que el RIA regule el uso policial de estos sistemas permite su despliegue en la Unión Europea y, aunque prevea una serie de limitaciones y requisitos, el riesgo que entraña en los derechos es muy elevado, tal vez excesivo. Habrá que esperar a ver cómo los Estados introducen las previsiones del RIA en su normativa interna, dado que se les deja potestad para ello, pues la seguridad continúa siendo competencia estatal.

APÉNDICE BIBLIOGRÁFICO

- Ad hoc Committee on Artificial Intelligence (CAHAI). *Possible elements of a legal framework on Artificial Intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law*, Strasbourg, December 3, 2021. CAHAI(2021)09rev
- Aguilar, Alberto.R. “Clearview AI, la controvertida compañía proveedora de reconocimiento facial a policías de todo el mundo, denunciada en cinco países europeos”, *Business insider*, 27 de mayo de 2021. Recurso electrónico obtenido a través de la Web: <https://www.businessinsider.es/politica/controvertida-clearview-ai-denunciada-5-paises-europeos-872917>, [recuperado el día 20 de noviembre de 2025].
- Alto Comisionado de Naciones Unidas para los Derechos Humanos. *Impact of new technologies on the promotion and protection of human rights in the contexts of assemblies, including peaceful protests*, 2020. A/HRC/44/24.
- BBC, *Twitter demands AI company stops 'collecting faces'*, 23 de enero de 2020. Recurso electrónico obtenido a través de la Web: <https://www.bbc.com/news/technology-51220654>, [recuperado el día 18 de noviembre de 2025].
- Brey, Philip. “Ethical aspects of facial recognition systems in public places”, *Journal of Information, Communication and Ethics in Society*, vol. 2, n°. 2, 2004.
- CEPD-SEPD. *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*, 18 June 2021. Recurso electrónico obtenido a través de la Web: https://www.edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf, [última consulta el día 11 de noviembre de 2025].
- Campbell, Fiona. “Data Scraping-Considering the Privacy Issues”, *Fieldfisher*, 2019. Recurso electrónico obtenido a través de la Web: <https://www.fieldfisher.com/en/services/privacy-security-and-information/privacy-security-and-information-law-blog/data-scraping-considering-the-privacy-issues>, [recuperado el día 18 de noviembre de 2025].
- Christakis, Theodore. et al. *Facial Recognition in the Draft European AI Regulation: Final Report on the High-Level Workshop Held on April 26, 2021*, AI-Regulation.com, May 27th, 2021. Recurso electrónico obtenido a través de la Web: <https://ai-regulation.com/facial-recognition-in-the-draft-european-ai-regulation-final-report-on-the-high-level-workshop-held-on-april-26-2021/>, [última consulta el día 12 de noviembre de 2025].
- Clayton, James. “La controvertida tecnología que usa Ucrania para identificar a los muertos y enemigos en la guerra”, *BBC News*, 3 de abril de 2022. Recurso electrónico obtenido a través de la Web: <https://www.bbc.com/mundo/noticias-internacional-61096583>, [recuperado el día 20 de noviembre de 2025].
- Coghlan, Niall. “*Gaughran* [2015] UKSC 29 on police biometric data retention: Strasbourg unheard, Strasbourg unheeded”, *European Human Rights Law Review*, n°. 21, 2016.

- Colacuri, Marco. “Riconoscimento faciale e rischi per i diritti fondamentali alla luce delle dinamiche di relazione tra poteri pubblici, imprese e cittadini”, en Balbi, G., et al. (Dirs.), *Diritto penale e intelligenza artificiale “nuovi scenari”*, Torino: G. Giappichelli Editore, 2022.
- Comisión Europea. *Communication to the Commission. Approval of the content of the draft Communication from the Commission –Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)*, Bruselas, 4 de febrero de 2025. C(2025) 888 final.
- Consejo de Derechos Humanos de Naciones Unidas. *El derecho a la privacidad en la era digital. Informe de la Alta Comisionada de las Naciones Unidas para los Derechos Humanos*. 48º período de sesiones, 2021.
- Cotino Hueso, Lorenzo. “Reconocimiento facial automatizado y sistemas de identificación biométrica bajo la regulación superpuesta de inteligencia artificial y protección de datos”, en Balaguer Callejón, F. y Cotino Hueso, L. (Coords.), *Derecho Público de la Inteligencia Artificial*, Zaragoza: Fundación Manuel Jiménez Abad de Estudios Parlamentarios y del Estado Autonómico, 2023.
- Couchman, Hannah. “Mass Surveillance on Our Streets: Facial Recognition and the Threat to Our Rights”, *European Human Rights Law Review*, nº. 6, 2019.
- Donatti, Filippo. “Diritti fondamentali e algoritmi nella proposta di Regolamento sull’Intelligenza Artificiale”, en Pajno, A. et al. (Dirs.), *Intelligenza Artificiale e diritto: una rivoluzione?*, Bologna: Il Mulino, 2022.
- Dushi, Desara. “The use of facial recognition technology in EU law enforcement: Fundamental rights implications”, *Global Campus of Human Rights. South East Europe*, Policy Briefs, 2020.
- Finocchiaro, Giusella. *Intelligenza artificiale. Quali regole?*, Bologna: Il Mulino, 2023.
- Gates, Kelly.A. *Our biometric future: Facial recognition technology and the culture of surveillance*. Vol. 2, New York University Press, 2011.
- Goold, Benjamin. et al. “The banality of security. The curious case of surveillance cameras”, *British Journal of Criminology*, vol. 53, nº. 6, 2013.
- Grupo de Trabajo del Artículo 29. *Dictamen 3/2012 sobre la evolución de las tecnologías biométricas*, 00720/12/EN WP103, 27 de abril 2012.
- *Dictamen sobre determinadas cuestiones fundamentales de la Directiva sobre protección de datos en el ámbito penal*, 29 de noviembre de 2017.
- Herverg, Jean. & Van Gyseghem, Jean.Marc. “La protection des données à caractère personnel en droit européen- Chronique de jurisprudence (2020)”, *Journal Européen des Droits de l’Homme*, nº. 1, 2021.
- Jasserand, Catherine. “Legal Nature of Biometric Data: From ‘Generic’ Personal Data to Sensitive Data”, *European Data Protection Law Review*, vol. 2, 2016.
- Kausar, Shaban. et al. “Comparative study of forensic recognition and fingerprint during crime scene investigation and the role of artificial intelligence tools in forensics”, *Journal of Development and Social Sciences*, vol. 5, nº. 1, 2024.
- Kindt, Els.J. “Having yes, using no? About the new legal regime for biometric data”, *Computer Law & Security Review*, vol. 34, 2018.

- Lynch, Nessa. “Facial Recognition Technology in Policing and Security-Case Studies in Regulation”, *Laws*, vol. 13, n°. 35, 2024.
- Neroni Rezende, Isadora. “Facial recognition in police hands: Assessing the ‘Clearview case’ from a European perspective”, *New Journal of European Criminal Law*, vol. 11, n°. 3, 2020.
- *Surveillance Risks in IoT applied to Smart cities*, Tesis doctoral defendida en 2023 en la Universidad de Bolonia, Italia.
- Niklas, Pablo. & Barrera, Gabriel. “Biometría aplicada a la seguridad pública”, *Simposio Argentino sobre Tecnología y Sociedad (STS)-JAIIO 46*, 2017.
- North-Samardzic, Andrea. “Biometric technology and ethics: Beyond security applications”, *Journal of Business Ethics*, vol. 167, n°. 3, 2020.
- Prest, Erind.M. *Privacy Impact Assessment for the Next Generation Identification-Interstate Photo System*, Federal Bureau of Investigation, 2019.
- Purshousea, Joe. & Campbell, Liz. “Automated facial recognition and policing: A Bridge too far?”, *Legal Studies*, vol. 42, 2022.
- Smith, Marcus. & Miller, Seumas. *Biometric identification, Law and Ethics*, Springer, 2021.
- Soler, Adrián. “Clearview AI recibe una multa de 30 millones de euros por incumplir el RGPD”, *Paréntesis Media*, 5 de septiembre de 2024. Recurso electrónico obtenido a través de la Web: <https://www.parentesis.media/clearview-ai-recibe-una-multa-de-30-millones-de-euros-por-incumplir-el-rgdp/>, [recuperado el día 20 de noviembre de 2025].
- Stitilis, Darius. & Laurinaitis, Marius. “Treatment of biometrically processed personal data: Problem of uniform practice under EU personal data protection law”, *Computer Law & Security Review*, vol. 33, 2017.
- Sumer, Bilgesu. “When do the images of biometric characteristics qualify as special categories of data under the GDPR?: A systematic approach to biometric data processing”, *2022 Intenational Conference of the Biometrics Special Interest Group (BI-OSIG)*, 2022.
- Supervisor Europeo de Protección de Datos. *Artificial Intelligence Act: a welcomed initiative, but ban on remote biometric identification in public space is necessary*. Recurso electrónico obtenido a través de la Web: https://www.edps.europa.eu/press-publications/press-news/press-releases/2021/artificial-intelligence-act-welcomed-initiative_en, [última consulta el día 12 de noviembre de 2025].
- *Opinion of 19 October 2005 on three Proposals regarding the Second Generation Schengen Information System (SIS II) (Com(2005) 236 final y COM(2005) 237 final)*, OJ C 91,19.4.2006.
 - *Opinion of 18 July 2013 on the Proposals for a Regulation establishing an Entry/Exit System (EES) and a Regulation establishing a Registered Traveller Programme (RTP)*.
 - *Opinion 3/2016 of 13 April 2016, on the Exchange of information on third country nationals as regards the European Criminal Records Information System (ECRIS)*.
 - *Opinion of 07/2016 on the First reform package on the Common European Asylum System (Eurodac, EASO and Dublin Regulations)*.
 - *Opinion 7/2018 of 10 August 2018, on the Proposal for a Regulation strengthening the security of identity cards of Union citizens and other documents*.

- *Directrices 5/2022 sobre el uso de la tecnología de reconocimiento facial en el ámbito de cumplimiento del Derecho. Versión 2.0*, adoptadas el 26 de abril de 2023.
- Raffiotta, Edoardo.Carlo. & Baroni, Maximiliano. “Inteligencia artificial, herramientas de identificación y protección de la identidad”, in PAJNO, A. et al. (Dirs.), *Intelligenza artificiale e Diritto: Una rivoluzione?*, Bologna: Il Mulino, 2022.
- Raposo, Véra.Lucía. “(Do not) remember my face: uses of facial recognition technology in light of the general data protection regulation”, *Information & Communications Technology Law*, vol. 32, n°. 1, 2023.
- “Health, privacy and liberty: a call for digital governance during (and after) the pandemic”, *The International Journal of Human Rights*, vol. 27, n°. 3, 2023.
- Woodward, John.D. *Super Bowl surveillance. Facing up to biometrics*, RAND Arroyo Center, 2001.
- Yue Liu, Nancy. *Bio-Privacy. Privacy Regulations and the Challenge of Biometrics*, Oxon, Abingdon: Routledge, 2012.



CAPÍTULO 7.

ENTRE LA LEGALIDAD Y EL ALGORITMO: PRINCIPIOS PROCESALES E INTELIGENCIA ARTIFICIAL EN LA EJECUCIÓN DE MEDIDAS DEL PROCESO PENAL DE MENORES²⁵⁸

ELENA PINEROS POLO

Profª. Permanente Laboral
Universidad Rey Juan Carlos

Resumen

La ejecución de medidas en el proceso penal de menores constituye una fase procesal decisiva en la que se concreta la respuesta del ordenamiento jurídico frente a la delincuencia juvenil y se despliegan las finalidades reeducativas del sistema. Tomando como eje la Ley Orgánica 5/2000 y su desarrollo reglamentario, el capítulo examina en primer lugar, el catálogo de medidas y los criterios de individualización que condicionan su cumplimiento, con especial atención a la distinción entre medidas privativas y no privativas de libertad y a su impacto en la fase ejecutiva. Se estudia la arquitectura de la ejecución, articulada en torno al reparto funcional de competencias entre el Juzgado de Menores y la entidad pública de protección o reforma, así como el papel del Ministerio Fiscal, la defensa del menor y los equipos técnicos. Sobre esta base, se precisan los principales principios procesales que informan la ejecución: legalidad, jurisdiccionalidad, interés superior del menor, proporcionalidad, flexibilidad, contradicción, audiencia e inmediatez. Así como su incidencia en las decisiones sobre la modificación, sustitución o cese de las medidas. Finalmente, se explora la posible utilización de sistemas de inteligencia artificial como instrumentos de apoyo en la ejecución, identificando los riesgos asociados a la protección de datos, los sesgos y el automatismo decisorio, y proponiendo criterios de compatibilidad basados en un marco de garantías robusto y en la preservación del control humano fuerte sobre la decisión judicial.

Palabras clave

Ejecución medidas, justicia juvenil, principios procesales, inteligencia artificial.

Abstract

The enforcement of measures in Spanish Youth Justice system is a decisive phase within the judicial procedure. Focusing on the Organic Law 5/2000 and its

²⁵⁸ El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

implementing Regulation this chapter firstly analyses the catalogue of measures and the criteria for their individualisation, with particular attention to the distinction between custodial and non-custodial measures and to their specific impact at the enforcement stage. Secondly, it explores the architecture of enforcement, built on the functional allocation of competences between the Youth Court and the public authority for protection and re-education, as well as the role of the Public Prosecutor, the defence lawyer and the Youth Offending Teams. On this basis, the chapter clarifies the main procedural principles governing enforcement: legality, jurisdictionality, the best interest of the child, proportionality, flexibility, adversarial process, the right to be heard and immediacy. Along with how they affect the judicial decisions on modifying, substituting or terminating measures. Finally, it addresses the potential use of artificial intelligence systems as assisting tools in this phase, identifying the risks linked to data protection, discriminatory bias and automation, proposing compatibility criteria grounded in a particularly robust framework of safeguards and the preservation of strong human control over judicial decision-making.

Key words

Enforcement of measures, youth justice, procedural principles, artificial intelligence.

Sumario: INTRODUCCIÓN: EJECUCIÓN DE MEDIDAS EN CLAVE GARANTISTA Y DIGITAL 1. CATÁLOGO DE MEDIDAS Y CRITERIOS DE IMPOSICIÓN CON EFECTO EN LA EJECUCIÓN 2. ARQUITECTURA DE LA EJECUCIÓN: ROLES, INSTRUMENTOS Y PROCEDIMIENTOS 3. PRINCIPIOS PROCESALES RECTORES EN LA EJECUCIÓN DE MEDIDAS 4. INTELIGENCIA ARTIFICIAL Y EJECUCIÓN DE MEDIDAS: COMPATIBILIDAD Y SALVAGUARDAS 4.1. *La Inteligencia artificial como elemento de apoyo en la fase de ejecución* 4.2. *Riesgos procesales y criterios de compatibilidad* 5. CONCLUSIONES BIBLIOGRAFÍA

INTRODUCCIÓN: EJECUCIÓN DE MEDIDAS EN CLAVE GARANTISTA Y DIGITAL

La justicia de menores en España, de inequívoco perfil garantista²⁵⁹, se configura como un modelo formalmente penal y materialmente sancionador-educativo. Lo cual implica una doble finalidad: la exigencia de responsabilidad por los hechos tipificados como delito, y la orientación socioeducativa y de reintegración social de la respuesta jurídica, en la que la proporcionalidad opera al servicio de la finalidad educativa y del interés superior del menor.

²⁵⁹ El Tribunal Constitucional ya abordó el reconocimiento a los menores de todas las garantías procesales que reconoce al Constitución al enjuiciamiento de adultos con carácter general, concretamente en el Fundamento Jurídico 6 de la sentencia 36/1991, de 14 de febrero.

La ejecución de la sentencia firme constituye la materialización de esas finalidades. Si bien el Juzgado de Menores²⁶⁰ ordena las medidas como respuesta jurídico-penal a la conducta delictiva, tanto su adopción como su cumplimiento han de guiarse por el interés superior del menor²⁶¹. En consecuencia, se impone la necesidad de adoptar un enfoque holístico²⁶², que integre todas las dimensiones que afectan la vida del menor: familiar, social, educativa, emocional y de salud, así como una proyección temporal orientada a la efectiva reintegración social.

Desde esta premisa, la responsabilidad penal del menor se ventila a través de un proceso penal especial, con garantías reforzadas y control jurisdiccional.

Una vez desvirtuada la presunción de inocencia, un juez ordinario y predeterminado por la ley, independiente e inamovible²⁶³, estimará al menos parcialmente, la pretensión ejercitada. La reacción jurídica del Ordenamiento consiste en una intervención socioeducativa, que gira en torno al cumplimiento de unas “medidas”²⁶⁴ de distinta intensidad en función de la gravedad de los hechos cometidos y las necesidades del infractor; de modo que toda restricción o limitación de los derechos se concibe al servicio de la reeducación y reintegración social, con reconocimiento y respeto pleno de los derechos del menor.

La progresiva incorporación en todos los ámbitos de herramientas digitales, concretamente el uso de sistemas de apoyo basados en inteligencia artificial (IA) parece conducirnos inexorablemente al uso de estas herramientas en el entorno judicial. En la justicia de menores la IA tiene una gran proyección hacia el ámbito de la valoración del riesgo de reincidencia del menor y la evaluación criminológica del mismo respecto a la adopción de decisiones judiciales en ejecución de las medidas impuestas por la sentencia.

No obstante, el uso de esta tecnología no está en absoluto, exento de riesgos graves como consecuencia del manejo de datos sensibles y la afección de los derechos

²⁶⁰ Sección de menores tras la entrada en vigor de las reformas operadas por la Ley Orgánica 1/2024, de 2 de enero, de medidas en materia de eficiencia del Servicio Público de Justicia.

²⁶¹ Así lo anuncia la Exposición de Motivos de la Ley Orgánica 5/2000 en su expositivo II. Añade ese mismo apartado una mención sobre la necesidad de respetar las garantías derivadas de los derechos constitucionales de carácter procesal, así como de las especiales exigencias derivadas del principio de interés superior del menor.

²⁶² Montero Hernanz, T. *Derecho penal de menores. Una introducción a la legislación reguladora de la responsabilidad penal de los menores*. Reus, 2023, p. 59.

²⁶³ Garantías procesales que exige la Constitución Española en sus artículos 24 y 117.

²⁶⁴ La LORPM utiliza el vocablo “medidas” y no el de “penas” con la clara intención de enfatizar la finalidad socioeducativa y de reintegración que sirven de base al sistema de justicia de menores. En este sentido García Pérez, O. *Medidas y su ejecución en el sistema de justicia penal juvenil*, Tirant lo Blanch, 2019 pp. 43-45.

fundamentales. En consecuencia, la utilización de estos sistemas debe estar sustentada sobre un respeto absoluto a las garantías procesales y derechos del infractor, y en armonía con los principios nucleares del proceso penal del menor²⁶⁵.

Este capítulo se estructura en 3 secciones complementarias. En primer término, se ofrece un análisis sistemático de las medidas previstas en la LORPM con atención a su impacto específico en la fase de ejecución (criterios de graduación, seguimiento y control jurisdiccional). En segundo término, se examinan los principios procesales que informan el proceso penal de menores, con especial atención en el cumplimiento de las medidas, precisando su alcance y límites en la práctica ejecutiva. Finalmente se propone un marco de aplicación de la IA en la ejecución, atendiendo a la compatibilidad y salvaguardas que aseguren la centralidad del interés superior del menor y el respeto íntegro de sus derechos.

2. CATÁLOGO DE MEDIDAS Y CRITERIOS DE IMPOSICIÓN CON EFECTO EN LA EJECUCIÓN

El artículo 7 de la LORPM enumera taxativamente las medidas susceptibles de ser impuestas por el Juzgado de Menores, cuya elección y graduación han de responder a criterios de individualización y al interés superior del menor.

La clasificación doctrinal entre medidas privativas de libertad y no privativas resulta funcional para la fase de ejecución, y coherente con el estándar internacional²⁶⁶ que impone reservar como última ratio y por el tiempo más breve posible la privación de libertad, reafirmando el carácter reeducativo del sistema de justicia juvenil. En efecto, esta categorización de las medidas respeta las directrices recogidas por el

²⁶⁵ Martín Diz, F. *Aplicación de la inteligencia artificial en la evaluación criminológica en materia de ejecución de medidas impuestas a menores infractores*, en Bujosa Vadell y Martín Diz (Dirs), “Menores infractores: predicción, gestión del riesgo e intervención”, Aranzadi, 2022, p. 128.

²⁶⁶ El marco jurídico internacional ofrecía las pautas para el progresivo desarrollo del sistema de justicia juvenil actual. El primer instrumento relevante fue la Declaración de Derechos del Niño adoptada mediante resolución 1386 (XIV) de la Asamblea General de NNUU en 1959. En esa declaración se reconocieron unos principios fundamentales que debían orientar las legislaciones nacionales. El Pacto de Derechos Civiles y Políticos aprobado por la Asamblea General de NNUU por Resolución 2200 A (XXI) de 1966 tuvo un mayor impacto y concreción. En este texto se enuncia el fin reeducativo, así como los derechos procesales esenciales. Posteriormente, las Reglas para la Administración de Justicia a menores, adoptadas por la Asamblea de NNUU por Resolución 40/33 en 1985, conocidas como Reglas de Beijing, explícitamente enuncian el principio de intervención mínima y el de proporcionalidad. Por último, la Convención sobre los Derechos del Niño adoptada por la Asamblea General de NNUU en 1989 recoge todos los principios y derechos de los anteriores instrumentos, dando forma jurídica a la Declaración de 1959. Esta convención es la norma de referencia. En este sentido, la Exposición de Motivos de la LORPM menciona como cimentación de la ley española la Convención.

instrumento internacional de referencia en el ámbito de menores, la Convención sobre los Derechos del Niño de NNUU²⁶⁷.

El internamiento queda reservado a los supuestos de especial gravedad, donde los hechos delictivos son cometidos con violencia o intimidación, o comportan un grave riesgo para la vida o la integridad física²⁶⁸. La privación de libertad del menor se justifica en la necesidad de ubicar al infractor en un entorno diferente, controlado y seguro, donde sea posible encontrar las condiciones pedagógicas idóneas para su reeducación y reorientación, orientando al menor hacia el abandono de conductas antisociales²⁶⁹.

El internamiento se estructura en tres regímenes, cerrado, semiabierto o abierto, diferenciados por el lugar donde se realicen las actividades que integran el programa individualizado de ejecución de medidas (PIEM).

Junto a esas tres modalidades, la ley prevé el internamiento de carácter terapéutico, para supuestos que requieren intervención sanitario-asistencial específica médico-sanitario: deshabituación de alcohol, drogas o fármacos, así como problemas de salud mental. En tal caso, la adscripción a uno u otro régimen (cerrado, semiabierto y abierto) se determina conforme al criterio clínico; los facultativos elaboran un programa de tratamiento y las pautas de seguimiento, que se integran en el PIEM²⁷⁰. Ahora bien, este tratamiento exige la colaboración efectiva del menor, en procesos de deshabituación, el consentimiento y la adherencia a los controles son condición necesaria. Si el menor no consiente, abandona o incumple el plan terapéutico, la entidad pública no iniciará o lo suspenderá y lo comunicará al Juzgado de Menores.

En el régimen de internamiento cerrado, el menor reside en el centro y desarrolla en él las actividades formativas, educativas, laborales y de ocio previstas en su programa individualizado (PIEM). La finalidad es proveerle de los recursos y competencias necesarios para la rectificación de las conductas delictivas y para su reintegración social.

Conforme al artículo 9.2 de la LO 5/2000 se restringe la imposición de esta modalidad a los delitos graves y a los delitos menos graves ejecutados con violencia o intimidación, o que hayan generado grave riesgo para la vida o la integridad física.

²⁶⁷ Trata ampliamente el marco jurídico internacional Vázquez González, C. *Delincuencia Juvenil*, Dykinson, 2019, pp. 411-485; Garrido Carrillo, F. J. *Principios y garantías del proceso penal de menores*. Aranzadi, 2023 pp. 25-74.

²⁶⁸ *Vid.* art. 9 apartados 1 y 2 LORPM.

²⁶⁹ *Vid.* Expositivo III párrafo 16 de la Exposición de Motivos de la LORPM.

²⁷⁰ En efecto, al ser el internamiento terapéutico una modalidad específica de la medida de internamiento requiere intervención de profesionales adecuados a la situación sanitaria del menor en conflicto con la ley. A estos efectos, el art. 27 del Real Decreto 1774/2004, de 30 de julio por el que se aprueba el Reglamento de la Ley Orgánica de Responsabilidad Penal del Menor menciona la elaboración del programa de internamiento por estos profesionales médicos.

Asimismo, también procede el internamiento cerrado cuando los hechos se realizan en grupo o al servicio de una banda u organización criminal, incluso aunque sea con carácter temporal. Del artículo 10 LORPM se deduce la obligación de imponer la medida de internamiento cerrado si los hechos son constitutivos de los delitos tipificados en el CP como homicidio, asesinato, agresión sexual y delitos de terrorismo²⁷¹, con los márgenes y condiciones que la ley establece.

En el régimen de internamiento semiabierto, el menor está obligado a residir en el centro, no obstante, realizará en otras localizaciones una o más de las actividades formativas, educativas, laborales o de ocio incluidas en el PIEM conforme al horario y las condiciones fijadas en dicho programa. Todo ello condicionado a la evolución personal del joven y el cumplimiento de los objetivos fijados en su caso concreto. Es decir, la ejecución se adapta a la evolución del joven y el grado de cumplimiento de los objetivos establecidos, previéndose un margen de ajuste por parte de la entidad pública, que podrá intensificar o reducir las actividades externas modulando sus horarios.

El menor que cumple una medida de internamiento en régimen abierto desarrolla íntegramente fuera del centro las actividades de su proyecto educativo, en los servicios normalizados del entorno. Mantiene, no obstante, la obligación de pernocta en el centro con una permanencia mínima de ocho horas. Esta exigencia puede flexibilizarse por el juez de menores, a propuesta motivada de la entidad pública²⁷².

La permanencia de fin de semana es una modalidad específica de privación de libertad de alcance temporal limitado²⁷³. Cuando se acuerda esta medida, el menor debe permanecer en su domicilio o en el centro de internamiento, durante un intervalo que no podrá exceder de 36 horas, entre el viernes y el domingo. Con ello se persigue, en clave preventiva y educativa, intensificar el control en esos periodos acotados. Es posible combinar la permanencia de fin de semana con otras medidas de restricción de libertad de contenido prestacional o socioeducativo, como las prestaciones en beneficio de la comunidad o tareas formativas, educativas o culturales propuestas por el profesional de la entidad pública.

Junto a las modalidades estrictamente privativas de libertad, la LORPM prevé un conjunto heterogéneo de medidas que, sin comportar ingreso en el centro, restringen

²⁷¹ Homicidio (art. 138 CP), asesinato (art. 139 CP) agresión sexual (arts. 178 a 181 CP) y delitos de terrorismo (arts. 571 a 580 CP).

²⁷² *Id.* art. 26 RLORPM.

²⁷³ Aclara Bujosa Vadell que la LORPM no considera la permanencia de fin de semana una medida de internamiento pura, pese a que sí supone restricción de la libertad del menor, siendo la única diferencia entre la última y la primera el lugar donde el menor deba fijar su domicilio habitual. Bujosa Vadell, L. M. “Ejecución en el enjuiciamiento de menores: Principios aplicables y limitación de la discrecionalidad” en L.M. Bujosa Vadell y F. Martín Diz (Dirs.) *Menores infractores: Predicción, gestión del riesgo e intervención* Thomson Reuters Aranzadi, 2022, p. 21.

de forma relevante la libertad ambulatoria. En estos supuestos, el menor permanece en su entorno familiar o en un centro de acogida, quedando su libertad ambulatoria sometida a ciertas obligaciones y prohibiciones, tales como la asistencia obligatoria a determinados lugares, como un centro sanitario para recibir algún tratamiento ambulatorio, un centro de día, un centro educativo o de formación. Asimismo, la ley contempla restricciones de desplazamiento mediante prohibiciones de acercamiento o de comunicación con determinadas personas.

La medida de tratamiento ambulatorio integra al menor en un itinerario clínico extrainstitucional: debe comparecer²⁷⁴ con la periodicidad fijada en el dispositivo sanitario que determinen los facultativos asignados a su caso, a los efectos de seguir un programa terapéutico específico dirigido tanto a posibles anomalías o alteraciones de la esfera psicológica, como a trastornos de la percepción que afecten a la conciencia, o situaciones de dependencia de alcohol, drogas u otras sustancias tóxicas.

Nos encontramos ante un programa de intervención clínica dirigido a facilitar que el menor pueda superar procesos adictivos o perturbaciones psíquicas, cuya imposición se reserva a los supuestos en que concurren en el joven las condiciones personales que permitan prever un aprovechamiento real del tratamiento. Al igual que sucede con el internamiento terapéutico, el tratamiento ambulatorio en materia de adicciones requiere el consentimiento del menor. Si este no muestra conformidad o no mantiene una participación activa en el seguimiento del programa, el juez habrá de sustituir la medida por otra que se adecúe mejor a sus circunstancias personales.

La libertad vigilada se concibe como una medida de supervisión y control del cumplimiento de las directrices fijadas en el programa individualizado de ejecución de medidas. Implica el acompañamiento continuado por parte de un profesional de la entidad pública quien mantendrá con el menor entrevistas periódicas de carácter obligatorio. A través de estas entrevistas se comprueba el grado de observancia de las reglas de conducta impuestas judicialmente, que pueden incluir, entre otras, la asistencia regular a un centro docente para completar la enseñanza obligatoria; la participación en actividades formativas de naturaleza cultural, profesional, en materia de educación sexual o educación vial; la obligación de residir en un domicilio determinado; la prohibición de ausentarse del mismo sin autorización judicial previa; o la prohibición de acudir a ciertos espacios, establecimientos o eventos. Las actuaciones socioeducativas que conforman esta medida persiguen en última instancia, reforzar las competencias personales y sociales del menor, promoviendo procesos de aprendizaje que favorezcan su integración y reduzcan el riesgo de reincidencia.

²⁷⁴ Nótese que la LO 5/2000 regula dos tipos de medidas terapéuticas diferentes, el internamiento y el tratamiento ambulatorio. La primera implica una privación de libertad del menor, mientras que la segunda consiste en una limitación de su derecho a la libertad ambulatoria, que le obliga a acudir al establecimiento sanitario para ser tratado de la dolencia.

Otra de las medidas que restringen la libertad ambulatoria es la asistencia a centro de día. Su objetivo consiste en incorporar al menor a un entorno estructurado²⁷⁵ y adecuado para el desarrollo de su proyecto socioeducativo, durante la mayor parte del día. El menor debe acudir al centro integrado en la red de servicios sociales de la comunidad autónoma²⁷⁶ en la que resida, donde realizará total o parcialmente de las actividades formativas, educativas, laborales o de ocio previstas. Corresponde a la entidad pública seleccionar, ente los recursos próximos al lugar de residencia del menor, aquellos servicios más adecuados, atendiendo a las circunstancias personales del joven y a la concreta ejecución de las medidas impuestas.

La LO 5/2000 faculta al juez a acordar la prohibición de aproximarse o comunicarse respecto de la víctima, sus familiares u otras personas identificadas. Mediante esta medida se veda al menor la posibilidad de acercarse físicamente a esas personas, con independencia del lugar en que se encuentren, como al de entablar contacto con ellas por cualquier vía escrita, verbal o visual²⁷⁷.

La LORPM contempla, por último, como medida restrictiva de libertad la convivencia del menor con otra persona, familia o grupo educativo por el tiempo que judicialmente se establezca. A través de esta convivencia se le proporciona un entorno socioafectivo positivo, orientado a facilitar su proceso de reinserción social. Las personas que se ofrecen voluntariamente para acoger son objeto de una selección cuidadosa por parte de la entidad pública, que verifica que reúnen las condiciones personales, familiares y económicas idóneas para la correcta ejecución de la medida²⁷⁸. Corresponde al juez de menores decidir si el menor puede mantener el contacto con su familia.

La LORPM contempla otro tipo de medidas menos intensas en términos de restricción de libertad como las anteriores que consisten en la realización de tareas que el

²⁷⁵ Montero Hernanz, T. *Derecho penal de menores op. cit.* p. 127.

²⁷⁶ *Vid.* art. 17.3 RLORPM.

²⁷⁷ El art. 7.1 i) LORPM resuelve el conflicto que podría surgir cuando para ejecutar esta medida de prohibición de aproximarse o comunicarse con la víctima, sus familiares u otras personas, el menor se ve obligado a cambiar de residencia, y dejar de convivir con los padres. Ordena la ley que el MF debe remitir testimonio a la entidad pública de protección para que promueva las medidas de protección del menor que considere adecuadas a la situación del joven infractor, conforme a la Ley Orgánica 1/1996 de 15 de enero, de Protección Jurídica del Menor de modificación parcial del Código Civil y de la Ley de Enjuiciamiento Civil.

²⁷⁸ *Vid.* art. 19 RLORPM.

menor debe aceptar²⁷⁹ y carecen de contraprestación económica²⁸⁰. Se distinguen aquí dos modalidades: las actividades de carácter socioeducativo, orientadas a la adquisición de competencias personales y sociales²⁸¹ que favorezcan su reeducación y reinserción; y las prestaciones en beneficio de la comunidad, orientadas a procurar un interés social o realizadas en beneficio de personas en situación de precariedad. Aunque la LO 5/2000 no especifica el contenido de dichas tareas²⁸², el Reglamento²⁸³ vincula preferentemente su diseño a la naturaleza del bien jurídico lesionado, de modo que exista correspondencia entre la intervención y el desvalor de la conducta. Finalmente, tras la reforma operada por la Ley Orgánica 10/2022 de 6 de septiembre, (en vigor desde el 7 de octubre de ese mismo año), cuando las medidas traigan causa de delitos que atentan contra la libertad sexual, en particular, los tipificados en los arts. 178 a 183 del CP, el juez necesariamente, y con carácter accesorio, debe imponer la obligación al menor de participar en programas formativos de educación sexual y educación en igualdad²⁸⁴.

La amonestación constituye una de las medidas de menor intensidad. Consiste en una advertencia por parte del tribunal, con una finalidad eminentemente pedagógica, en la que se explicita al menor infractor la gravedad de la conducta, así como sus consecuencias tanto para él como para la víctima, instándole no repetir estos comportamientos en el futuro.

Finalmente, la LORPM contempla medidas de privación de derechos distintos a la libertad ambulatoria. En esta categoría encontramos la privación del permiso de conducir ciclomotores y vehículos a motor, o del derecho a obtenerlo; así como la privación de licencias administrativas de uso de armas y de caza. Estas medidas pueden acordarse con carácter accesorio a otras impuestas en procesos por delitos vinculados a la seguridad vial y a las armas.

²⁷⁹ En relación con el necesario consentimiento del menor, Bujosa Vadell apunta que la falta del mismo implicaría una vulneración de la prohibición constitucional de trabajos forzados del art. 25.2 CE. Bujosa Vadell, L.M. *La ejecución en el enjuiciamiento de menores op. cit.* p. 22.

²⁸⁰ La LO 5/2000 no lo plantea el asunto de la falta de retribución, sin embargo, el Reglamento aclara que la realización de estas prestaciones en beneficio de la comunidad no puede estar supeditada a intereses económicos, y en consecuencia no podrán ser retribuidas. Sin embargo, se podrá indemnizar al menor que las ejecute en relación con los gastos de transporte, manutención en su caso. No siendo posible percibir esta indemnización si se realizan en la entidad pública o esta se hace cargo de asumir ese coste. Art. 20.2 d) y 20.3 RLORPM.

El art. 7.1 k) LORPM no menciona esta conexión de las medidas impuestas con el bien jurídico protegido por el delito cometido.

²⁸¹ Montero Hernanz, T. *Derecho penal de menores op. cit.* p.138.

²⁸² El art. 7.1 k) LORPM no añade esta conexión entre las medidas impuestas y el bien jurídico protegido por el delito cometido.

²⁸³ *Vid.* art. 20 RLORPM.

²⁸⁴ *Vid.* art. 7.5 LORPM.

Asimismo, la inhabilitación absoluta comporta la privación de honores, empleos o cargos públicos de todo tipo, y la imposibilidad de acceder a ellos; cuya ejecución corresponde a la autoridad administrativa competente, salvo cuando son ejecutadas directamente por el tribunal²⁸⁵.

Excepto la medida de la libertad vigilada, la prohibición de acercarse o comunicarse con la víctima u otras personas y la convivencia con personas designadas, que se pueden adoptarse cautelarmente, el resto de las medidas restrictivas de libertad únicamente podrán imponerse mediante sentencia firme.

De entre el amplio y variado catálogo de medidas, el juez de menores individualiza la respuesta seleccionando una o varias, de distinta naturaleza, idóneas para el menor concreto. A tal fin, no solo considerará la prueba y calificación jurídica de los hechos, sino además pondera de forma destacada la edad, las circunstancias familiares y sociales, la personalidad y el interés del menor, con apoyo en los dos últimos informes del equipo técnico y de las entidades públicas de protección y reforma²⁸⁶. Esta elasticidad del sistema permite una valoración holística de la situación del menor, sin confundir el principio de flexibilidad con el de arbitrariedad. La sentencia debe ofrecer una motivación²⁸⁷ suficiente que justifique tanto la elección de la medida o medidas ordenadas, como su duración, desde la perspectiva del interés del menor²⁸⁸.

3. ARQUITECTURA DE LA EJECUCIÓN: ROLES, INSTRUMENTOS Y PROCEDIMIENTOS

De acuerdo con el principio de legalidad, la ejecución de medidas solo puede tener lugar en virtud de sentencia firme²⁸⁹ que estime total o parcialmente la pretensión ejercitada, y se dicte conforme a lo dispuesto en la LORPM²⁹⁰. En dicha resolución condenatoria han de consignarse de forma expresa los hechos probados y los medios probatorios que han permitido al órgano judicial alcanzar un grado de convencimiento suficiente para desvirtuar la presunción de inocencia, así como la medida o medidas impuestas, con indicación de su contenido, duración y objetivos que se persiguen con

²⁸⁵ Montero Hernanz, T. *Derecho penal de menores op. cit.* p. 140.

²⁸⁶ *Ibid.* p. 146.

²⁸⁷ *Vid.* art. 7.3 LORPM.

²⁸⁸ Establece la ley algunas limitaciones relevantes a la flexibilidad mencionada. En ocasiones el juez se verá limitado a elegir una medida determinada para ciertos actos delictivos. Los artículos 9 a 11 establecen esos contextos. A modo de ejemplo: las acciones u omisiones imprudentes no se pueden sancionar con internamiento en régimen cerrado (art. 9.4).

²⁸⁹ Martín Diz, F. “Aplicación de la inteligencia artificial en la evaluación criminológica en materia de ejecución de medidas impuestas a menores” en L.M. Bujosa Vadell y F. Martín Diz (Dirs.) *Menores infractores, predicción, gestión del riesgo e intervención*, Thomson Reuters Aranzadi, 2022, p. 131.

²⁹⁰ Así se lo enuncia el art. 43.1 LORPM. Lo identifica como garantía jurisdiccional Garrido Carrillo, F. J. *Principios y garantías op. cit.* p. 89.

su cumplimiento. Todo ello debe formularse con un lenguaje accesible y comprensible para el menor²⁹¹, que le permita conocer el alcance de la decisión y de las consecuencias que de ella se derivan.

Los artículos 44 y 45 de la LORPM articulan un reparto funcional de competencias entre el órgano judicial y la entidad pública de protección o reforma. La segunda asume la ejecución material de las medidas impuestas, mientras que el Juzgado de Menores mantiene, desde el inicio y durante toda la fase de cumplimiento de la sentencia, un control jurisdiccional que se proyecta sobre las decisiones esenciales del proceso ejecutivo. Así, corresponde al juez de menores la adopción de múltiples decisiones relativas al desarrollo y la evolución de la ejecución de las medidas impuestas por sentencia firme, decisiones que puede adoptar de oficio o a instancia del Ministerio Fiscal o del Letrado del menor.

El artículo 44.2 LORPM enumera las actuaciones reservadas a la autoridad judicial: la resolución de las propuestas de revisión de las medidas, la aprobación del programa individualizado de ejecución (PIEM), la recepción y valoración de los informes de seguimiento, la resolución de los recursos que se interpongan contra las resoluciones dictadas en la fase de ejecución, la decisión sobre las peticiones o quejas del menor, realizar regularmente visitas a los centros de internamiento y entrevistas con los menores sancionados, ofrecer sugerencias y recomendaciones a la entidad pública en relación con la organización y régimen de ejecución de las medidas, resolver las cuestiones disciplinarias. En definitiva, la ley configura una intervención judicial activa y continuada que permite garantizar el respeto a los principios de legalidad²⁹², proporcionalidad y flexibilidad²⁹³.

La ejecución de las medidas se concibe como un proceso dinámico sometido a un control jurisdiccional intenso, en cuyo marco las herramientas tecnológicas ya tienen una presencia significativa, y como se verá más adelante, están llamadas a asumir una significación mayor en relación con la mejora de la eficacia del sistema.

En esta arquitectura, el Letrado de la Administración de Justicia (LAJ) desempeña un papel específico, al LAJ le corresponde practicar la liquidación de la medida una vez aprobado judicialmente el PIEM²⁹⁴. En su liquidación se fijan las fechas de inicio y fin de la medida, así como, en su caso, el abono de periodos ya cumplidos. El Reglamento refuerza de nuevo el control jurisdiccional al exigir que la liquidación efectuada

²⁹¹ *Vid.* art. 39 LORPM.

²⁹² El principio de legalidad en la LORPM es ampliamente tratado por Garrido Carrillo, F. J. *Principios y garantías op. cit.* pp. 86-92.

²⁹³ Martín Diz, F. *Aplicación de la inteligencia artificial op. cit.* p. 147.

²⁹⁴ Es el RLORPM el que detalla en los artículos 9 y 10 cómo comienza el proceso de ejecución de las medidas, concretamente: aprobación del PIEM art. 10.1 5º; aprobación de liquidación del LAJ art. 10.1 7º segundo párrafo.

por el LAJ sea sometida a la aprobación del juez de menores, previa emisión de un informe del Fiscal y del Letrado del menor.

El Ministerio Fiscal y la defensa intervienen también de forma significativa en la fase de ejecución. El fiscal puede instar actuaciones ante el Juzgado de Menores²⁹⁵ cuando lo estime necesario para la correcta ejecución de la sentencia o para la protección del interés del menor o de la víctima. Asimismo, el fiscal recibe, junto con el juez, los informes de seguimiento remitidos por la entidad pública.

El abogado defensor, por su parte, puede formular solicitudes relativas al desarrollo de la ejecución; obtener copia de los informes, siempre que se lo solicite expresamente a la entidad pública y promover la revisión de medidas en los términos previstos en la ley.

Además, tanto el Ministerio Fiscal como la defensa, deben ser oídos en los procedimientos de modificación o sustitución de las medidas²⁹⁶ y la defensa del menor está legitimada para recurrir en apelación²⁹⁷ los autos que se dicten en esta fase, de acuerdo con el régimen general de recursos.

La entidad pública de protección o reforma constituye el otro pilar de la ejecución. Le corresponde la ejecución material de las medidas acordadas por el Juzgado de Menores. A tal efecto, recibe la ejecutoria, los informes del equipo técnico, la identificación del Letrado de la defensa y cuantos testimonios considere relevantes el juez a efectos de ejecución. La entidad pública de la Comunidad Autónoma competente procede a la designación de un profesional responsable del caso, extremo que comunica al órgano judicial, y dicho profesional elabora el PIEM²⁹⁸, que concreta la forma de cumplimiento de la medida.

Paralelamente, la entidad pública abre un expediente personal único para cada menor, en el que se integran tanto los documentos remitidos al tribunal, como los elaborados por los profesionales que intervienen en la ejecución. Toda esta información se custodia con gran cautela por quienes tienen acceso a la misma.

²⁹⁵ *Vid.* art. 44.3 LORPM.

²⁹⁶ *Vid.* arts. 50.2 y 51.1 LORPM.

²⁹⁷ De la redacción de los artículos 41, 52 y 42.3 de la LORPM se puede inferir que el Ministerio Fiscal y el abogado de la defensa están legitimados para recurrir las resoluciones dictadas en ejecución, aunque formalmente el art. 52 concede legitimación para recurrir al menor a través de su Letrado. No obstante, el fiscal aparece como parte recurrente en casación (art. 42.3 LORPM) y conforme al régimen general de los recursos penales.

²⁹⁸ Los detalles sobre el expediente personal único quedan regulados en el art. 12 RLORPM.

Asimismo, la entidad pública debe remitir al juez y al Ministerio Fiscal informes de seguimiento²⁹⁹ en los que se detallan las incidencias y la evolución personal del menor durante el cumplimiento de la medida, con la correspondiente remisión al Letrado de la defensa cuando lo solicite.

Tales informes presentan una periodicidad ordinaria, que con carácter general es trimestral, aunque el propio texto legal y reglamentario contemplan frecuencias específicas para determinadas medidas, como la permanencia de fin de semana o las prestaciones en beneficio de la comunidad³⁰⁰. Se completan con un informe final al término de la ejecución y con informes extraordinarios cuando se produce alguna incidencia relevante, como la fuga, la falta de asistencia o el no retorno al domicilio en el momento establecido³⁰¹.

Sobre esta base fáctica y documental, configurada por los roles de los distintos operadores y por los instrumentos sobre los que se asienta (PIEM, expediente personal e informes de seguimiento), resulta necesario examinar, en coherencia con el título de este epígrafe, los procedimientos a través de los cuales se articula la revisión, modificación y sustitución de las medidas impuestas por sentencia firme.

A partir de los informes de seguimiento puede articularse una clasificación funcional, atendiendo a si incorporan o no una propuesta de revisión de la medida³⁰². Ahora bien, la iniciativa para revisar lo acordado en sentencia no se agota en la actividad de la entidad pública. La solicitud de revisión puede ser promovida también por el Ministerio Fiscal, por el Letrado del menor o por la propia Administración competente, o incluso puede ser acordada de oficio por el juez³⁰³. La modificación de las medidas inicialmente impuestas puede adoptar, en esencia, tres formas: dejar sin efecto la medida, reducir su duración o sustituirla por otra que se considere más adecuada, siempre con sujeción prioritaria al interés del menor.

La primera de las alternativas (dejar sin efecto la medida), pese a no constar así expresamente en el art. 51 LORPM, solo resulta justificada cuando se ha alcanzado el

²⁹⁹ *Vid.* arts. 49.1 LORPM y 13 RLORPM.

³⁰⁰ Respecto a la ejecución de la medida de permanencia de fin de semana, tanto en el domicilio como en un centro, los informes deben ser remitidos transcurridos cuatro fines de semana; y para la medida de prestaciones en beneficio de la comunidad se debe remitir cada 25 horas de prestación si la medida impuso un total de horas inferior a 50, o un informe cada 50 horas si la sanción establece prestaciones por tiempo superior.

³⁰¹ En efecto, el artículo 14 del RLORPM regula estos incumplimientos causados por fuga o falta de presentación en el lugar indicado, ya sea el centro o el domicilio, o incluso no acudir a una entrevista o al lugar donde se debe realizar la actividad socioeducativa o la prestación en beneficio de la comunidad.

³⁰² Montero Hernanz, T. *Derecho penal de menores op. cit.* p. 172.

³⁰³ El artículo 51 LO 5/2000 regula la sustitución de las medidas en concordancia con el art. 13 regulador de los informes.

objetivo propuesto antes de que finalice el tiempo inicialmente fijado en la sentencia, o en sentido inverso, cuando la medida deviene ineficaz para la consecución de la finalidad para la que fue impuesta. En ambos supuestos el juez de menores ha de acordar el cese de la medida mediante auto motivado, en el que se expongan las razones que fundamentan la decisión. Con todo, la ley sí contempla de forma expresa³⁰⁴ la posibilidad de dejar sin efecto la medida en los supuestos de conciliación entre agresor y víctima³⁰⁵.

La reducción temporal de la medida se justifica cuando la evolución del menor durante el tiempo de ejecución ha sido claramente positiva y pone de manifiesto que los objetivos inicialmente previstos pueden considerarse alcanzados en un lapso temporal más breve que el fijado en la sentencia. En tales circunstancias, la continuación de la medida en sus términos originarios dejaría de resultar desproporcionada, lo que legitima su abreviación.

Por último, el juez de menores puede de oficio o a instancia de la entidad pública, sustituir las medidas acordadas en la sentencia por otra que resulte más adecuada y eficaz, siempre que la nueva medida hubiera podido ser impuesta inicialmente atendiendo a los delitos cometidos, y que su duración no exceda el tiempo de ejecución previsto para la medida originaria. La sustitución puede operar tanto en el sentido de pasar de un régimen más restrictivo o limitativo de derechos a otro menos restrictivo; como en sentido contrario. Esta segunda dirección (el tránsito hacia una medida más restrictiva) únicamente es posible cuando, de acuerdo con las previsiones legales, la medida sustitutoria pudo ser impuesta desde el inicio, y los hechos delictivos se encuadran en los supuestos del art. 9.2 LORPM (delitos graves, delitos menos graves cometidos con violencia o intimidación, hechos que hayan generado un grave riesgo para la vida o a la integridad física, o delitos cometidos en grupo o al servicio de una banda)³⁰⁶. De este modo, la sustitución puede dar lugar a la aplicación de una medida más restrictiva si la evolución del menor es desfavorable³⁰⁷ o, por el contrario, a una medida

³⁰⁴ Tanto el art 51.3 LORPM como el art. 15 RLORPM mencionan esta opción, dejando al juez valorar si es conveniente a la luz del tiempo de ejecución de la medida y el propio acto de conciliación son manifestación suficiente del reproche que merecen los actos delictivos cometidos por el menor.

³⁰⁵ Analizan la mediación y la justicia restaurativa en el ámbito de la justicia de menores Jullien de Asís, J. “Reparación de las víctimas en la justicia juvenil” en *La víctima en el proceso penal de menores. Tratamiento procesal e intervención socioeducativa*, Farto Piay, T. (Coord.), Dykinson, Madrid, 2021 p.257, y Palma Chapparra, L. “La mediación” en Monge Fernández. A. (Dir.) *La protección jurídica del menor*, Tirant Lo Blanch, Valencia, 2024, pp. 277-333.

³⁰⁶ Martín Diz, F. *Aplicación de la inteligencia artificial op. cit.* p. 133.

³⁰⁷ Nótese que nos referimos a “evolución desfavorable” que no necesariamente implica quebrantamiento de la medida inicialmente impuesta. Este segundo supuesto, está contemplado en el art. 50.2 LORPM en términos similares, pero con ciertos matices diferenciales en relación con la posibilidad de sustituir una medida menos restrictiva por otra más severa si se produce un quebrantamiento. En efecto, dicho precepto permite al juez de menores, en casos excepcionales, y a instancia del Ministerio Fiscal, oídos el Letrado y representante legal del menor, así como el equipo técnico, sustituir una medida de

menos restrictiva cuando la ejecución de la medida original ha tenido un efecto positivo, y el infractor presenta una evolución favorable, siempre que la medida sustitutiva resulte jurídicamente aplicable al hecho delictivo cometido.

A título ilustrativo, si en la sentencia se impuso una medida de internamiento en régimen cerrado, el juez podrá sustituirla por un internamiento en régimen semiabierto, cuando la evolución del menor sea favorable y así lo aconseje la consecución de los fines reeducativos. No obstante, si tras la sustitución la evolución del infractor resulta negativa y el menor no progresa conforme a los objetivos fijados, el juez podrá dejar sin efecto la sustitución y restablecer el internamiento en régimen cerrado, retomando el régimen originario de cumplimiento.

El procedimiento para la sustitución de las medidas exige que se oiga a las partes, al equipo técnico y en su caso, al representante de la entidad pública de protección o reforma. Se resuelve mediante auto motivado susceptible de ser recurrido en apelación³⁰⁸.

La sustitución de las medidas constituye, en suma, una manifestación paradigmática de la flexibilidad que preside no solo la ejecución de la sentencia sino todo del proceso penal de menores, al permitir adaptar la respuesta sancionadora-educativa a la evolución real del infractor³⁰⁹. Esta flexibilidad, que en el plano formal se plasma en los mecanismos de revisión, reducción o sustitución descritos, será objeto de análisis específico en relación con el ámbito administrativo de la ejecución de medidas a cargo de la entidad pública.

4. PRINCIPIOS PROCESALES RECTORES EN LA EJECUCIÓN DE MEDIDAS

La arquitectura de la ejecución de medidas descrita en los epígrafes anteriores se encuentra atravesada por un haz de principios procesales que condicionan de manera decisiva el margen de actuación del juez de menores. En esta fase, las decisiones

libertad por otra de internamiento en régimen semiabierto, en caso de quebrantamiento. Esta previsión de la ley ha suscitado mucha polémica en torno al principio de legalidad en la ejecución. Diversos autores han manifestado que el art. 50.2 LORPM vulnera el principio de proporcionalidad, así como el principio de legalidad. No obstante, coincidimos con Garrido Carrillo en puntualizar que la sustitución no constituiría vulneración alguna si la sentencia emitida por el juez de menores incluyese ambos tipos de medidas, tanto la privativa de libertad como la no privativa, habiendo quedado en suspenso la primera a expensas del correcto cumplimiento de la segunda. Garrido Carrillo, F. J. *Principios y garantías op. cit.* pp. 90-92 citando a Olmedo Cardenete, M. “Reglas generales y especiales para la ejecución de medidas en la LORPM. Especial referencia al quebrantamiento de la medida de privación de libertad” en *El menor como víctima y victimario de la violencia social*, Morillas Cueva, L. (Dir.) Madrid, Dykinson, 2010, p. 508.

³⁰⁸ *Vid.* art. 13 y 41.3 LORPM.

³⁰⁹ Coincidimos plenamente con el parecer de Bujosa Vadell, L.M. *La ejecución en el enjuiciamiento de menores op. cit.* p. 26.

relativas a la modificación, sustitución y verificación del cumplimiento de las medidas impuestas en sentencia no constituyen un espacio libre de discrecionalidad, sino que han de adoptarse bajo el prisma de la legalidad, la jurisdiccionalidad, el interés superior del menor, la proporcionalidad y la flexibilidad, entre otros.

La puesta en práctica de estos principios resulta especialmente relevante en un ámbito, como el de la ejecución, caracterizado por la necesidad de adaptación constante a la evolución del menor y, cada vez más, por la progresiva incorporación de herramientas tecnológicas de apoyo a la decisión.

En esta fase, los principios de legalidad y jurisdiccionalidad constituyen el eje sobre el que se articula la adopción y el eventual ajuste de las medidas sancionadoras impuestas al menor infractor³¹⁰. La ejecución se concibe como una extensión jurisdiccional del proceso penal de menores, en la que el juez mantiene su posición central de control y garantía en relación con las decisiones que afectan al contenido, intensidad y duración de las medidas.

Estos dos principios no operan solo como parámetro técnico, sino que se configuran como verdaderos valores de fondo que orientan y limitan la actuación del juez. En el ámbito de la ejecución, encontramos la concreta positivación de la legalidad y la jurisdiccionalidad en los artículos 43³¹¹ y 44 LORPM. Atribuyen al Juzgado de Menores la competencia para la adopción y el control de medidas, exigiendo que toda respuesta sancionadora se incardine en el catálogo legalmente predeterminado. Asimismo, la LORPM prevé un cauce específico para su revisión, modificación o sustitución, sometido al imperio de la ley y al control jurisdiccional efectivo, incluso cuando su desarrollo material se encomienda a la entidad pública de protección o reforma.

El Tribunal Constitucional ha configurado el interés superior del menor como un principio de rango constitucional, anclado en el artículo 39 de la Constitución Española³¹² y en el 3.1 de la Convención sobre los Derechos del Niño³¹³. Este concepto jurídico indeterminado, que la jurisprudencia ha ido dotando de contenido³¹⁴, exige que toda decisión que afecte al menor tome como consideración primordial el interés del menor. Asimismo, este principio se configura como un parámetro de orden público

³¹⁰ Martín Diz, F. *Aplicación de la inteligencia artificial op. cit.* p. 131.

³¹¹ Identifica Garrido Carrillo tres exigencias relativas al principio de legalidad en el ámbito de la ejecución de medidas: seguir un procedimiento legalmente previsto; la decisión sobre la ejecución corresponde al juez de menores; quien emite una sentencia motivada cuyo cumplimiento no se puede iniciar hasta que la misma adquiere firmeza. Garrido Carrillo, F. J. *Principios y garantías op. cit.* p. 89.

³¹² Sentencia del Tribunal Constitucional 178/2020, de 14 de diciembre.

³¹³ La Convención sobre los Derechos del Niño de NNUU fue probada por resolución 44/25 de 20 de noviembre de 1989.

³¹⁴ Bujosa Vadell, L.M. *La ejecución en el enjuiciamiento de menores op. cit.* p. 29 citando a Rivero Hernández F. *El interés del menor*, Dykinson, 2007 p 131.

que puede justificar la atenuación de la rigidez de las normas procesales, y en su caso el sacrificio de otros intereses jurídicos en conflicto. Ello impone a los órganos jurisdiccionales un canon de motivación especialmente intenso, puesto que deben explicar la ponderación de las circunstancias concretas del caso y las razones por las que la solución adoptada se estima más beneficiosa para el menor³¹⁵.

El interés superior del menor se concibe en nuestro ordenamiento jurídico como un principio jurídico, un derecho y una norma de procedimiento³¹⁶. En la fase de ejecución se erige como el punto de referencia central³¹⁷ a partir del cual han de interpretarse y adoptarse las decisiones que inciden sobre las medidas impuestas. Este principio no se agota en la proclamación abstracta de una finalidad reeducativa, sino que exige que cada decisión sobre el mantenimiento, la reducción, el cese o la sustitución de una medida se adopte tras una valoración individualizada de la situación del menor, de su evolución durante la ejecución y de sus necesidades concretas. Desde esta perspectiva, la ejecución no puede limitarse a comprobar el mero cumplimiento formal de las medidas, sino que ha de cuestionarse constantemente si la modalidad e intensidad de la intervención siguen siendo las más adecuadas para favorecer la maduración personal del menor, la responsabilización por el hecho cometido y la reintegración social del joven infractor³¹⁸, incluso cuando ello obligue a reajustar el contenido inicialmente fijado en la sentencia.

Desde esta perspectiva, la proporcionalidad actúa en la ejecución como un criterio de equilibrio entre la necesidad de exigir responsabilidad penal del menor, y la obligación de modular la intervención a las necesidades educativas concretas del joven infractor³¹⁹.

El catálogo de medidas previsto en la LORPM y las posibilidades legales de revisión, modificación o sustitución durante la ejecución ofrecen al juez un margen de individualización considerable, pero no arbitrario, pues tiene la obligación de justificar la imposición inicial de la medida, así como cualquier modificación que considere adecuada. En definitiva, la ley equilibra la intervención penal sobre el menor infractor,

³¹⁵ La Sentencia del Tribunal Constitucional 178/2020, de 14 de diciembre, en los FFJJ 3 y 4 explicita la ponderación de las circunstancias del caso y los derechos en conflicto.

³¹⁶ Garrido Carrillo, F. J. *Principios y garantías op. cit.* p. 109.

³¹⁷ *Ibid.* p. 113.

³¹⁸ Como apunta Colás Turégano la respuesta que persigue el proceso penal de menores especialmente es contribuir a que el menor supere los motivos que le han llevado a delinquir, utilizando la solución que resulte más educativa. Colás Turégano, M.A. *Hacia una humanización de la justicia penal: la mediación en la justicia juvenil española. Principios y ámbito aplicativo en la LO 5/2000, reguladora de la responsabilidad penal de los menores* Revista Boliviana de Derecho n° 20, 2015 p. 159.

³¹⁹ En este sentido Bujosa Vadell, L.M. *Proceso Penal de menores y derecho comparado: Novedades y desafíos* en Martín Ríos, P. y Pérez Martín M.A. (Dir.) "La administración de justicia en España y América" Astigi, 2021 p. 276.

articulando mecanismos destinados a evitar que la respuesta del ordenamiento jurídico resulte desproporcionada en relación con los fines socioeducativos propios de la justicia de menores.

Íntimamente ligado a la proporcionalidad se sitúa el principio de la flexibilidad³²⁰, que la ley erige como rasgo distintivo del proceso penal de menores y que se proyecta con especial intensidad en la fase de ejecución. La posibilidad de revisar, modificar, dejar sin efecto o sustituir las medidas, así como modular sus condiciones de cumplimiento, obedece a la necesidad de ajustar en cada momento la intervención a la evolución del menor y a su contexto familiar y social³²¹. Esta flexibilidad actúa como técnica de articulación dinámica del interés superior del menor, de la finalidad reeducativa y de la proporcionalidad, evitando respuestas estáticas que ignoren los cambios que se producen en las necesidades del infractor durante el cumplimiento de la medida.

Ahora bien, la flexibilidad se despliega dentro de contornos normativos precisos y no puede confundirse con una facultad discrecional ilimitada. El juez de menores está limitado al catálogo tasado de medidas previstas en la LORPM y a sus limitaciones de contenido y duración³²², de modo que cualquier revisión, modificación o sustitución debe respetar tanto la tipología de las medidas como los límites máximos legalmente establecidos. Del mismo modo, las eventuales modificaciones en sentido más restrictivo solo son admisibles cuando la propia ley lo autoriza expresamente y siempre que la nueva medida hubiera podido imponerse desde el inicio, en particular en los supuestos más graves contemplados en la normativa. A ello se añade la exigencia de motivación³²³ y la posibilidad de recurso, así como la intervención del Ministerio Fiscal, del Letrado de la defensa y del equipo técnico, que actúan como contrapesos imprescindibles para impedir que la flexibilidad derivada del interés superior del menor y de la finalidad reeducativa degenere en decisiones arbitrarias.

Junto a estos principios que orientan el contenido y la intensidad de la intervención, en la fase de ejecución adquieren especial relevancia las garantías estructurales del proceso, en cuanto conforman el propio procedimiento como una garantía en sí

³²⁰ Resulta muy interesante la reflexión de Colás Turégano sobre la consideración de excepción al principio de flexibilidad en la decisión, imposición y ejecución de las medidas la obligación incluida en el art. 10 de la LORPM de imponer el internamiento en régimen cerrado para los delitos sexuales. Colás Turégano, M.A. “Punitivismo y justicia de menores: La reforma de la Ley Reguladora de la Responsabilidad Penal de los Menores (LO 5/2000) por la Ley del “solo sí es sí” (LO 10/2022 de Garantía Integral de la Libertad Sexual”, *Revista Electrónica de Ciencia Penal y Criminología*. 25-25 (2023) p. 7.

³²¹ En este sentido Blanco Barea, J. A. “Responsabilidad penal del menor: principios y medidas judiciales aplicables en el Derecho Penal español”, *Revista de Estudios Jurídicos*, nº 8/2008 (Segunda Época) p. 15

³²² En este sentido se pronuncia Cano Paños al analizar si resultaría adecuado endurecer las sanciones en la LO 5/2000. Cano Paños, M.A. “¿Es conveniente un endurecimiento del Derecho penal juvenil? Una toma de posición crítica”, *Anuario de Derecho penal y Ciencias Penales*, vol. LV 2002, p. 306.

³²³ *Vid.* art. 7.3 LORPM.

mismo³²⁴. La contradicción y el derecho de audiencia, junto con la igualdad de partes y la dualidad de posiciones resultan particularmente significativos en relación con las decisiones judiciales durante la ejecución. La posibilidad de revisar, sustituir o dejar sin efecto las medidas no puede articularse como un mero trámite burocrático entre el órgano judicial y la entidad pública, sino que se trata de una actuación que necesariamente debe preservar un espacio efectivo para la intervención del menor, de su defensa y del Ministerio Fiscal. Por ello, las decisiones que inciden sobre el contenido o la intensidad de las medidas deben adoptarse, como regla general, previa audiencia de las partes y del equipo técnico, de modo que las valoraciones profesionales sobre la evolución del menor y las alegaciones de quienes ejercen su representación y defensa puedan ser tenidas en cuenta por el juez antes de redefinir las condiciones de cumplimiento.

En este sentido, la LORPM articula un esquema de garantías que proyecta al ámbito de la ejecución el derecho del menor a ser oído en todo procedimiento judicial que le afecte. La previsión de informes periódicos y extraordinarios de la entidad pública, su remisión al juez y al Ministerio Fiscal, la posibilidad de acceso a dicha documentación por el Letrado del menor, así como la exigencia de motivación de las decisiones que modifiquen en algún sentido las medidas y previa audiencia de las partes y del equipo técnico, dotan de contenido real al principio de contradicción y audiencia en la fase de ejecución.

En suma, el principio de contradicción y el derecho de audiencia del menor encuentran anclaje normativo claro en la LORPM³²⁵, cuando exige la emisión periódica de informes por la entidad pública y su remisión al juez y al Ministerio Fiscal³²⁶, así como la audiencia a las partes y la resolución motivada por auto para cualquier modificación, reducción o sustitución de las medidas inicialmente impuestas; previsiones que se complementan con el desarrollo reglamentario del deber de informar sobre la evolución del menor³²⁷, y con el derecho del menor a ser oído en todo procedimiento judicial que le afecte³²⁸.

Finalmente, estas garantías se completan con la exigencia de una cierta inmediatez judicial en la fase de ejecución. La LORPM no concibe al juez de menores como un mero homologador de las propuestas formuladas por la entidad pública, sino como un órgano que mantiene el control efectivo sobre el cumplimiento de las medidas. De

³²⁴ Garrido Carrillo, F. J. *Principios y garantías op. cit.* p. 80-81.

³²⁵ En este sentido Bernuz Beneitez, M. “El derecho a ser escuchado: el caso de la infancia en conflicto con la norma” *Revista Derechos y Libertades* nº 33, 2015, p. 90.

³²⁶ *Vid.* arts. 49 y 51 LORPM.

³²⁷ *Vid.* art. 13 RLORPM.

³²⁸ El derecho del menor a ser oído no solo se encuentra regulado en el ámbito nacional con la Ley Orgánica 1/1996 de 15 de enero, de Protección Jurídica del Menor (art. 9) sino además y con carácter previo en la Convención sobre los Derechos del Niño de 1989 (art. 12).

ahí que, junto a la competencia para decidir sobre su revisión, reducción o sustitución, la ley le atribuya facultades como la visita periódica a los centros de internamiento y la realización de entrevistas con los menores³²⁹, así como la solicitud de informes adicionales cuando lo considere necesario³³⁰. Con ello se pretende evitar que la valoración de la evolución del menor durante el cumplimiento de las medidas se limite a una lectura distante del expediente, garantizando que las decisiones que redefinan su situación jurídica se adopten sobre la base de una percepción lo más completa posible de su realidad personal, familiar y social, en coherencia con el derecho del menor a ser escuchado, así como con la finalidad socioeducativa que informa el sistema de justicia juvenil.

5. INTELIGENCIA ARTIFICIAL Y EJECUCIÓN DE MEDIDAS: COMPATIBILIDAD Y SALVAGUARDAS

Sobre el entramado de principios y garantías descrito se proyecta la posible incorporación de sistemas de inteligencia artificial a la fase de ejecución de medidas. Lejos de constituir un ámbito neutro desde el punto de vista procesal, la utilización de herramientas algorítmicas para valorar el riesgo de reincidencia, apoyar decisiones de revisión, sustitución o cese de medidas, o gestionar el seguimiento de su cumplimiento incide directamente en los núcleos esenciales del modelo: la legalidad y jurisdiccionalidad de la respuesta, la consideración primordial del interés superior del menor, la proporcionalidad y flexibilidad de la intervención y las garantías de contradicción, audiencia e inmediación. La cuestión no es, por tanto, si la inteligencia artificial puede mejorar la eficacia del sistema, sino en qué condiciones su empleo resulta compatible con las exigencias propias de un proceso penal de menores de inequívoco perfil garantista.

5.1. La IA como elemento de apoyo en la fase de ejecución

La posible utilización de herramientas de inteligencia artificial en la fase de ejecución de medidas impuestas al menor infractor plantea interrogantes de notable calado jurídico y operativo. El uso de sistemas de IA en el ámbito de la justicia ya es una realidad, aunque, hasta el momento, su despliegue como instrumento de apoyo en la toma de decisiones se ha concentrado fundamentalmente en fases anteriores del proceso penal, muy en particular, en las labores de investigación y en la fase intermedia³³¹. Ello evidencia el potencial de estas soluciones tecnológicas avanzadas para procesar grandes volúmenes de datos y generar recomendaciones útiles para los operadores

³²⁹ *Vid.* art. 44.2 g LORPM.

³³⁰ En efecto, el control jurisdiccional se apoya además en la posibilidad de intensificar el seguimiento mediante la solicitud de nuevos informes a la entidad pública, al amparo de lo previsto en el art. 13.3 del Real Decreto 1774/2004 (RLORPM), que permite tanto al juez como al Ministerio Fiscal, requerir nuevos informes de seguimiento adicionales fuera de la periodicidad ordinaria.

³³¹ Sobre el uso de herramientas a disposición de abogados y fiscales para asistirles en la preparación de asuntos, y valoración sobre la presentación de acusaciones Martín Diz, F. *Aplicación de la inteligencia artificial op. cit.* p. 127.

jurídicos, lo que invita a preguntarse por la eventual traslación de esa capacidad al ámbito específico de la ejecución de medidas en el proceso penal de menores.

Desde la perspectiva de la justicia juvenil, la cuestión no puede plantearse en términos puramente instrumentales. Cualquier reflexión sobre la aplicación de la IA en la fase de ejecución debe partir del reparto competencial existente entre el órgano judicial y la entidad pública de protección o reforma, así como de los principios rectores que informan el sistema: legalidad, jurisdiccionalidad, interés superior del menor, proporcionalidad, flexibilidad y garantías de contradicción y audiencia, entre otros.

El objeto no es tanto determinar si la tecnología puede incrementar la eficiencia administrativa, sino examinar si puede aportar valor en términos de mejor adaptación de las medidas al perfil individual del menor y sobre todo, si puede hacerlo de manera compatible con las exigencias de un proceso penal de menores fuertemente garantista.

Para abordar esta cuestión resulta necesario delimitar, siquiera de forma sintética, qué ha de entenderse por inteligencia artificial a efectos de este trabajo³³². El Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, del 13 de junio de 2024, por el que se establecen las normas armonizadas en materia de inteligencia artificial define los sistemas de IA atendiendo especialmente a su autonomía funcional y a su capacidad para inferir información a partir de datos y generar resultados como predicciones, contenidos, recomendaciones o decisiones.

Esta caracterización resulta suficiente para examinar la eventual utilización de tales sistemas en la fase de ejecución de medidas, particularmente en relación con las decisiones de revisión, modificación, sustitución o extinción de medidas, que en nuestro ordenamiento corresponden al juez de menores mediante resolución motivada.

Como se ha expuesto en los epígrafes precedentes, la ejecución de las medidas se configura por la LORPM como un ámbito marcado por la flexibilidad y la necesidad de adaptación constante a la evolución del menor. La intervención no se agota en la mera comprobación del cumplimiento formal de la medida, sino que requiere un seguimiento continuado de la situación personal, familiar, educativa y social del joven infractor, a partir del cual, el juez puede valorar, con apoyo en los informes de la entidad pública y del equipo técnico, la conveniencia de mantener, suavizar, intensificar o sustituir la respuesta inicialmente acordada. Esta estructura dinámica, que permite tanto la mitigación como el endurecimiento de la medida en función de la evolución del menor, convierte la fase de ejecución en un terreno particularmente propicio para plantear el eventual uso de sistemas de IA orientados a valorar, entre otros aspectos,

³³² No es el objetivo de este trabajo ofrecer una definición precisa, sino más bien una aproximación humilde y respetuosa del concepto para explorar su utilización en el ámbito de la justicia, concretamente en la justicia de menores.

el riesgo de reincidencia o la probabilidad de éxito del proceso educativo durante el tiempo de cumplimiento.

En este contexto, la pregunta que se plantea no es si la inteligencia artificial puede operar en abstracto en la fase de ejecución, lo que tecnológicamente resulta indiscutible, sino si resulta jurídicamente aceptable que el juez de menores y la entidad pública se sirvan de herramientas algorítmicas como apoyo para valorar la evolución de la conducta del joven infractor, siempre que esas herramientas se integren en un marco que preserve intactas las garantías procesales y los principios que informan el sistema de justicia juvenil.

En este sentido, el punto de partida es inequívoco: no se propone en modo alguno la sustitución de las decisiones judiciales en materia de ejecución de medidas. La doctrina ha puesto de relieve, de forma mayoritaria, la imposibilidad de desplazar el ejercicio de la potestad jurisdiccional en favor de sistemas de IA³³³, por cuanto el razonamiento jurídico del juez, en particular la ponderación de los derechos e intereses en juego no es reducible a los patrones de cálculo propios de los algoritmos. Los sistemas de IA pueden procesar de manera ágil grandes cantidades de datos y generar salidas en forma de recomendaciones o predicciones, sin embargo, esas operaciones no generan ni constituyen *pensamiento*³³⁴ ni equivalen a la formación de un juicio jurídico ni suponen la comprensión semántica³³⁵ del caso.

Desde esa premisa, la eventual utilización de herramientas de IA en la fase de ejecución solo puede concebirse, en un horizonte de futuro próximo, como un instrumento estrictamente asistencial de la función jurisdiccional³³⁶. El uso de modelos predictivos o de sistemas de análisis de riesgo podría, en su caso, integrarse como un recurso adicional a disposición del juez de menores, junto con las audiencias, las entrevistas personales, los informes de seguimiento y las valoraciones del equipo técnico y de la entidad pública. Lejos de desplazar el elemento humano de juicio cognitivo, la IA se limitaría a ofrecer una panorámica complementaria³³⁷ sobre determinados aspectos de la evolución del menor, que el juez deberá contrastar críticamente antes de

³³³ Gómez Colomer, J.L. *El juez robot. La independencia judicial en peligro*, Tirant Lo Blanch, 2023, p. 190; Nieva Fenoll, J. *Inteligencia artificial y proceso judicial*, Marcial Pons, 2018, p.16; Pérez Daudí, V. *De la justicia a la ciberjusticia*, Atelier, 2022, p. 43.

³³⁴ Barona Vilar, S. *Algoritmización del Derecho y de la Justicia. De la inteligencia artificial a la Smart Justice*, Tirant Lo Blanch, 2021, p. 556-557.

³³⁵ Ídem p. 559.

³³⁶ Se plantea Planchadell Gargallo dos alternativas en relación con la implantación de sistemas de IA en el ámbito de la justicia, la utilización de herramientas que ayudan a tomar decisiones de un lado, y el uso de herramientas que sustituyen al humano en la toma de decisiones judiciales. Planchadell Gargallo, A. *Inteligencia artificial y proceso penal; especial referencia a las medidas cautelares*, Teoría y Derecho: revista de pensamiento jurídico, n° 37, 2024, pp. 182-183.

³³⁷ En esta misma línea Martín Diz, F. *Aplicación de la inteligencia artificial op. cit.* p. 147.

decidir qué opción resulta más adecuada para favorecer la reeducación y reinserción del menor.

5.2. Riesgos procesales y criterios de compatibilidad

La eventual incorporación de herramientas de inteligencia artificial a la fase de ejecución de medidas no está exenta de riesgos. La clave no reside únicamente en comprobar si el sistema es técnicamente robusto, sino en determinar si su diseño y funcionamiento son compatibles con las exigencias de un proceso penal de menores presidido por el respeto a los derechos fundamentales y a las garantías procesales. Entre los riesgos que deben ser objeto de especial atención destacan, de un lado, los relacionados con la protección de datos³³⁸ personales del menor³³⁹ y la posible introducción o amplificación de sesgos discriminatorios; y, de otro, el peligro de automatismo decisorio y de debilitamiento de la motivación judicial y del control humano efectivo sobre la decisión.

Desde la perspectiva de la protección de datos³⁴⁰, el tratamiento de la información relativa a menores en conflicto con la ley penal a través de sistemas de inteligencia artificial exige un grado especialmente elevado de seguridad y de control normativo. La naturaleza sensible de muchos de los datos manejados en esta fase (información sobre la salud mental, consumo de sustancias, contexto familiar, entorno social o pertenencia a determinados grupos) obliga a garantizar un entorno tecnológico seguro y plenamente respetuoso con el marco normativo vigente en materia de protección de datos³⁴¹ y de administración de justicia³⁴². Cuestiones como la procedencia de los datos,

³³⁸ Pillado González, E. “Posibilidades de aplicación de herramientas de inteligencia artificial en la imposición y ejecución de medidas a menores infractores” en A. Pérez Cruz y S. Calaza López (Dirs.) *Desafíos del Derecho procesal del S. XXI: prueba prohibida, inteligencia artificial y digitalización de la administración de justicia*, Dykinson, 2024, p. 125.

³³⁹ González Pulido, I. “Derecho a la protección de datos personales: herramientas de valoración del riesgo de violencia y proceso penal de menores” en Bujosa Vadell, L. y Martín Diz, F. (Dirs.) *Menores y Justicia juvenil*. Thomson Reuters Aranzadi, 2021, pp. 273-304.

³⁴⁰ Respecto al tratamiento de los datos personales en la justicia de menores Pillado González, E. “Implicaciones de la Directiva (UE) 2016/800 relativa a las garantías procesales de los menores sospechosos o acusados en los procesos penales en la Ley de Responsabilidad Penal del Menor”. *Revista General de Derecho Europeo*, n° 48, 2019.

³⁴¹ La Ley Orgánica 7/2021 de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y ejecución de sanciones penales, así como la Ley Orgánica 3/2018 de 5 de diciembre, de Protección de Datos Personales conforman en marco normativo principal en nuestro país al respecto.

³⁴² Aborda con detalle la protección de los datos personales en el procedimiento judicial Hernández Alfaro, M. A. *El tratamiento de datos personales con fines jurisdiccionales y el ejercicio del derecho de autodeterminación informativa “Habeas data”* Anuario de la Facultad de derecho. Universidad de Extremadura, n° 40, 2024, pp. 485-511.

la forma de su incorporación al sistema, los perfiles de acceso, la conservación y eventual reutilización de la información adquieren una relevancia decisiva cuando se integran en modelos algorítmicos susceptibles de influir en decisiones sobre la intensidad o duración de las medidas.

En este contexto, la opción de modelos de gestión basados en proveedores públicos, dependientes de la Administración de Justicia, aparece como una vía especialmente coherente con las exigencias de control y responsabilidad institucional. La experiencia de sistemas como VioGén³⁴³ en el ámbito de la violencia de género pone de relieve la capacidad de la Administración para desarrollar y gestionar herramientas tecnológicas complejas en materia de seguridad y protección de víctimas, bajo estrictos estándares de seguridad, interoperatividad y protección de datos.

Asimismo, el marco normativo de la jurisdicción permite el tratamiento de datos personales obtenidos en el ejercicio de la función jurisdiccional sin necesidad de recabar el consentimiento del afectado³⁴⁴, e incluso prevé la posibilidad de utilizar datos no personales para el desarrollo de sistemas automáticos de apoyo a la tramitación procesal, siempre bajo el estricto cumplimiento de las normas de seguridad e interoperatividad³⁴⁵.

Estas previsiones apuntan a que el legislador contempla la utilización de tecnologías avanzadas, incluida la IA, como posibles instrumentos de apoyo a la función judicial, sin que ello suponga una relajación de las garantías³⁴⁶.

Junto a la vertiente estrictamente relativa a la protección de datos, la utilización de sistemas de inteligencia artificial en este ámbito debe afrontar el riesgo de los sesgos discriminatorios³⁴⁷. La incorporación de los modelos de variables vinculadas a la pertenencia a determinados colectivos, bandas o grupos, al origen nacional o étnico, o al consumo de alcohol y otras sustancias puede dar lugar a patrones de tratamiento que, aun siendo formalmente neutros, en la práctica reproduzcan o acentúen situaciones de discriminación estructural³⁴⁸.

³⁴³ López García-Nieto, I. “Protección de las víctimas de violencia de género”, *Diario La Ley* n° 9721, *Sección Tribuna*, 2020, pp. 2-6.

³⁴⁴ *Vid.* art. 236 ter LOPJ.

³⁴⁵ *Vid.* art. 236 sexies LOPJ

³⁴⁶ Así lo defiende Martín Diz, F. *Aplicación de la inteligencia artificial op. cit.* p. 141-145.

³⁴⁷ Se ocupa de analizar los sesgos Castillejo Manzanares, R. “Una pregunta difícil de responder: ¿es la IA imparcial?” en R. Castillejo Manzanares y L. Noya Ferreiro *Inteligencia artificial y proceso penal: un reto para la justicia*. Aranzadi, 2023 p. 82-83.

³⁴⁸ En este sentido Pérez Estrada, M. J. *Fundamentos jurídicos para el uso de la inteligencia artificial en los órganos judiciales*, Tirant Lo Blanch, 2022, pp. 32-33.

La discusión doctrinal en torno a la legitimidad del uso de herramientas predictivas se ha centrado precisamente en la necesidad de evitar que categorías históricamente estigmatizadas se conviertan en factores de riesgo automatizados que condicionen, de manera opaca³⁴⁹, la evaluación de la probabilidad de reincidencia o de éxito de los itinerarios educativos. Cualquier diseño de sistemas de IA aplicados a la ejecución de medidas deberá, por tanto, incorporar salvaguardas específicas de auditoría, corrección y transparencia orientadas a identificar y neutralizar dichos sesgos.

Un tercer foco de preocupación lo constituye el riesgo de automatismo decisorio. La experiencia comparada ha puesto de relieve el peligro de que, en la práctica, el juez termine por delegar de forma acrítica en la herramienta tecnológica su razonamiento, de modo que las recomendaciones algorítmicas adquieran un peso determinante en la decisión, aun cuando formalmente sigan presentándose como meros elementos de apoyo. Ello podría derivar en decisiones estandarizadas, poco sensibles a las particularidades del caso concreto, y en un debilitamiento de la exigencia de motivación reforzada que caracteriza las resoluciones que afectan a derechos fundamentales del menor. El peligro no radica únicamente en la posible *pereza cognitiva* del operador jurídico, sino en la opacidad técnica de algunos modelos, que dificulta explicar de forma comprensible por qué se ha asignado a un menor un determinado nivel de riesgo o se ha recomendado una determinada intensidad en la intervención.

Ahora bien, el alcance de este riesgo de automatismo se ve atemperado, en el ámbito de la ejecución de medidas, por la configuración legal de la fase de ejecución en el proceso penal de menores. El reparto competencial entre el juez de menores y la entidad pública de protección o reforma garantiza que la decisión jurisdiccional se adopte sobre la base de un haz plural de fuentes de información: informes de seguimiento, valoraciones del equipo técnico, percepciones de los profesionales que intervienen en el cumplimiento de la medida, entrevistas con el menor, y en su caso, con su entorno familiar. La inteligencia artificial, en caso de incorporarse, se integraría en este entramado como una fuente adicional de análisis, pero no sustitutiva de las restantes. Desde esta óptica, la compatibilidad de la IA con las garantías procesales pasa necesariamente por preservar un núcleo irreductible de control humano sobre la decisión, asegurando que el juez conozca y comprenda el funcionamiento básico de las herramientas utilizadas, someta sus resultados a un contraste crítico y motive de manera expresa la medida en términos comprensibles para el menor, con independencia de que haya contado o no con apoyo algorítmico.

En suma, la utilización de sistemas de inteligencia artificial en la ejecución de medidas solo puede considerarse compatible con el modelo garantista del proceso penal de menores si se articula sobre tres ejes: un marco de protección de datos

³⁴⁹ Realiza un completo análisis del proceso de creación del modelo de predicción y cómo pueden aparecer los sesgos en las distintas fases de dicho proceso Bravo Volado, A. *Justicia algorítmica: un enfoque sociotécnico*, Revista de Estudios Penales y Criminológicos, n° 44, 2023, pp. 5-8.

especialmente robusto, que tenga en cuenta la condición de persona menor de edad y el carácter sensible de la información; mecanismos efectivos de prevención, detección y corrección de sesgos discriminatorios; y la preservación de un control humano fuerte, que evite cualquier forma de automatización encubierta de la decisión y garantice la vigencia plena de las exigencias de motivación, contradicción, audiencia e intermediación que caracterizan la justicia juvenil.

6. CONCLUSIONES

El análisis desarrollado en este trabajo ha permitido mostrar que la ejecución de medidas en el proceso penal de menores constituye un momento decisivo, en el que se concreta la respuesta del ordenamiento jurídico frente a la delincuencia juvenil. En esta fase el juez de menores adopta decisiones que inciden directamente en los derechos fundamentales del menor, y por tanto quedan sometidas a un sólido sistema de garantías.

La arquitectura de la ejecución diseñada por la LORPM permite una intervención flexible e individualizada, basada en un catálogo tasado de medidas y en mecanismos de revisión, sustitución o cese, pero reserva un núcleo de control jurisdiccional sobre las incidencias relevantes en la situación jurídica del menor.

Se ha puesto de manifiesto cómo la ejecutoria se encuentra atravesada por un haz de principios procesales que informan su desarrollo (legalidad, jurisdiccionalidad, interés superior del menor, proporcionalidad, flexibilidad, contradicción, audiencia e intermediación), permitiendo articular una intervención individualizada, gobernada por el interés superior del menor y orientada a su reeducación y reinserción social.

La IA ofrece un potencial asistencial formidable para procesar gran cantidad de datos, y aportar valoraciones útiles. No obstante, su uso plantea riesgos significativos. De ahí que cualquier utilización de esta tecnología deba apoyarse en una base normativa clara que afirme su carácter estrictamente asistencial, y garantice el control humano efectivo respecto de la decisión judicial.

En definitiva, la compatibilidad entre el algoritmo y la legalidad en la ejecución de medidas, pese a ser tecnológicamente factible, entraña peligros que no pueden ignorarse. Solo si la IA se concibe como un instrumento asistencial más, subordinado a los principios y fines que persigue el proceso penal de menores, podría integrarse de forma coherente en nuestro modelo de justicia juvenil inequívocamente garantista.

BIBLIOGRAFÍA

- Alba Robles, J. L., Aroca Montolío, C. y López Latorre, M. J. “El modelo criminológico de evaluación e intervención para menores en conflicto con la ley”, *Archivos de Criminología, Seguridad Privada y Criminalística*, nº 16, 2016.
- Barona Vilar, S., *Algoritmización del Derecho y de la Justicia. De la inteligencia artificial a la Smart Justice*, Tirant Lo Blanch, 2021.
- Bernuz Beneitez, M., “El derecho a ser escuchado: el caso de la infancia en conflicto con la norma”. *Revista Derechos y Libertades* nº 33, 2015.
- Blanco Barea, J. A., “Responsabilidad penal del menor: principios y medidas judiciales aplicables en el Derecho Penal español”, *Revista de Estudios Jurídicos*, nº 8/2008 (Segunda Época).
- Bravo Volado, A., “Justicia algorítmica: un enfoque sociotécnico”, *Revista de Estudios Penales y Criminológicos*, nº 44, 2023.
- Bujosa Vadell, L. M., “Ejecución en el enjuiciamiento de menores: Principios aplicables y limitación de la discrecionalidad” en L.M. Bujosa Vadell y F. Martín Diz (Dir.) *Menores infractores: Predicción, gestión del riesgo e intervención*, Thomson Reuters Aranzadi, 2022.
- Bujosa Vadell, L. M., “Proceso Penal de menores y derecho comparado: Novedades y desafíos” en P. Martín Ríos y M.A. Pérez Martín (Dir.) *La administración de justicia en España y América*, Astigi, 2021.
- Colás Turégano, M.A., “Punitivismo y justicia de menores: La reforma de la Ley Reguladora de la Responsabilidad Penal de los Menores (LO 5/2000) por la Ley del “solo sí es sí” (LO 10/2022 de Garantía Integral de la Libertad Sexual”, *Revista Electrónica de Ciencia Penal y Criminología*, 25-26, 2023.
- Colás Turégano, M.A., “Hacia una humanización de la justicia penal: la mediación en la justicia juvenil española. Principios y ámbito aplicativo en la LO 5/2000, reguladora de la responsabilidad penal de los menores”, *Revista Boliviana de Derecho* nº 20, 2015.
- Cano Paños, M.A., “¿Es conveniente un endurecimiento del Derecho penal juvenil? Una toma de posición crítica”, *Anuario de Derecho penal y Ciencias Penales*, vol. LV 2002.
- Castillejo Manzanares, R., “Una pregunta difícil de responder: ¿es la IA imparcial?” en R. Castillejo Manzanares y L. Noya Ferreiro *Inteligencia artificial y proceso penal: un reto para la justicia*, Aranzadi, 2023.
- García Pérez, O., *Medidas y su ejecución en el sistema de justicia penal juvenil*, Tirant lo Blanch, 2019.
- Garrido Carrillo, F. J., *Principios y garantías del proceso penal de menores*, Aranzadi, 2023.
- González Pulido, I., “Derecho a la protección de datos personales: herramientas de valoración del riesgo de violencia y proceso penal de menores” en L.M. Bujosa Vadell y F. Martín Diz (Dir.) *Menores y Justicia juvenil*. Thomson Reuters Aranzadi, 2021.
- Gómez Colomer, J. L., *El juez robot. La independencia judicial en peligro*, Tirant Lo Blanch, 2023.

- Hernández Alfaro, M. A., “El tratamiento de datos personales con fines jurisdiccionales y el ejercicio del derecho de autodeterminación informativa “Habeas data”, *Anuario de la Facultad de Derecho. Universidad de Extremadura*, nº 40, 2024.
- Jullien de Asís, J., “Reparación de las víctimas en la justicia juvenil” en T. Farto Piay (Coord.) *La víctima en el proceso penal de menores. Tratamiento procesal e intervención socioeducativa*, Dykinson, 2021.
- López García-Nieto, I., “Protección de las víctimas de violencia de género.” *Diario La Ley nº 9721, Sección Tribuna*, 2020.
- Martín Diz, F., “Aplicación de la inteligencia artificial en la evaluación criminológica en materia de ejecución de medidas impuestas a menores” en L.M. Bujosa Vaddell y F. Martín Diz (Dir.) *Menores infractores, predicción, gestión del riesgo e intervención*, Thomson Reuters Aranzadi, 2022.
- Montero Hernanz, T., *Derecho penal de menores. Una introducción a la legislación reguladora de la responsabilidad penal de los menores*, Reus, 2023.
- Nieva Fenoll, J., *Inteligencia artificial y proceso judicial*, Marcial Pons, 2018.
- Palma Chaparra, L., “La mediación” en A. Monge Fernández (Dir.) *La protección jurídica del menor*, Tirant Lo Blanch, 2024.
- Pérez Estrada, M. J., *Fundamentos jurídicos para el uso de la inteligencia artificial en los órganos judiciales*, Tirant Lo Blanch, 2022.
- Pérez Daudí, V., *De la justicia a la ciberjusticia*, Atelier, 2022.
- Pillado González, E., “Posibilidades de aplicación de herramientas de inteligencia artificial en la imposición y ejecución de medidas a menores infractores” en A. PÉREZ CRUZ y S. CALAZA LÓPEZ (Dir.) *Desafíos del Derecho procesal del S. XXI: prueba prohibida, inteligencia artificial y digitalización de la administración de justicia*, Dykinson, 2024.
- Pillado González, E., “Implicaciones de la Directiva (UE) 2016/800 relativa a las garantías procesales de los menores sospechosos o acusados en los procesos penales en la Ley de Responsabilidad Penal del Menor”. *Revista General de Derecho Europeo*, nº 48, 2019.
- Planchadell Gargallo, A., *Inteligencia artificial y proceso penal; especial referencia a las medidas cautelares*, Teoría y Derecho: revista de pensamiento jurídico, nº 37, 2024.
- Vázquez González, C., *Delincuencia Juvenil*, Dykinson, 2019.



CAPÍTULO 8. JUSTICIA EN PLAZO, SIN DILACIONES, E INTELIGENCIA ARTIFICIAL: ALGUNAS PROPUESTAS³⁵⁰

AURELIO DE PRADA GARCÍA³⁵¹

Universidad Rey Juan Carlos
aurelio.deprada@urjc.es

“For who would bear the whips and scorns of
time, th'oppressor's wrong, the proud man's contumely,
the pangs of dispriz'd love, the law's delay, the insolence
of office, and the spurns that patient merit of
th'unworthy takes, when he himself might his quietus
make with a bare bodkin?”

Shakespeare, *Hamlet*, III;1.

Resumen

En trabajos anteriores y a partir de un caso en el que se habrían violado presuntamente aspectos esenciales del primer derecho humano procesalmente hablando: el derecho a la tutela judicial efectiva, ya analizamos las virtualidades de la implementación de la inteligencia artificial (IA) en la administración de justicia. Ahora bien, no enfatizamos lo suficiente las ventajas de tal implementación en cuanto a la corrección de la pérdida de legitimidad y del descrédito de los órganos jurisdiccionales concernidos y tampoco tuvimos en cuenta lo que dicha implementación supondría de cara a solventar un defecto en la administración de justicia ya señalado por la sabiduría popular desde hace siglos y también por alguno de los autores más importantes de la literatura universal

³⁵⁰ El presente trabajo se realiza en el desarrollo de las actividades propias del Grupo de Investigación, Derechos Humanos e Inteligencia Artificial (IAIURIS) de la Universidad Rey Juan Carlos. Encuentra su origen en las Jornadas Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar organizadas en el marco de las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

³⁵¹ Licenciado en Derecho (UCM), Doctor en Derecho (UCM) y en Filosofía (UVA). Ha sido profesor de Teoría y Filosofía del Derecho en la UCM, UNED, UB y URJC de la que actualmente es Profesor colaborador honorífico. Estancias de investigación en la Universidade de Coimbra, National Taipei University, La Sapienza, Georg August Universität Göttingen, Tor Vergata y Javeriana de Cali. Entre sus últimas publicaciones: *El silogismo jurídico y la inteligencia artificial* (2025) *¿El Leviatán o Las Meninas? Hacia una nueva mirada política* (2025), *¿Hobbes o Confucio? El cambio climático y las formas de identidad* (2024); *Greta y el Leviatán: un análisis del movimiento adolescente contra el cambio climático* (2023) y *El proceso de Sócrates: Del “nosotros” al “yo”. Contra el prejuicio individualista no percibido* (2022). ORCID: 0000-0002-5160-3325

como Shakespeare en la cita que encabeza estas líneas: la tardanza en la administración de justicia (“the law’s delay”). Tardanza que el autor inglés incluso llega a equiparar con males que justificarían el suicidio (“he himself might his quietus make with a bare bodkin”).

En este trabajo procederemos a completar dicho análisis hasta llegar a la conclusión de que, si bien sería de gran ayuda, la implementación de la IA en la administración de justicia, tardanza incluida, no sería sin embargo suficiente para corregir todos los defectos que se siguen de ese análisis. Habría que combinar el uso de la IA con un replanteamiento general de la consideración de los juristas y del propio derecho, incluyendo su consideración peyorativa y, por tanto, un replanteamiento de la enseñanza jurídica, así como del acceso a la carrera judicial para el que sería requisito *sine qua non* el haber sido parte en un juicio.

Palabras clave

Administración de justicia, concepto de derecho, enseñanza jurídica, IA, justicia en plazo, tutela judicial efectiva.

Abstract

In earlier works, drawing on a case in which core elements of what may be regarded as the first procedural human right—the right to effective judicial protection—were allegedly violated, we examined the potential role of artificial intelligence (AI) in the administration of justice. What we did not sufficiently underscore, however, were the benefits that such implementation could offer in helping to restore legitimacy and credibility to the judicial bodies concerned. Nor did we consider how it might address a longstanding defect in judicial administration, one captured for centuries in popular wisdom and famously articulated by Shakespeare in the epigraph to this article: the chronic delay of the law (‘the Law’s delay’). Shakespeare goes so far as to place such delay among the afflictions that might drive a person to contemplate suicide (‘he himself might his quietus make with a bare bodkin’).

In this article, we revisit and expand that earlier analysis and argue that, although the incorporation of AI into judicial processes—including the mitigation of delay—would be of substantial value, it would nonetheless fall short of remedying all the shortcomings identified. Addressing these deficiencies would require not only the use of AI but also a broader re-evaluation of the role of legal professionals and of the nature of law itself—including its often pejorative social perception. This, in turn, would call for a reconsideration of legal education and of the criteria for admission to the judiciary, for which having been a party to legal proceedings would have to be regarded as a *sine qua non* condition.”

Keywords

Administration of justice, AI, concept of law, justice in time, legal teaching, right to effective judicial protection.

Sumario: INTRODUCCIÓN 1. DE NOBIS IPSIS NON SILEMUS 2. THE FIRST THING WE DO, LET'S KILL ALL THE LAWYERS 3. ALGUNAS PROPUESTAS. 4. ¿PLAZOS RAZONABLES Y DILACIONES INDEBIDAS? 5. A MODO DE CONCLUSIÓN BIBLIOGRAFÍA UTILIZADA

INTRODUCCIÓN

No parece necesario recordar que la inteligencia artificial (IA) supone una de las mayores revoluciones tecnológicas de todos los tiempos. Y ello por mucho que aún se halle en sus comienzos, -en “los albores de la era de la IA” por decirlo con la Sala de lo Contencioso Administrativo, sección tercera, del Tribunal Supremo³⁵²- y también por mucho que hoy por hoy presente varios modelos, en liza por lo demás: la IA de la llamada “barra libre” de los EEUU de D. Trump; la IA “estatalista” de la China de Xi Jinping o la “garantista” de la UE. Una batalla que ciertamente no impide sino todo lo contrario que se vaya implementando en diferentes campos, incluso en el de la gestión pública, si bien con mayores o menores resistencias.

Y es que tampoco parece necesario recordar que la revolución tecnológica de la IA, al igual que todas las demás, presenta tanto ventajas como inconvenientes dando lugar a múltiples reflexiones y propuestas al respecto como, sin ir más lejos y entre nosotros, la del jurista español S. Carretero defensor del llamado Humanismo tecnológico³⁵³ en el que se establecen los principios prácticos que deben guiar a las sociedades contemporáneas hacia un uso ético, responsable y humanista de la IA.

Pues bien, uno de los campos en los que, pese a encontrarse aún en un estadio inicial y pese a algunas reticencias, la IA estaría ya implementándose es precisamente el de la gestión pública, como antes señalábamos y, más en concreto, el de la Administración de Justicia. Y en efecto, está ya presente en los sistemas judiciales de la inmensa mayoría de los países del mundo, el nuestro incluido³⁵⁴.

³⁵² Sentencia 1.119/2025, caso Bosco, de la Sección tercera, Sala de lo Contencioso del Tribunal Supremo de 11 de septiembre de 2025.

³⁵³ Carretero, S., *Humanismo tecnológico: justicia predictiva y derechos humanos en la era digital*, Colex, Madrid 2025.

³⁵⁴ “Actualmente, los sistemas judiciales del mundo entero utilizan inteligencia artificial (IA) para analizar grandes cantidades de datos legales a fin de ayudar a los abogados a identificar precedentes en la jurisprudencia, permitir que las administraciones hagan más eficientes los procesos judiciales y apoyar a los jueces con predicciones sobre cuestiones que abarcan la duración de sentencias y los registros de reincidencia”. <https://en.unesco.org/artificial-intelligence/mooc-judges>. Consultado el 3.11.2025.

Una implementación que, por lo demás, nosotros mismos hemos defendido en trabajos anteriores³⁵⁵ a propósito de un caso concreto: una *reformatio in peius* que, tras pasar por la Audiencia Nacional (AN) y el Tribunal Constitucional (TC) llegó hasta el Tribunal Europeo de Derechos Humanos (TEDH) y en el que se habrían vulnerado presuntamente dimensiones y principios básicos de la tutela judicial efectiva, del primer derecho humano procesalmente hablando. Y así, en tales trabajos habríamos demostrado que, con la aplicación de la IA, se habrían evitado tales presuntas vulneraciones del derecho a la tutela judicial efectiva, si bien no enfatizamos lo suficiente que tal aplicación serviría también para evitar el descrédito e incluso la deslegitimación que tales presuntas vulneraciones suponían en los órganos judiciales concernidos. Todo ello, por supuesto, dicho con el máximo respeto y en estrictos términos de defensa si es que cabe recurrir a esa fórmula procesal en sede académica

Así las cosas, en lo que sigue, partiendo de esos trabajos previos que ahora completaremos en el sentido recién señalado, trataremos de mostrar además la conveniencia de la implementación de la IA en un punto crucial de dicho derecho a la tutela judicial efectiva que entonces tampoco tratamos: el derecho a la justicia en plazo, sin dilaciones. Algo que por lo demás y como es bien sabido, viene reclamando no solo la sabiduría popular desde hace siglos en dichos como “la justicia tardía no es justicia” o el de “justicia retardada, justicia denegada” sino también algunos de los autores más importantes de la literatura universal como Shakespeare en la cita de *Hamlet* que encabeza estas líneas y en la que se equipara la tardanza de la justicia (“the law’s delay”) con otros males que cabría evitar recurriendo al puñal (a bare bodkin), esto es, al suicidio.

Un punto crucial decimos de la tutela judicial efectiva y cabría objetar que, al menos en nuestro ordenamiento jurídico, tal afirmación resulta inexacta pues, como es asimismo bien sabido, en el art 24.1 de la Constitución Española (CE) se recoge el derecho a la tutela judicial efectiva mientras que en el 24.2 CE se habla de un derecho a un proceso público sin dilaciones indebidas. Esto es, se trataría de dos derechos distintos como, por otro lado, el TC ha puesto de relieve en sentencias ya antiguas como la del recurso de amparo 26/83 de 13 de abril de 1983 en la que se señala que jurídicamente se trataría de dos derechos distintos.

Ahora bien, en esa misma sentencia se señala también que «desde el punto de vista sociológico y práctico, puede seguramente afirmarse que una justicia tardíamente concedida equivale a una falta de tutela judicial efectiva»³⁵⁶, incluyendo así el derecho a un

³⁵⁵ Vid., por ejemplo, Prada García, A. de, *¿Más allá de Rashomon? En busca de los cadáveres perdidos*. En Cobo Del Rosal G. Navalpotro Sánchez, J.M y Cano Linares A. (Dirs.) ARS IURIS. Estudios sobre Derecho y Arte. Editorial Dykinson, Madrid 2023, pp.168-184.

³⁵⁶ Marín Castán, M. L., *La polémica cuestión de la determinación del «plazo razonable» en la administración de justicia* (Comentario a la Sentencia del Tribunal Europeo de Derechos Humanos de 13 de julio de 1983) *Revista española de derecho constitucional*, I Año n° 4, N° 10, 1984, p 225.

proceso sin dilaciones indebidas dentro del derecho a la tutela judicial efectiva y ello desde la sabiduría popular a la que nosotros apelábamos hace un momento.

Pero a todo esto, no estamos siendo todo lo claros que debiéramos pues de lo anterior cabría pensar que defendemos la implementación de la IA para hacer efectivo el derecho a una justicia sin dilaciones indebidas que establece el 24.2 CE, derecho que sería parte del derecho a la tutela judicial efectiva recogido en el 24.1, cuando nuestra propuesta es más radical. Es *de lege ferenda* y no *de lege lata*, toda vez que defendemos que la implementación de la IA en la administración de justicia llevaría a una justicia sin dilaciones, obligando así a modificar el art. 24.2 CE y, por lo mismo, también el art. 6.1 del Convenio Europeo de Derechos Humanos (CEDH), en el que se establece el derecho a un proceso “dentro de un plazo razonable”. Llevaría a una justicia en plazo, sin dilaciones, tal y como reza el título dado a estas líneas en el que se pretende sintetizar las modificaciones legales que nuestra propuesta comportaría.

1.- *DE NOBIS IPISIS NON SILEMUS*

Antes de nada y tal y como hicimos ya en alguno de esos trabajos anteriores a los que más arriba aludíamos, resulta obligado hacer constar que en la re-exposición del caso que nos servirá de referencia violaremos el viejo brocardo “*de nobis ipsis silemus*” de vigencia general en sede académica. Una vulneración forzosa, por lo demás, como se verá a continuación, sobre todo en lo que respecta a la AN en relación con la omisión en la sentencia de la prueba admitida y practicada y al TEDH cuyas directrices obligan a la destrucción, al cabo de un año, del expediente de las demandas inadmitidas de modo que efectivamente el expediente del caso sobre el que hablaremos habría sido ya destruido. Circunstancias que ciertamente nos obligan a violar dicho principio.

Yendo a ello, el caso se generó a raíz de una “*reformatio in peius*”, en concreto, en el hecho de que dos comités asesores de la Comisión Nacional de Evaluación de la Actividad Investigadora (CNEAI) aprobaron primero dos aportaciones científicas presentadas a evaluación (convocatoria de 2006, comité asesor n. 9 Derecho), suspendiéndolas en la convocatoria del año siguiente (2007, comité asesor n. 9 Derecho), sin que hubieran variado los criterios de evaluación y por el mero hecho de la modificación en la composición de los comités asesores. Cambio que, literalmente, podía llevar a modificar la calificación previa.

Ahora bien, los dos comités eran materialmente idénticos en cuanto a la mayoría de sus miembros, -presidente y especialista concernido incluidos-, toda vez que el cambio en la composición se limitó a la entrada de tres nuevos miembros y a la salida de dos del comité anterior, sin que variara la mayoría en ambos comités. Así las cosas, la motivación esgrimida para el suspenso de las aportaciones previamente aprobadas: mero cambio en la composición del comité, suponía afirmar implícitamente que la sola presencia de tres nuevos miembros en el segundo comité asesor, sin necesidad de hacer afirmación o argumentación alguna al respecto, habría sido suficiente para que la

mayoría presente en los dos comités y que había aprobado esas aportaciones aceptara modificar la calificación previa y suspender las aportaciones aprobadas en la convocatoria anterior, sin necesidad de hacer afirmación o argumentación alguna al respecto .

No parece necesario insistir en lo irracional de tal motivación, ni tampoco en el desprestigio e incluso deslegitimación completa, -dicho sea, una vez más, con el máximo respeto y en estrictos términos de defensa-, que de ello se deriva para el organismo presuntamente científico, en cuanto que se encarga de la evaluación de la actividad investigadora, que la sostuvo: la CNEAI. Y en efecto no parece compatible con cualquier mínimo criterio científico, por muy laxamente que se entienda el término ciencia, modificar a la baja la calificación de dos aportaciones científicas con la motivación recién señalada: la mera presencia de tres nuevos miembros en el comité asesor que, sin utilizar palabra o argumentación alguna, suspenden (“reformatio in peius”) dos aportaciones previamente aprobadas por la mayoría presente en ambos comités y en la que figuraba el especialista concernido.

Y no solo para el organismo presuntamente científico involucrado pues, como es bien sabido, los “sexenios” son un punto fundamental en la carrera académica universitaria que, a su vez, constituye la base de nuestro sistema educativo. Carrera académica y sistema educativo que obviamente se ven contaminados, por así decirlo, en su totalidad por la irracionalidad de esa motivación como creemos haber hemos demostrado en otro lugar³⁵⁷, sin que por razones de espacio proceda entrar ahora en ello.

Por lo demás y en lo que aquí importa en último término, no parece preciso señalar que con la implementación de cualquier sistema de IA se hubiera detectado la irracionalidad de tal decisión evitándose así el desprestigio y la deslegitimación tanto de la CNEAI cuanto de nuestro sistema universitario y de nuestro sistema educativo recién señalados³⁵⁸.

Un desprestigio y una deslegitimación que, *mutatis mutandis* y de nuevo con el máximo respeto y en estrictos términos de defensa, se aplicarían al siguiente órgano jurisdiccional al que, tras diversos avatares en los que, de nuevo por razones de espacio, no procede detenerse³⁵⁹, llegó el caso; la sección Sexta de la Sala de lo Contencioso de la AN (Sentencia de 14 de febrero de 2014 P.O. 439/2012) a la que se recurrió al haberse tenido conocimiento informal de que la ANECA-CNEAI había instaurado de hecho

³⁵⁷ Prada García, A. de, *Sombras en la evaluación de la actividad investigadora*. Claves de razón práctica, n. 256, enero-febrero 2018, p. 90-99

³⁵⁸ Vid. Domínguez Angulo, J. P. y Ortiz Bolaños, L., *Perspectiva Ontológico-hermenéutica de la Inteligencia Artificial. Una colección de métodos para su mejoramiento y delimitación en la interpretación del derecho* en R. Opin. Jur., Fortaleza, año 23, n. 44, p. 35 y ss.

³⁵⁹ Vid al respecto, Prada García, A. de, *Por una reforma del derecho a obtener la tutela efectiva de Jueces y Tribunales: A propósito de dos sentencias de la Audiencia Nacional y una providencia de inadmisión del TC*. Boletín de Ministerio de Justicia. Año LXX, n. 2193, noviembre de 2016.

un sistema por el que ya no era posible “*reformatio in peius*” alguna. Un sistema informático detectaba si las aportaciones presentadas lo habían sido previamente y supuesto que sí, se mantenía la calificación o se mejoraba no pudiendo empeorarse en ningún caso.

A la vista de tal conocimiento informal, se pidió prueba al respecto; prueba que fue admitida en parte y cuya práctica resultó en unos términos muy favorables al recurrente. Términos en los que, una vez más por razones de espacio, no podemos entrar limitándonos a indicar el lugar donde los hemos expuesto en detalle³⁶⁰. Por el contrario, sí que es obligado recordar que, como hemos señalado más arriba al justificar la violación del brocardo “*De nobis ipsis silemus*”, esa prueba admitida y practicada no aparecía en lugar alguno de la sentencia desestimatoria del recurso y con condena en costas (1500 euros) para el recurrente. La única alusión a la misma eran las palabras “habiendo practicado las declaradas pertinentes” sin que tal práctica y su resultado entrase en la motivación de la sentencia, vulnerándose así, presuntamente, los preceptos legales al respecto: art.209.2 LEC y 218.2 (LEC).

No parece necesario insistir en que esa presunta incongruencia omisiva, por decirlo técnicamente, desprestigia y deslegitima a un órgano jurisdiccional como la AN en cuanto que, como acaba de señalarse, incumplía en los términos vistos todos los preceptos legales al respecto. Más en concreto, incumplía uno de los puntos clave, esenciales de la tutela judicial efectiva, del primer derecho humano procesalmente hablando en que cuanto que la garantía de los demás derechos dependen de él. Y en efecto resulta evidente que no tiene sentido alguno hablar de tutela judicial efectiva cuando la prueba admitida y practicada no figura en la sentencia.

Por lo demás, tampoco parece necesario insistir de nuevo en lo que aquí importa en ultimo termino que cualquier sistema de IA habría detectado inmediatamente esa incongruencia omisiva obligando a incluir la prueba practicada en la motivación de la sentencia y evitando el desprestigio y la deslegitimación recién aludidos.

Un desprestigio y una deslegitimación, dicho una vez más con el máximo respeto y en estrictos términos de defensa, que *mutatis mutandis* se seguirían también de la actuación del siguiente órgano que conoció el asunto: el TC ante el que, -tras el pertinente incidente de nulidad de actuaciones ante la AN, desestimado como era previsible-, se presentó recurso de amparo (5497/2014) por violación del derecho a la tutela judicial efectiva (art. 24.1 CE) al no incluirse en la Sentencia de la AN la prueba admitida y practicada. Y es que, transcurridos quince meses, la Sección primera de la Sala primera del TC dictó providencia (16 de diciembre de 2015) por la que inadmitía el recurso,

³⁶⁰ Ibidem, p. 9 y ss.. Vid. También, Prada García, A. de, *Tomando la prueba, (los viejos y los nuevos derechos) en serio. A propósito de una Sentencia de la Audiencia Nacional P.O. 439/2012*, Revista General de Derecho Procesal 39, 2016.

dada la “manifiesta inexistencia de violación de derecho fundamental tutelable en amparo.”

Ciertamente no parece necesario abundar en el desprestigio y falta de legitimidad que semejante decisión comportan. Y en efecto, el TC no solo reconoce que ha necesitado un año y tres meses para apreciar algo que literalmente, según sus propias palabras, sería manifiesto, sino que asimismo afirma implícitamente que la no inclusión en una sentencia de la prueba admitida y practicada no supone violación alguna del primer derecho humano procesalmente hablando: el derecho a la tutela judicial efectiva.

Una vez más, tampoco parece necesario insistir en que la aplicación de cualquier sistema de IA hubiera detectado inmediatamente tanto la incongruencia de utilizar el término “manifiesto” en una providencia que ha tardado 15 meses en materializarse como la de sostener que la prueba practicada no debe figurar ni, por tanto, motivar una sentencia.

Pero donde el desprestigio y la pérdida de legitimidad llegarían al culmen, dicho sea, una vez más, con el máximo respeto y en estrictos términos de defensa, sería en el TEDH, último órgano que conoció el caso y ello precisamente por cuanto que, como es bien sabido, constituye la cima de uno de los sistemas de protección de los Derechos Humanos presuntamente más avanzados del mundo.

Pues bien, una formación de juez único del TEDH inadmitió la demanda contra el Reino de España (D. 32.905/16), planteada en base al art 13 del CEDH: derecho a un recurso efectivo que, obviamente, exige la inclusión en la sentencia de la prueba admitida y practicada. Decisión de inadmisión en la que se aducía que no se cumplían las condiciones de admisibilidad y ello no solo sin precisar en qué puntos concretos se incumplían sino también advirtiendo que no se darían explicaciones adicionales al respecto y que el expediente sería destruido al cabo de un año como señalábamos más arriba a propósito de la necesidad de incumplir el principio “de nobis ipsis silemus”.

Una vez más no parece necesario insistir en que cualquier sistema de IA hubiera recordado al juez único el art 45.1 CEDH en el que se establece la obligación de motivar tanto las sentencias como las decisiones de admisión o inadmisión de las demandas, lo que le habría obligado a especificar mínimamente los puntos en los que tales condiciones de admisibilidad no se cumplían.... Pero no procede seguir abundando en ello porque bien pudiera calificarse de superficial, si es que no de ingenuo, el análisis realizado hasta aquí y en el que hemos concluido repetidamente en que la aplicación de la IA habría evitado tanto las actuaciones reseñadas cuanto el desprestigio y la falta de legitimidad consiguientes.

Y es que, dicho una vez más con el máximo respeto y en estrictos términos de defensa, resulta difícil imaginar que la ANECA/CNEAI no fuera plenamente consciente de la irracionalidad de la motivación esgrimida para justificar la “reformatio in

peius” origen del pleito. Igualmente resulta difícil pensar que la AN no fuera plenamente consciente de la incongruencia omisiva realizada al omitir la práctica de la prueba admitida en la sentencia y ello tanto más cuanto que se planteó un incidente de nulidad de actuaciones al respecto. Asimismo, resulta difícil creer que el TC no fuera plenamente consciente de lo inapropiado de utilizar el término “manifiesta” en una decisión que había tardado quince meses en alcanzarse, así como de lo indefendible que resulta sostener que la omisión en una sentencia de la prueba admitida y practicada no constituye violación de derecho fundamental tutelable en amparo. Finalmente, también resulta difícil imaginar que el juez único del TEDH no fuera plenamente consciente de que el art. 45.1 del CEDH le obligaba a especificar mínimamente en qué puntos concretos la demanda inadmitida no cumplía las condiciones de admisibilidad.

Así las cosas, y dicho una vez más con el máximo respeto y en estrictos términos de defensa, parece obligado concluir que la implementación de cualquier sistema de IA no habría modificado la actuación de los diversos órganos involucrados. Por lo mismo, parece obligado concluir que tal actuación no habría sido todo lo imparcial que debiera. Más aún, dada la relevancia de los organismos involucrados y sobre todo la del TEDH, -máxima instancia de uno de los sistemas de protección de los DH más avanzados del mundo-, parece obligado concluir también que la característica de la imparcialidad que tradicionalmente se predica como típica de los jueces no sería, hoy por hoy, tal,

Ahora bien, no solo resulta obligado hablar de una presunta ausencia de imparcialidad en los órganos involucrados en el caso, con los corolarios recién señalados, sino también de algo más. Y es que ciertamente lo anterior nos devuelve a la cita del Hamlet de Shakespeare reproducida al inicio de estas páginas, pero no en lo referido a la tardanza de la justicia que, por lo demás, ya hemos tratado implícitamente a propósito de los 15 meses que se tomó la Sección primera de la Sala primera del TC para decidir la inadmisión del recurso de amparo, sino a otro punto de esa misma cita: la insolencia de la autoridad, “the insolence of office”.

En efecto y dicho una vez más con el máximo respeto y en estrictos términos de defensa, resulta obligado calificar de “insolente” la actuación tanto de la CNEAI, cuanto la de la AN, la del TC y la del TEDH. La de la CNEAI suspendiendo, en los términos vistos, las aportaciones científicas de un investigador sin una motivación mínimamente razonable. La de la AN, permitiéndose desatender todos los preceptos legales relativos al papel de la prueba admitida y practicada en la motivación de la sentencia. La del TC tomándose 15 meses para adoptar una decisión de inadmisión “manifiesta” que además no considera violación de derecho fundamental la omisión en una sentencia de la prueba admitida y practicada. La del TEDH, en fin, al permitirse desatender la obligación de motivación de las decisiones de inadmisión establecida literalmente en el art. 45.1 CEDH.

Unas insolencias tanto más evidentes, dicho sea, una vez más, con el máximo respeto y en estrictos términos de defensa, cuanto que se producen en un contexto presuntamente democrático y es que no parece necesario recordar que, desde 1978,

España es un país democrático, -una democracia avanzada incluso-, en el que “La soberanía nacional reside en el pueblo español, del que emanan los poderes del Estado.” y “La justicia emana del pueblo.” según rezan el art. 1.2 y el 117.1 CE respectivamente.

Pero esta última referencia al contexto democrático en el que todas esas insolencias se habrían producido nos obliga, paradójicamente, a salir de la cita de Hamlet a la que acabábamos de volver y ello porque, literalmente, en ella la única solución para el que sufre:

“los ultrajes y desdenes del tiempo, el agravio del opresor, la afrenta del soberbio, los sufrimientos del amor desairado, la tardanza de la justicia, la insolencia de la autoridad, o el maltrato que de los indignos recibe el mérito paciente”

sería el suicidio:

“¿... él mismo podría acabar con todos sus males con un puñal?”

Una solución puramente pasiva que ciertamente resulta impensable, inasumible para cualquier ciudadano de una democracia. Ahora bien y afortunadamente en lo que aquí toca, Shakespeare no escribió solo Hamlet.

2.- “THE FIRST THING WE DO, LET’S KILL ALL THE LAWYERS.”.

Y en efecto, también escribió Henry VI en cuya parte segunda, escena segunda, figura ese “*The first thing we do, let’s kill all the lawyers.*”, que acabamos de reproducir. Una afirmación archiconocida que, por lo demás, se inserta en una larga tradición de lemas y proclamas contra los juristas que cabría remontar a los orígenes mismos del derecho romano.

Y es que ya en la misma Roma podría apreciarse esa consideración negativa de los juristas en el paso de la jurisprudencia pontifical a la laica hacia el 300 a.C., con el agradecimiento que el pueblo mostró a Gnaeus Flavius. Un liberto que robó el libro en el que se recogían las acciones procesales entregándolo al pueblo, tal y como se

relata en el texto de Pomponio recogido en el Digesto³⁶¹. Algo que, al parecer, no dejó de suscitar enconada resistencia entre los juristas-sacerdotes³⁶².

Asimismo, en la Edad Media cabe encontrar esa consideración negativa de los juristas, especialmente virulenta en lo que respecta a los abogados como puede seguirse de la propia glosa que acompaña a su patrón en países como Canadá o Estados Unidos, San Ivo (1253-1303), “abogado, pero no ladrón, cosa increíble para el pueblo”: “*Sanctus Iuus erat britio, advocatus et non latro, res miranda populo*”³⁶³,

Y también cabe apreciar esa consideración negativa a lo largo de la Edad Moderna hasta la Edad contemporánea tanto en la proclama de la España de Felipe V: “*Todo el mal nos viene de los togados*”³⁶⁴, como en la Revolución Francesa con la supresión en 1793 de los centros de formación de juristas: las facultades de derecho³⁶⁵, cuanto en el lema: “*Pas d’avocats*” durante la Comuna de París³⁶⁶.

Una consideración peyorativa, negativa de los juristas que, desde luego, sigue presente en nuestros días con afirmaciones como la famosa de R. Frost: “*A jury consists of twelve persons chosen to decide who has the better lawyer.*”; la no menos famosa de W.S. Gilbert: “*That whether you’re an honest man or whether you’re a thief, depends on whose solicitor has given me my brief.*”³⁶⁷, o, en fin y entre nosotros, las de A. del Moral:

³⁶¹ « ... Luego, derivadas de estas leyes, se formaron casi a un tiempo las acciones con las que los hombres pudieran pleitear entre sí; las cuales acciones, para que el pueblo no las practicara a su arbitrio, quisieron que fueran determinadas y solemnes; y esta parte del derecho se denomina «acciones de la ley» es decir, acciones legítimas. Y así, casi en la misma época nacieron estos tres derechos a partir de la ley de las Doce Tablas: de ellas derivó el derecho civil y, de ellas procedió la formación de las acciones de la ley. Pero la ciencia de la interpretación de estas leyes, así como el conocimiento de las acciones se hallaban en el colegio de los pontífices, de los que se designaba uno que cada año atendiera a los pleitos privados. Y el pueblo se atuvo a esta costumbre durante casi cien años. Posteriormente, habiendo Apio Claudio reunido y puesto por escrito estas acciones, su escriba Gneo Flavio, hijo de un liberto entregó al pueblo el libro que había sustraído y hasta tal punto agradó al pueblo este regalo que le nombró tribuno de la plebe, senador y edil curul.” DIGESTO 1,2,9

³⁶² «Cicerón insinúa que después de la publicación del calendario por Cneo Flavio, los pontífices redactaron las fórmulas procesales con el fin de tener todavía un secreto que guardar.» Vid. García Garrido, M. J., y Eugenio, F.: *Estudios de Derecho y Formación de Juristas.*, Dykinson, Madrid 1988, p. 38.

³⁶³ Del Moral García, A.: Prólogo a Aparisi Miralles, A.: *Ética y deontología para juristas*. Eunsa, Pamplona, 2006, p. 21.

³⁶⁴ Capella, J.R.: “Sobre la extinción del derecho y la supresión de los juristas” en *Materiales para la crítica de la Filosofía del Estado*, Fontanella, Barcelona, 1976, p. 66.

³⁶⁵ Koschaker, R. P., *Europa y el Derecho Romano*, Ed. Revista de Derecho Privado, Madrid, 1955, p. 291.

³⁶⁶ Ibid.

³⁶⁷ Gilbert, W.S. cit. por Rivlin, G.: *Understanding the Law*. Oxford University Press, Oxford 2004, p. 85.

“La imagen del jurista que maneja las leyes a su antojo para que se acomoden a sus intereses, sin freno para manipularlas puede haber sido objeto de caricaturización o exageración pero responde a una cierta realidad, a una percepción social cronificada, más o menos matizada o matizable y desde luego muy simplificadora pero con fundamento real.”³⁶⁸

Pero no procede seguir abundando en la historia de esa animadversión contra los juristas, si bien resulta obligado tomar nota de su antigüedad y persistencia. Y, por supuesto, tampoco procede tomar literalmente la cita del Henry VI, y propugnar el exterminio -“Let’s kill”- de los letrados. Ni siquiera procede adoptar una versión más suave como sería la de defender la supresión del derecho y de los juristas, sin exterminarlos físicamente y es que todo ello sería contradictorio con la confianza en el derecho y en los juristas que se sigue del hecho de haber agotado, en el caso que hemos reexaminado, todas las vías llegando hasta el TEDH.

Ahora bien y una vez vueltos al caso, sí que procede tomar nota de la actitud activa implícita que se sigue tanto de la cita del Henry VI como de las que acabamos de reproducir al ilustrar la milenaria consideración peyorativa de los juristas. Una actitud activa en claro contraste con la pasiva: el suicidio que se seguía de la cita de Hamlet y que ciertamente, en cuanto ciudadanos de una democracia y además teóricos del derecho, nos obliga a proponer soluciones alternativas para que una situación semejante a la del caso expuesto no vuelva a producirse. Propuestas que completen la implementación de la IA en la administración de justicia, insuficiente como se sigue de lo anterior.

Pero, a todo esto, quizás se objete que estamos yendo demasiado lejos y que del análisis de un solo caso estamos derivando la necesidad de una reforma jurídica general. Y es que parece razonable pensar que aducir un solo caso en el que además somos parte no es suficiente para realizar propuestas *de lege ferenda* como las que presuntamente estaríamos obligados a hacer en nuestra doble condición de ciudadanos de una democracia y teóricos del derecho. Serían necesarios más casos de una gravedad semejante y ello dando por supuesto que la del presente haya quedado probada.

A esa improbable objeción cabría responder ante todo mencionando otros casos en los que no hemos sido parte, como, sin ir más lejos, la reciente condena del Fiscal general del Estado por parte de la Sala Segunda del Tribunal Supremo (TS). Condena de la que, en el momento en que escribimos, aún no tenemos sentencia, por lo que no podemos entrar en un análisis pormenorizado si bien resulta suficiente con señalar al respecto que se han cumplido los pronósticos según los cuales los 5 miembros conservadores del TS acabarían condenándole frente a los 2 miembros progresistas. Algo que permite profetizar, por decirlo en términos del realismo jurídico norteamericano, lo que ocurrirá en el previsible recurso de amparo ante el TC y que, desde, luego abunda

³⁶⁸ Del Moral García, A., *Prólogo*, cit. p 21.

en la presunta falta de imparcialidad de los órganos concernidos a la que hemos llegado más arriba.

Por lo demás y también frente a esa improbable objeción cabría aducir otro caso en el que asimismo nos hemos visto involucrados: una recusación por cuestión litigiosa pendiente que sorprendentemente no fue admitida pese a haberse aportado toda la documentación al respecto. Un caso en el que, sin embargo, no vamos a entrar no solo por razones de espacio sino porque no se llevó tan lejos como el aquí reexaminado, dejándolo en el nivel del Juzgado de lo Contencioso, sin recurrir en su momento. Decisión que, ahora a la vista de todo lo anterior, parece más que acertada toda vez que casi con certeza habría llevado al mismo resultado.

Por lo que toca a la objeción de que somos parte en el caso aquí re-expuesto, ya la habríamos justificado más arriba al exponer que nos veíamos obligados a incumplir el brocardo “de nobis ipsis silemus”. Y en efecto, resulta obvio, por una parte, que la omisión en la sentencia de la AN del resultado de la prueba admitida y practicada impedía, dicho sea, una vez más, con el máximo respeto y en estrictos términos de defensa, constatar que nos fue muy favorable, cosa que hemos podido señalar nosotros gracias al incumplimiento de dicho principio. Por otro lado, también resulta evidente que las directrices del TEDH, como también señalamos en su lugar, según las cuales los expedientes de las demandas inadmitidas se destruyen al cabo de un año, hacían imposible constatar que el juez único incumplió, dicho sea, una vez más, con el máximo respeto y en estrictos términos de defensa, la obligación de motivar las decisiones de inadmisión que le impone el 45.1 del CEDH. Todo lo cual nos obligó a incumplir dicho principio “de nobis ipsis silemus”.

Finalmente, y en lo que respecta a la prueba de la gravedad del caso, estaría más que demostrada dada la, por así decirlo y volviendo a Hamlet, naturalidad de las “insolencias” de los órganos involucrados en el caso lo que ciertamente exime de la necesidad de aportar más casos de gravedad parecida. Y en efecto resulta más que reveladora tanto la ligereza con la que se tomó la decisión irracional de la “reformatio in peius” origen del caso por parte de la CNEAI, cuanto la de la omisión de la prueba admitida y practicada en la sentencia de la AN y ello tanto más cuanto que se planteó un incidente de nulidad de actuaciones al respecto. Asimismo, resulta más que reveladora la naturalidad tanto de la decisión de inadmisión del recurso de amparo por parte del TC en los términos vistos, cuanto la del TEDH al inadmitir la correspondiente demanda contra el Reino de España, también en los términos visto.

Y es que, desde luego, decisiones tan naturales no solo resultan inasumibles, inaceptables en un contexto democrático en el que el derecho y la justicia emana del pueblo sino en cualquier contexto. Parafraseando a Hart, resulta evidente que las decisiones mínimamente racionales y la inclusión en una sentencia de la prueba admitida y

practicada son partes obligadas del contenido mínimo no ya del derecho natural³⁶⁹, sino de cualquier derecho.

Así las cosas, podemos abandonar ya la actitud pasiva de la cita de Hamlet a la que nos habría devuelto la objeción de que precisaríamos de mas casos de parecida gravedad para proponer reformas. Podemos volver, pues, a la actitud activa del Henry VI y ello tanto más cuanto que los propios jueces habrían reconocido la necesidad de una ética judicial más allá de la deontología profesional³⁷⁰. Algo que ciertamente constituye un reconocimiento, por mínimo que sea, de que su actuación no se corresponde con la que debiera.

Un reconocimiento mínimo de mala praxis por parte de los jueces que, desde luego, no nos libera de formular propuestas de reforma por nuestra parte y ello tanto más cuanto que una de las obligadas que se sigue de lo anterior sería precisamente la de que uno de los requisitos *sine qua non* para ser juez, más allá de la mera oposición o del ejercicio de un número determinado de años como letrado, sería el de haber sido parte en un proceso. Idealmente, haber sufrido avatares como los relatados aquí o semejantes: decisiones irracionales, incumplimiento de la legislación, falta de motivación... Requisito que, parece razonable pensar, contribuiría a que no se reprodujeran.

3.-ALGUNAS PROPUESTAS

Pero no es esa la única propuesta de reforma jurídica que se sigue de lo anterior y es que, sin que podamos entrar a fondo, por razones de espacio, debiendo limitarnos a esbozarlas, cabría indicar las siguientes propuestas que se deducen lógicamente de la actuación de la CNEAI, AN y TC en el contexto de un país democrático y de la del TEDH, cúspide de uno de los sistemas de protección de los Derechos Humanos más avanzados del mundo. Propuestas que vendrían a complementar la implementación de la IA ya referida que, como asimismo se ha señalado resultaría insuficiente.

Ante todo, y paradójicamente la implementación de la IA, -una tecnología recentísima que apenas si está en sus comienzos-, se vería acompañada de la recuperación de esa milenaria consideración peyorativa de los juristas a la que el análisis de la actuación de los diversos órganos involucrados en el caso nos ha devuelto y que, como vimos, se remontaría al comienzo mismo de derecho romano.

Esa recuperación se daría en varios niveles. En primer lugar, en el de la propia denominación de las facultades de Derecho que, como es bien sabido, han ido cambiando en los últimos años pasando a denominarse de Ciencias Jurídicas como sin ir más lejos se sigue de la denominación que recibe en la Universidad desde la que escribimos “Facultad de ciencias jurídicas y políticas”.

³⁶⁹ Hart, H. L. A., *The Concept of Law*, Clarendon Press, Oxford 1981, p 189 y ss..

³⁷⁰ Vid. Hermida del Llano, C., “De la ética a la ética judicial a través del Estado de Derecho” en Hermida del Llano, C., *La libertad de expresión del juez a la luz de ética judicial*. Dykinson, Madrid 2025, pp. 18 a 24

Una vez más por razones de espacio, no procede entrar en el tema de la cientificidad del derecho que, como es bien sabido, cabría remontar a J. von Kirchmann³⁷¹ así como tampoco procede entrar en las diversas acepciones del término ciencia alguna de las cuales ciertamente, de un modo u otro, avalarían la consideración científica del derecho. Ahora bien, sí que resulta obligado señalar que desde esa denominación “Facultad de ciencias jurídicas” no cabe tratamiento alguno de la milenaria consideración peyorativa de los juristas a la que acabamos de llegar, por lo que procede reivindicar ese cambio de denominación a Facultades de Derecho. Denominación que permite incluir tanto la consideración del derecho como ciencia cuanto la milenaria consideración peyorativa de los juristas.

Y no solo de ellos, pues, esa denominación “Facultades de Derecho” permite incluir también la consideración peyorativa del Derecho mismo. Algo a lo que no hemos hecho mención hasta aquí pero que se sigue lógicamente de la propia consideración negativa de los juristas. Y en efecto, esa consideración negativa del derecho aparecería ya en el momento inaugural de uno de los pilares de la cultura europea; el cristianismo como se sigue inmediatamente de la célebre cita del Evangelio de S. Mateo:

*“Habéis oído que se dijo: ojo por ojo diente por diente. Pero yo os digo: No resistáis al mal, antes bien, al que te abofetee en la mejilla derecha ofrécele también la otra; al que quiera pleitear contigo para quitarte la túnica, déjale también el manto; al que te obligue a andar una milla, vete con él dos. A quien te pida da y al que desee que le prestes algo no le vuelvas la espalda.”*³⁷²

Una cita desde la que, ciertamente, resulta muy difícil si es que no imposible no sólo equiparar al derecho con una ciencia y a sus cultores con científicos, sino incluso justificar la existencia misma de jueces, abogados, fiscales.... Una cita que, por lo demás, sigue resonando en nuestros días.³⁷³

Esa misma consideración peyorativa del derecho aparece también en otra de las constantes históricas en el pensamiento occidental: el anarquismo, la defensa de una sociedad sin normas³⁷⁴. Y también en el pensamiento comunista que, como es bien sabido, considera el derecho como una superestructura incompatible con la sociedad comunista tanto en su forma de ley como en su forma de derecho en general³⁷⁵.

³⁷¹ Vid, Kirchmann, J.H. von, *Die Wertlosigkeit der Jurisprudenz als Wissenschaft*. Rudolf Huafe Verlag, Berlin 1990.

³⁷² Mateo 5, 38-48

³⁷³ Vid. La Torre, M., *Los juristas malos cristianos*, en *Derecho y libertades*, 2003. Año 8, n. 12 pp. 72-108.

³⁷⁴ Vid. Rivaya, B., *Anarquismo y Derecho*, Revista de Estudios Políticos, 112, Madrid 2001, p.77-108.

³⁷⁵ “... en la *Ideología alemana* escrita por Marx y Engels conjuntamente, no solamente es declarado incompatible con el comunismo el Derecho en forma de ley, sino el Derecho en general. Entre el Derecho y el comunismo existe una antítesis (Gegensatz) aún cuando se presente en su forma más general como derecho del hombre”. FASSO, G., *Historia de la Filosofía del Derecho*. Pirámide, Madrid 1979, p. 115.

Constantes históricas de nuevo difícilmente compatibles con la consideración del derecho como ciencia y la consiguiente conversión de las facultades de “derecho” en facultades de “ciencias jurídicas”.

Así las cosas, resulta obligada la propuesta de cambio en la denominación de las actuales Facultades de Ciencias Jurídicas por Facultades de Derecho toda vez esa última denominación incluye las dos consideraciones del fenómeno jurídico; la científica a la que se habría llegado en los últimos siglos y la peyorativa milenaria de la maldad del derecho y de los juristas a la que la actuación de la CNEAI, AN, TC y TEDH nos ha devuelto.

La segunda de las propuestas de reforma jurídica que se sigue lógicamente de dicha actuación se situaría en el nivel de enseñanza del derecho, en el de la formación jurídica. Nivel en el que, junto a las consideraciones científicas, por así decirlo, se incluirían también las correspondientes a las de la maldad del derecho y de los juristas. Los estudiantes aprendices de derecho, los futuros operadores jurídicos además de la consideración del derecho que, por ejemplo, se sigue de la definición de Celso, recogida por Ulpiano en el Digesto:

*“Conviene que el que ha de darse al estudio del derecho, conozca en primer lugar de donde viene la palabra ius. Llámase así de iustitia pues, como elegantemente lo define Celso el derecho es el arte de lo bueno y de lo justo. Por cuya razón se nos llama sacerdotes, ya que cultivamos la justicia y profesamos el conocimiento de lo bueno y de lo justo, separando lo justo de lo injusto, discerniendo lo lícito de lo ilícito, deseando hacer buenos a los hombres, no sólo por la amenaza de las penas sino también por la invitación de los premios. Buscando con ansia si no me equivoco la filosofía verdadera no la aparente.”*³⁷⁶,

tendrían información cumplida de la consideración contraria implícita, por ejemplo, en *El Príncipe* de Maquiavelo:

“Pero, siendo mi propósito escribir algo útil para quien lo lea, me ha parecido más conveniente ir directamente a la verdad real de la cosa que a la representación imaginaria de la misma. Muchos se han imaginado repúblicas y principados que nadie ha visto jamás ni se ha sabido que existieran realmente; porque hay tanta distancia de cómo se vive a cómo se debería vivir, que quien deja a un lado lo que se hace por lo que se debería hacer, aprende

³⁷⁶ *Iuri operam daturum prius nosse oportet, unde nomen iuris descendat. est autem a iustitia appellatum: nam, ut eleganter Celsus definit, ius est ars boni et aequi. . Cuius merito quis nos sacerdotes appellet: iustitiam namque colimus et boni et aequi notitiam profitemur, aequum ab iniquo separantes, licitum ab illicito discernentes, bonos non solum metu poenarum, verum etiam praemiorum quoque exhortatione efficere cupientes, veram nisi fallor philosophiam, non simulatam affectantes* Digesto 1,1,1.

*antes su ruina que su preservación, porque un hombre que quiera hacer en todos los puntos profesión de bueno, labrará necesariamente su ruina entre tantos que no lo son.”*³⁷⁷ .

Más en concreto y en los términos que aquí interesan. De una parte conocerían el derecho que, en este caso, debería ser: la CNEAI dictando una resolución mínimamente razonable; la AN cumpliendo los preceptos legales e incluyendo la prueba admitida y practicada en la motivación del fallo; el TC dictando una resolución congruente con el tiempo, -esto es, no manifiesta tras 15 meses-, y defendiendo que la no inclusión en una sentencia de la prueba admitida y practicada viola el derecho a la tutela judicial efectiva y, finalmente, el TEDH cumpliendo la legalidad vigente, el art. 45.1 CEDH, y motivando las decisiones de admisión/inadmisión.

De otra, serían informados del derecho que es: la CNEAI dictando una “reformatio in peius” completamente irracional en los términos vistos; la AN incumpliendo los preceptos legales y omitiendo en la sentencia la prueba admitida y practicada; el TC afirmando que algo es manifiesto tras 15 meses de deliberación y defendiendo que la no inclusión de la prueba en la sentencia no vulnera derecho fundamental alguno y, finalmente, el TEDH incumpliendo el 45.1 CEDH y no motivando la decisión inadmisión.

Así las cosas, no parece necesario insistir en que todo ello exige, hoy por hoy, un nuevo modelo de profesor de derecho con la metodología docente correspondiente. Un profesor y una metodología duales, bifrontes si se prefiere. Y es que, de una parte, seguiría presente el profesor y la metodología docente implícitos en la consideración científica del derecho; del derecho que debería ser en los términos recién vistos. De otra parte, ese mismo profesor tendría que ser capaz de explicar el derecho que es también en los términos recién vistos más arriba-

Un profesor dual, bifronte, Celso y Maquiavelo reunidos, por así decirlo. Un científico, explicando el derecho que debe ser, el legalmente establecido en este caso, junto a propuestas *de lege ferenda* elaboradas conjuntamente por los estudiantes y a la vez un “abogado del diablo”, también por así decirlo, explicando el derecho que es. Un abogado del diablo que enseñaría a los futuros operadores jurídicos a ser precavidos, a ser conscientes de las trampas, trucos, triquiñuelas... que encontrarán, hoy por hoy, en el derecho que es y ello de modo que, en la propia enseñanza, tendría que usar trampas, trucos, triquiñuelas.... al efecto. Algo en lo que, por razones de espacio, no podamos entrar de manera pormenorizada, debiendo remitir al lugar donde lo hemos esbozado

³⁷⁸ .

³⁷⁷ Maquiavelo, *El príncipe*, Alianza, Madrid 1993, p. 83.

³⁷⁸ Prada García, A. de, *El profesor como Abogado del Diablo: Sobre la enseñanza de la Teoría del Derecho* en Santos Arnaiz, J.A. (Coord.), *Tendencias y propuestas de innovación educativa en Derecho*, Atelier Madrid 2026, p. 331-345 (en prensa).

Así las cosas cabría concluir que las propuestas de reforma recién apuntadas unidas a la implementación de la IA en los términos vistos más arriba, así como a la propuesta de que para ser juez habría que haber sido parte previamente e, idealmente, haber sufrido avatares como los aquí relatados o semejantes, (esto es, haber experimentado el derecho que, hoy por hoy, es), harían posible un acercamiento del derecho que es al que debe ser, el legalmente establecido en este caso, impidiendo situaciones como las aquí descritas.

4.- ¿PLAZOS RAZONABLES Y DILACIONES INDEBIDAS?

Pero, a todo esto, no hemos incluido en el anterior listado de propuestas nada relativo al retraso en la justicia, al tema de la justicia en plazo y sin dilaciones que figura en el título de estas líneas... Ciertamente que hemos tratado el tema al hablar de los 15 meses que se tomó la Sección primera de la Sala primera del TC para inadmitir el recurso de amparo basándose en la “manifiesta inexistencia de violación de derecho fundamental tutelable en amparo”, pero lo hicimos de pasada, tangencialmente sin poner en el tema el énfasis que correspondía a su presencia en el título de estas líneas.

Una omisión tanto más grave cuando que, como vimos, la justicia tardía, retrasada sería uno de los defectos más graves del derecho que es, por utilizar la terminología a la que hemos llegado. Y ello siquiera sea por las repetidas quejas secularmente formuladas al respecto y a los que nos referimos más arriba: tanto la sabiduría popular en dichos como “la justicia tardía no es justicia”, o “justicia retrasada justicia denegada” cuanto citas de autores fundamentales de la historia de la literatura como Shakespeare en Hamlet, Una cita a la que hemos ido y vuelto a lo largo de estas líneas y en la que, como se recordará, se señalaba el retraso de la justicia como uno de los hechos que justificarían incluso el suicidio. Todo lo cual ciertamente lleva a la conclusión obligada de la reivindicación de la justicia en plazo, sin dilaciones, como derecho que debe ser, por volver a utilizar la terminología a la que hemos llegado.

Ahora bien, afortunadamente, aún podemos tratar el tema con la atención debida. Y es que, frente a lo que pudiera parecer, el caso no se cerró definitivamente con la decisión de inadmisión por parte del TEDH de la demanda contra el Reino de España en 2016 a la que nos hemos referido más arriba, sino que se ha vuelto a abrir estando ahora mismo en la sección Sexta de la Sala de lo Contencioso Administrativo (P.O. 0000989/2023). La misma sección en la que se sustanció el recurso al que nos hemos referido más arriba y en el que no se consiguió probar que la CNEAI prohibía de facto la “reformatio in peius”, toda vez que la prueba practicada, muy favorable al recurrente, como se apuntó, desapareció de la Sentencia como también se vio.

Pues bien, gracias a la Ley 19/2013 de Transparencia, acceso a la información pública y buen gobierno (LTAIBG) y a una larga batalla de siete años, desde 2016 a 2023, con varias intervenciones del Consejo de Transparencia y Buen Gobierno

(CTBG) e incluso del Defensor del Pueblo, se consiguió probar que efectivamente no cabe *reformatio in peius* alguna dado el reconocimiento expreso al respecto por parte de la ANECA/CNEAI en respuesta a una solicitud formulada a través del portal de transparencia el 14 de febrero de 2023. Literalmente:

“Hemos recibido un escrito con número de registro REGAGE23e00004128702 en el que nos pregunta si se mantiene la nota de las aportaciones que recupera de una evaluación negativa. Le informamos que es así, la nota debe mantenerse y mejorarse, nunca puede ser inferior (reformatio in peius).”

Obviamente y una vez más por razones de espacio no procede entrar en detalle en toda esta última batalla y cómo se ha llegado a esa solicitud debiendo limitarnos a remitir a los escritos donde la hemos reflejado³⁷⁹. Batalla y escritos que, ciertamente y violando de nuevo el brocardo “de nobis ipsis silemus”, no dejan de volver sobre nuestra confianza en el derecho que debe ser y en los juristas a la que más arriba aludíamos y que nos llevó a agotar todas las instancias posibles hasta llegar al TEDH y después a seguir peleando, por así decirlo, de la mano de la LTAIBG. Una confianza en que el derecho que debe ser acabe siendo, obligada en nuestra doble condición de ciudadano de una democracia y teórico del derecho como también señalamos más arriba.

Ahora bien, sí que resulta obligado llamar la atención sobre el más que dilatado período de tiempo transcurrido hasta ese reconocimiento por parte de la CNEAI. Y es que resulta sorprendente, por decirlo de algún modo, que con la LTAIBG en vigor desde 2013, la ANECA-CNEAI se tomase nada menos que siete años, con sucesivas intervenciones del CTBG e incluso del Defensor del Pueblo, para reconocer finalmente, por escrito, que no cabe “reformatio in peius” alguna.

Algo que, de un modo u otro, no solo nos devuelve al tema de la tardanza de la justicia sino a las propuestas de reforma jurídica a las que más arriba llegamos y, más en concreto, a la de un profesor “abogado del diablo” que ilustre a sus alumnos de las trampas, trucos, triquiñuelas... que encontrarán en el derecho que, hoy por hoy, es. Todo ello por supuesto con el máximo respeto y en estrictos términos de defensa.

³⁷⁹ Vid. Prada García, A. de, *Un banco de pruebas para la Ley de Transparencia: La evaluación de la actividad investigadora*, en Serrano Maillo, I, (Coord.) *La aplicación de la transparencia*. Universidad Complutense, Madrid 2018, pp. 15-24. Prada García, A. de, *Evaluación de la actividad investigadora y posjusticia: contra la “reformatio in peius”* en López Ramón, F. y Valero Torrijos, J. (Coords.): *20 años de la ley de lo contencioso administrativo. Actas del XIV Congreso de la Asociación Española de Profesores de Derecho Administrativo*. INAP, Madrid 2019, pp.147-158. Prada García, A. de, *Posverdad y derecho a los hechos: A propósito de una Resolución del CTB (R 0198/2017)* en Sánchez de Diego, M, y Sierra Rodríguez, J, (Coords): *Transparencia y participación para un gobierno abierto*. Wolters Kluwer, Madrid 2020, p.385-400 y Prada García, A. de, *Posverdad y derecho a los hechos: A propósito de una Resolución del CTB (R 0198/2017)* en Sánchez de Diego, M, y Sierra Rodríguez, J, (Coords): *Transparencia y participación para un gobierno abierto*. Wolters Kluwer, Madrid 2020, p.385-400.

Pero si ese más que dilatado período de tiempo que se tomó la ANECA-CNEAI para ese reconocimiento resulta sorprendente, todavía lo es más su actuación posterior. Y es que ante el consiguiente recurso para que corrigiera la “reformatio in peius” de 2007, del caso que nos ocupa, no procedió de acuerdo con el derecho que debe ser; con el principio general del derecho de no ir contra los propios actos, “nemo contra factum proprium venire potest”. Principio que, una vez reconocida la improcedencia de cualquier “reformatio in peius”, obligaba ciertamente a corregir la de 2007 atendiendo al recurso planteado al efecto.

Pues bien, en lugar de ello, en lugar de atenerse a ese principio, la ANECA-CNEAI desestimó por vía de silencio negativo el recurso planteado incumpliendo, además, la obligación de resolver que imponen los artículos 21 y 24 de la Ley 39/2015 según los cuales la administración está obligada a dictar resolución expresa y a notificarla en todos los procedimientos cualquiera que sea su forma de iniciación.

Obviamente todo ello no viene sino a reforzar una vez más las propuestas que más arriba formulamos a partir de la distinción entre el derecho que debe ser (el legalmente establecido, en este caso) y el derecho que es. Distinción a la que llegamos tras el análisis de la actuación de la CNEAI, AN, TC y TEDH. Pero no es momento de abundar en ello ya que el caso ahora mismo, tras la interposición de un nuevo recurso contencioso administrativo, está de nuevo en la misma Sección Sexta de la Sala de lo Contencioso de la AN, pendiente de fijación de fecha para votación y fallo. Algo que nos permite abordar, por fin, con la atención debida, la tardanza de la justicia que figura en el título de estas líneas.

Y en efecto, la diligencia en la que se declaraban las actuaciones conclusas y pendientes de fijación de fecha para señalamiento y fallo fue notificada el 22 de enero de 2024. Así las cosas, a fecha de hoy, en el momento en que escribimos, diciembre de 2025, casi dos años después, y tras un escrito de impulso procesal, por decirlo técnicamente, de fecha 17 de noviembre de 2025, seguimos a la espera de la fijación de fecha para votación y fallo pese al plazo máximo de 6 meses que establece la ley 39/2015. Casi dos años, pues, esperando a que se fije la fecha para votación y fallo. Algo tanto más sorprendente cuanto que las partes, recurrente y abogado del estado, estamos de acuerdo en la petición subsidiaria: que se dé traslado al Consejo de Estado para que evacúe el informe correspondiente.

Ciertamente no procede extenderse en el descrédito y la pérdida de legitimidad que, dicho una vez más con el máximo respeto y en estrictos términos de defensa se sigue de esos, -de momento-, 23 meses, esperando se fije la fecha para votación y fallo, cuando el plazo legalmente establecido, el derecho que debe ser, son 6 meses. Ahora bien, dado que hemos llegado al punto en que nuestra escritura y el momento procesal del caso coinciden, también es el momento de precisar nuestras propuestas de reforma jurídica al respecto, toda vez que ese punto se refiere al tema último de este escrito, el retraso en la impartición de justicia y la IA.

Algo que, por lo demás, no plantea excesiva dificultad pues no parece necesario insistir en que la aplicación de cualquier sistema de IA habría acortado hasta casi hacer innecesario plazo alguno superior al legalmente establecido: 6 meses, mediante la presentación de borradores de sentencia a ser corregidos por el ponente antes de someterlo a la votación de la Sala y ello tanto más cuanto que las partes, como se ha señalado, estamos de acuerdo en la petición subsidiaria.

Más aún, creemos que la implementación de las propuestas de reforma jurídica esbozadas más arriba, junto con la que sigue de esta última circunstancia: elaboración de borradores de sentencia por parte de la IA para ser corregidos por el ponente y el resto de los magistrados, harían que el derecho que debe ser, sea y en lo que toca a la justicia en plazo, sin dilaciones se refiere facilitarían al máximo la labor de los jueces y llevaría a dictar sentencia en plazo haciendo innecesarias precisiones como las que figuran en nuestra Constitución o en el CEDH.

Y en efecto, tras lo anterior, pensamos que la aplicación generalizada de la IA junto con las demás propuestas aquí realizadas acabaría obligando, antes o después, a suprimir el término “razonable” con el que, en el art. 6.1 del CEDH, se califica el plazo en el que toda persona tiene derecho a que su causa sea oída, dejando simplemente “en plazo”, esto es en el legalmente establecido. Por lo mismo también creemos que, antes o después, decaería la elaborada doctrina del TEDH respecto del “plazo razonable”, manifestada entre otras, en la sentencia del TEDH de 20 de diciembre de 2016, asunto Ruiz Villar contra España, en la que reiteró que: “la razonabilidad en la duración del procedimiento debe evaluarse a la vista de las circunstancias del caso y en referencia a los siguientes criterios: complejidad del caso, comportamiento del demandante y autoridades, y aquello que el demandante arriesga en el litigio”.

Igualmente pensamos, tras lo anterior, que esa aplicación generalizada de la IA junto con las demás propuestas aquí esbozadas acabaría obligando, antes o después, a suprimir el adjetivo “indebidas” con el que se califica a las dilaciones en el art. 24.2 CE, dejándolo simplemente como un derecho “a un proceso público sin dilaciones”; esto es, ni debidas ni indebidas. Una supresión que conllevaría, claro es, el decaimiento de la elaborada doctrina al respecto y de la que puede servir de muestra lo afirmado en la STC 178/2007, de 23 de julio: “...el juicio sobre el contenido concreto de las dilaciones, y sobre si son o no indebidas, debe ser el resultado de la aplicación a las circunstancias específicas de cada caso de los criterios objetivos que a lo largo de nuestra jurisprudencia se han ido precisando, y que son la complejidad del litigio, los márgenes ordinarios de duración de los litigios del mismo tipo, el interés que arriesga el demandante de amparo, su conducta procesal y la conducta de las autoridades.”³⁸⁰

³⁸⁰ Rojo Torrecilla, E., *Derecho a la tutela judicial efectiva por excesiva dilación en la fecha de fijación del juicio: los vaivenes del TC*. Revista de Jurisprudencia Laboral. Número 6/2023, p. 6 y 7.

5. A MODO DE CONCLUSIÓN

Así pues, tras lo anterior, pensamos que, antes o después, la aplicación generalizada de la IA junto con las demás propuestas aquí esbozadas llevaría a contribuir a que el derecho que debe ser, sea y, más en concreto, a una justicia en plazo y sin dilaciones cumpliendo así aspiraciones centenarias de los ciudadanos expresadas en dichos y refranes como los recogidos más arriba: “la justicia tardía no es justicia”, “justicia retardada, justicia denegada” o en la cita de Hamlet que encabeza estas líneas y en la que se señalaba la tardanza de la justicia como uno de los hechos que incluso justificarían el suicidio.

Antes o después decimos, pero la naturalidad con la que, dicho sea por última vez con el máximo respeto y en estrictos términos de defensa, la CNEAI, la AN, el TC y el TEDH han incumplido el derecho que debe ser, el legalmente establecido en un contexto democrático en el que el derecho y la justicia emanan del pueblo, obliga a concluir que la tarea, -la lucha más bien-, por conseguir que el derecho que debe ser, sea y, más en concreto, que se llegue a una justicia en plazo es, si se nos permite la licencia, no ya a corto o a medio, sino a largo, a muy largo plazo.

BIBLIOGRAFÍA UTILIZADA

- Capella, J.R., “Sobre la extinción del derecho y la supresión de los juristas” en *Materiales para la crítica de la Filosofía del Estado*, Fontanella, Barcelona, 1976.
- Carretero, S., *Humanismo tecnológico: justicia predictiva y derechos humanos en la era digital*, Co-lex, Madrid 2025.
- Del Moral García, A., *Prólogo* a Aparisi Miralles, A.: *Ética y deontología para juristas*. Eunsa, Pamplona, 2006
- Domínguez Angulo, J. P. y Ortiz Bolaños, L., *Perspectiva Ontológico-hermenéutica de la Inteligencia Artificial. Una colección de métodos para su mejoramiento y delimitación en la interpretación del derecho* en R. Opin. Jur., Fortaleza, año 23, n. 44, 2025.
- García Garrido, M. J., y Eugenio, F.: *Estudios de Derecho y Formación de Juristas.*, Dykinson, Madrid 1988.
- Koschaker, R. P., *Europa y el Derecho Romano*, Ed. Revista de Derecho Privado, Madrid, 1955.
- Hart, H. L. A., *The Concept of Law*, Clarendon Press, Oxford 1981.
- Hermida del Llano, C., “De la ética a la ética judicial a través del Estado de Derecho” en Hermida del Llano, C., *La libertad de expresión del juez a la luz de ética judicial*. Dykinson, Madrid 2025
- Kirchmann, J.H. von, *Die Wertlosigkeit der Jurisprudenz als Wissenschaft*. Rudolf Huafel Verlag, Berlin 1990.
- La Torre, M., *Los juristas malos cristianos*, en Derecho y libertades, 2003. Año 8, n. 12.
- Marín Castán, M. L., *La polémica cuestión de la determinación del «plazo razonable» en la administración de justicia (Comentario a la Sentencia del Tribunal Europeo de Derechos Humanos de 13 de julio de 1983)* Revista española de derecho constitucional, I Año nº 4, Nº 10, Madrid, 1984.
- Maquiavelo, *El príncipe*, Alianza, Madrid 1993
- Prada García, A. de, *Por una reforma del derecho a obtener la tutela efectiva de Jueces y Tribunales: A propósito de dos sentencias de la Audiencia Nacional y una providencia de inadmisión del TC*. Boletín de Ministerio de Justicia. Año LXX, n. 2193, noviembre de 2016.
- Prada García, A. de, *Tomando la prueba, (los viejos y los nuevos derechos) en serio. A propósito de una Sentencia de la Audiencia Nacional P.O. 439/2012*, Revista General de Derecho Procesal 39, 2016.
- Prada García, A. de, *Sombras en la evaluación de la actividad investigadora*. Claves de razón práctica, n. 256, enero-febrero 2018, p. 90-99.
- Prada García, A. de, *¿Más allá de Rashomon? En busca de los cadáveres perdidos*. En Cobo Del Rosal G., Navalpotro Sánchez, J.M. y Cano Linares A. (Dir.) *ARS IURIS. Estudios sobre Derecho y Arte*. Editorial Dykinson, Madrid 2023.
- Rivaya, B., *Anarquismo y Derecho*, Revista de Estudios Políticos, 112, Madrid 2001,
- Rivlin, G., *Understanding the Law*. Oxford University Press, Oxford 2004
- Rojo Torrecilla, E., *Derecho a la tutela judicial efectiva por excesiva dilación en la fecha de fijación del juicio: los vaivenes del TC*. Revista de Jurisprudencia Laboral. Número 6/2023.
- Shakespeare, *Hamlet* Atlantis, London 1980.
- Shakespeare, *King Henry the Sixth* Atlantis, London 1980.

CAPÍTULO 9. GEOPOLÍTICA, GOBERNANZA E INTELIGENCIA ARTIFICIAL³⁸¹

FERNANDO GIL GONZÁLEZ³⁸²

Universidad de Alcalá de Henares

fernando.gilg@uah.es

Abstract

This chapter explores the global context of the deformation of political reality in relation to Artificial Intelligence (AI). The chapter discusses the need to understand the importance of geopolitics for good governance. Likewise, the chapter examines, from a historical perspective, the planetary governance structures that have been characterized through the following three fundamental pillars: The exercise of power over a specific territory; the inequality between the central executive and administered regional structures, and finally, the implementation of a political project through various forms of influence from the political, economic, institutional, social, ideological, and technological perspectives. The chapter concludes with an analysis of the future of AI in public governance practices and processes.

Keywords

Geopolitics, Digital Empires, Challenges, Law and A.I.

Resumen

Este capítulo explora el contexto global de la deformación de la realidad política en relación con la Inteligencia Artificial (I.A). Se aborda la necesidad de comprender la importancia de la geopolítica para la buena gobernanza. Asimismo, se examinan, desde una perspectiva histórica, las estructuras de gobernanza planetaria, caracterizadas por tres pilares fundamentales: El ejercicio del poder sobre un territorio específico; la desigualdad entre el ejecutivo central y las estructuras regionales administradas, y, finalmente, la implementación de un proyecto

³⁸¹ El presente trabajo parte de la ponencia defendida en la I Jornada de Derechos Humanos en I.A, que se celebraron en el Campus de Vicálvaro de la URJC (Madrid) los días 28 y 29 de octubre de 2025 cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content> Es por ello que las referidas Actas en parte recogen de forma extractada algunas partes de este texto que ahora se publica, pp. 17-21.

³⁸² Doctor en Derecho. Profesor Ayudante Doctor de Historia del Derecho en el Departamento de Ciencias Jurídicas en la Facultad de Derecho de la Universidad de Alcalá, Escritor con más de una quincena de libros como autor, Miembro de la Royal Historical Society (London, UK) y del Instituto de Estudios Jurídicos Internacionales de la URJC (Madrid).

político mediante diversas formas de influencia desde las perspectivas política, económica, institucional, social, ideológica y tecnológica. El capítulo concluye con un análisis del futuro de la I.A. en las prácticas y procesos de gobernanza pública.

Palabras clave

Geopolítica, Imperios Digitales, Desafíos, Derecho e Inteligencia Artificial.

Sumario: INTRODUCCIÓN: LA INTELIGENCIA ARTIFICIAL Y LA TRANSMISIÓN DEL CONOCIMIENTO 1. LA CONVERGENCIA DE BIG DATA: APRENDIZAJE AUTOMÁTICO 2. LA CUESTIÓN DE LA SOBERANÍA DE LOS GOBIERNOS 3. EL DUOPOLIO CHINO-ESTADOUNIDENSE: ESTADOS UNIDOS Y CHINA 4. ¿GEOPOLÍTICA DE BLOQUES? LA SITUACIÓN DIGITAL DE LA UNIÓN EUROPEA 5. EPÍLOGO: ¿HACIA DÓNDE NOS DIRIGIMOS? 6. CONCLUSIONES BIBLIOGRÁFICA

INTRODUCCIÓN: LA INTELIGENCIA ARTIFICIAL Y LA TRANSMISIÓN DEL CONOCIMIENTO

La geopolítica está orientada a la cooperación a través del intercambio, las transacciones, negociación y otros elementos que están relacionados. En cuanto a la esfera pública, existe el interés y la voluntad común de los Estados-nación, las reglas de coexistencia o, incluso, las instituciones políticas, que constituyen elementos organizativos de la política internacional (Mialhe et al., 2020: 56-64). De hecho, el sistema global comprende un conjunto de estrategias implementadas por diferentes sujetos y países. En este caso, el papel del liderazgo global y la formulación de la agenda enfatiza las estrategias de trabajo social. Por otro lado, la geopolítica, en el ámbito público, es un instrumento de poder, gobierno, control y dominación que evita subyugación por otro elemento institucional (Mesa, 2014: 107). En otro orden de ideas, el concepto de posverdad (Morales, 2018: 29) nos permite comprender el contexto global encarnada en la deformación de la realidad política. Aun así, es necesario conocer y integrar algunas de las realidades geopolíticas, a pesar de estar gobernadas por la hipocresía, que suele ser de naturaleza despiadada y traicionera (Baños, 2017: 12). Por otro lado, cabe añadir que, históricamente, las estructuras de gobernanza planetaria han sido caracterizado por tres pilares fundamentales. En primer lugar, el ejercicio del poder sobre un grupo específico. territorio. En segundo lugar, la desigualdad entre el ejecutivo central y el administrado estructuras regionales y, finalmente, la implementación de un proyecto político a través de diversas formas de influencia desde el ámbito político, económico, institucional, social, ideológico y puntos de vista tecnológicos. La IA se centra en la concentración del poder en las manos de unos pocos agentes políticos, económicos, militares y sociales

(Bárcena et al., 2018: 33-34) De hecho, los imperios digitales se beneficiarían de las economías de escala y, de esta manera, acelerarían su concentración de poder. Con ello, tales imperios digitales, se convertirían en los principales instrumentos gobernar los asuntos internacionales, y así volver a una lógica de bloques. Tales redes y sus correspondientes instrumentos digitales se extenderían a escala continental, especialmente en los casos de la Estados Unidos y China, mientras que otros actores políticos y supranacionales, como Europa, adoptaría estrategias más conservadoras. Además, la vigilancia está asociada con el uso de cámaras, adentrándose en los lugares públicos como se observa en numerosas mega-ciudades como Londres, Nueva York, Nueva Delhi, Shanghái o Río de Janeiro (Bauman y Lyon, 2013: 9). De hecho, la vigilancia, desde el punto de vista de control, existe en forma de pasaportes, escáneres biométricos y otros modelos de control. Asimismo, existe un destacado control de seguridad en materia de compras online o participación en la política de Internet. Otros métodos de monitoreo social para ahondar en la IA en aspectos de geopolítica comprenden el documento de identificación fiscal, el uso de códigos o contraseñas para acceder a un edificio compra de productos en línea y, por último, búsquedas en internet (Lyon, 2005: 66). Frente a esto, en este contexto, para comprender la geopolítica del siglo XXI, es vital considerar la siguiente pregunta: ¿cuál es la importancia de las nuevas tecnologías? o ¿cuál es su vínculo con la política en el mundo actual? En primer lugar, la idea de la fenomenología. Es necesario destacar la magnitud, entendida como una estrategia coercitiva en el abuso de poder. A continuación, la idea de expansión en el procesamiento de datos (Saura y Palacios, 2021b) para explicar vigilancia o analizar elementos de control en geopolítica e Inteligencia Artificial debe profundizarse. De hecho, entre los siglos XX y XXI, ¡hay que repensar el tema; hasta qué punto nos permiten conocer y tener una visión crítica. Entonces habría que preguntarse ¿qué visión tiene la vigilancia en la actualidad? La respuesta es el enfriamiento líquido, que es un instrumento de guía, además de servir para identificar cambios en seguridad y vigilancia. Sin embargo, la vigilancia del consumo en la sociedad actual es difusa. Por otro lado, son los bits, que comprenden los datos personales extraídos (Janssen, 2020) para un fin o un propósito sin un objetivo fijo, como resultado de la presión derivada de las exigencias de seguridad, publicidad o propaganda. Además, sería necesario analizar el concepto de pos-modernidad desde la época de las teorías marxistas. Posteriormente, para comprender el concepto, es fundamental realizar un análisis pormenorizado en relación con algunos expertos como G. Deleuze, que discute la idea de la sociedad de control crece vertiginosamente como una enredadera (Deleuze, 1992: 3) Sin embargo, la vigilancia actual está incentivada por los productos básicos, es decir, todo lo que ha sido establecido como líquido, termina disolviéndose en algún momento de la historia. Por otro lado, algunos expertos consideran que el concepto de vigilancia consiste en un sistema panóptica (Rhodes, 1998: 308), entendido como un modelo de información comprimida que se utiliza en iPhones o iPads. Sin embargo, existen otras tecnologías electrónicas como los firewalls o ventanas, integradas en la arquitectura del software. Otras

formas de control en la geopolítica están orientadas al uso del teléfono inteligente o PNR (Passenger Name Record o, traducido del castellano, Registro del Nombre del Pasajero). En resumen, en estos momentos, surge la necesidad de comprender la seguridad en los negocios mediante el uso de técnicas digitales para analizar la geopolítica en un contexto actual en un contexto histórico. En este sentido, es posible explicar que la política consiste en reunir los intereses públicos y privados a nivel local, regional, estatal, supranacional o incluso a nivel del mundo globalizado. Asimismo, es importante entender que el control político es evidente en el ámbito de las instituciones, la administración o, incluso, los estados policiales. Por esa razón, hay un conjunto de redes consideradas como elementos de la fragmentación social. Por lo tanto, debemos expurgar, *in illo tempore*, las redes sociales, especialmente para vigilancia no controlada y, de esta manera, utilizarlas de forma más ordenada y actualizado con el propósito de evitar problemas ciudadanos como el movimiento Occupy (Chomsky, 2012).

La adiaforización (neutralizar actos morales haciendo que dejen de ser juzgados por el bien o el mal) existe en los datos personales como ocurre con el ADN o con las contraseñas (Flanagan, 2010: 93). En numerosas ocasiones, surge la necesidad de profundizar en la vida personal, y los datos o, incluso, en la forma de obtener conocimiento de los datos fragmentados que se duplican, lo que lleva a una mayor credibilidad que la que podría aportar una persona. De hecho, los diseñadores de software insisten en la idea de una modernidad líquida que consiste en el uso de criterios de búsqueda, según el contexto histórico-jurídico y tecnológico o, incluso, en los métodos de socialización control como se ve en el ejemplo orwelliano del Gran Hermano, que es un órgano instrumental que domina las ideologías y las mentes de los ciudadanos. También tenemos rastreo de teléfonos inteligentes con los códigos de barras, los chips, los códigos QR y un largo etcétera. Sin embargo, los chips localizan dónde están los productos mediante de la venta. En resumen, en la vida cotidiana, como ciudadanos, estamos controlados por los movimientos y la velocidad de las señales eléctricas. En cuanto al nivel de secretismo, debemos profundizar en el análisis nacional, la seguridad o, por el contrario, en la competencia comercial. En resumen, en el mundo de la sociedad líquida, dicha información personal existe porque una proporción de la de los datos se guardan en las esferas privada y pública (Bauman, 2000: 11). Los drones son sofisticados dispositivos de visualización que son casi tan imperceptibles como otros dispositivos gubernamentales. Cabe destacar su uso en enfrentamientos, para invisibilizar el conflicto armado. país o Estado-nación que lo inicia o lo lleva a cabo (Parker, 2010: 23). Los drones están programados para volar por sí solos para iniciar el espionaje o, en su defecto, para llevar a cabo la vigilancia geopolítica. Esto genera preocupación por la existencia de una vigilancia cada vez mayor. creciente tsunami de datos (Whitlock, 2011). Cabe destacar que, en geopolítica hay que reforzar el anonimato. No podemos sacrificar voluntariamente nuestro derecho como ciudadanos a la privacidad, por un precio razonable, como si fuéramos un rebaño de ovejas. En segundo lugar,

hay acceso a internet a través de páginas web con la inteligencia colectiva de dos mil millones de usuarios que utilizan su correo electrónico, mensajes, fotos o intercambio de datos. De hecho, renunciamos a nuestra autonomía personal a modo de un acto civil o contrato. Así, existe un consumo público que utiliza dispositivos más fáciles de manejar. Pero lo cierto es que hay un uso importante de drones u operadores de red, intentando influir en la realidad. Sin embargo, Internet no nos roba, en general, nuestra identidad; más bien refleja nuestra identidad porque, como consumidores, dejamos nuestros datos a merced de los datos. comerciantes, sin ningún control, y apenas nos preocupamos por nuestra privacidad. En resumen, debemos prestar atención a lo que Internet y las redes sociales revelan sobre relaciones sociales, ya que la violación de la privacidad debería importarnos. Es decir, la privacidad tiene convertirse en un lugar de confinamiento y el dueño de un espacio privado sentenciado, condenado y abandonado a su suerte que quiere ser comprendido por los consumidores de las redes sociales a cambio de compartir secretos (Saura, 2021b). Todo esto se debe al ego de los ciudadanos que brindan una enorme cantidad de información, a un precio ridículo, para presentadores de programas de entrevistas y televisión. o colaboradores de radio. De no ser así, podría resultar en muerte en vida o exclusión social, ya que uno no pertenece al club de la información selecta. El campo sociológico (Marx y Muschert, 2008) se relaciona principalmente con la digitalización y las redes sociales, ya que existen dos universos: online y offline, y en cuanto a los mundos, está el mundo virtual que es muy rápido y el mundo real, con un ritmo más lento, aludiendo stricto sensu a la fábula de la liebre y la tortuga. De hecho, las relaciones humanas avanzan según la ley del mínimo esfuerzo ya que el camino digital y virtual permite conectarse con otros sin apenas interacción con la gente de los diferentes grupos sociales. Sin embargo, para la tecnología, el mundo real es anacrónico y casi claustrofóbico porque evita, en un gran número de casos, las prácticas de socialización en la vida real. Huelga destacar los panópticos, que constituyen el elemento fundamental de la arquitectura tecnológica en el desarrollo, durante los siglos XX y XXI. Este mecanismo se basa en el cifrado de los mensajes de texto, WhatsApp, los correos electrónicos, etc., mientras que, en el caso de los drones con forma de libélulas mecánicas después del 11 de septiembre de 2001, que son más eficaces para perseguir los delitos tipificados como homicidios, en los diferentes códigos penales nacionales o internacionales. En el ámbito de la seguridad en cada país, los horrores y la tortura en Siria deben ser explicados y analizados. Esto lo demuestra los grupos terroristas del nacionalismo islámico. Asimismo, para comprender la seguridad, las enfermedades, los riesgos de la vida nocturna, padres que protegen a los niños o incluso la existencia de riesgos y las vulnerabilidades deben ser mitigadas. Para ello, es fundamental contar con elementos de vigilancia y técnicas de seguimiento que permitan a las personas en la esfera pública ejercer sus propias poder ejecutivo, legislativo y judicial discutible a través de la individualización, los recursos, y habilidades sociales. En resumen, el momento de la caída de las Torres Gemelas desató una obsesión desenfrenada que intentó desarrollar con la vigilancia

digital y era esencial hacerlo con la ayuda de la tecnología. Por otra parte, existe un temor creciente en las decisiones en un halo de secretos, mientras que, en el caso de Internet, la vigilancia es oculta. En resumen, la sofisticación de la tecnología de almacenamiento de datos permite obtener una respuesta razonada en el menor tiempo posible (Tester, 2004: 147). La civilización se encuentra actualmente en una encrucijada debido al uso de fuentes de energía obsoletas, como el petróleo o los combustibles fósiles; esto debe cambiarse radicalmente si no queremos destruir el planeta o queremos evolucionar en consecuencia como sociedad modernizada. Además, existen problemas graves como el cambio climático o incluso el comienzo de una sociedad o Era poscarbónica (Ayres, 2010: 11). En 1990, comenzó la Tercera Revolución Industrial, y en el Tercer Milenio se fomentó el uso y la regulación de las políticas digitales. En resumen, entender que la Tercera Revolución Industrial debe sentar las bases de la era colaborativa entre agentes sociales, políticos y económicos, de lo contrario la sociedad deriva en una dirección incierta. En la industria o en los mercados empresariales es posible colaborar utilizando nuevas herramientas como internet o las energías renovables para fortalecer el futuro de la sociedad a través de resultados científico-tecnológicos. Sin embargo, en cuanto a la revolución de las tecnologías y la comunicación, existen grandes empresas como Apple y AOL (Rifkin, 2011:191193) que ahondan en este propósito. De hecho, gracias a la digitalización, las emisiones de CO₂ han disminuido en nuestro planeta. Hay elementos o formas de cambio, tras la digitalización, en aspectos de comunicación, organización o incluso gestión, etc. Así, en el siglo XXI, los pilares fundamentales de la tercera Revolución Industrial son: ¡la transición de las energías renovables o la energía verde; la transformación del parque inmobiliario, el despliegue de tecnología lo que llevó al aumento de la Inteligencia Artificial o el almacenamiento de hidrógeno que comenzó a desarrollarse en los primeros años del Tercer Milenio. Entonces, ¿necesitamos gestionar, por ejemplo, ¿cuáles son las necesidades energéticas de una economía global compleja basada en energías blandas? Es importante ahondar en la idea de la existencia de vehículos motorizados electrónicos en la red de suministro. Sin embargo, la estabilización del régimen energético de los edificios fue fundamental, desde el primer momento de su creación como el aspecto de promover la sostenibilidad del planeta. En segundo lugar, cuentan con las plataformas tecnológicas de la Unión Europea, entre las que se encuentran las públicas y las iniciativas de investigación privadas, en relación con nuevas áreas como la promoción de transporte, electricidad o incluso tecnología (GWEC, 2009: 10). En resumen, la Tercera Revolución Industrial es un aspecto fundamental para consolidar las prácticas de la política internacional y nacional a través de la nueva mentalidad política entre las nuevas generaciones conectadas a las diferentes redes sociales e Internet. De ahí la internautica y los vínculos tienden a dividir el mundo, consolidando así un marcado giro en la economía internacional. la geopolítica. El desempeño político, en relación con la creación real, existe en la economía; la transformación y el cambio en los valores políticos o incluso cambios en la globalización como considera al continente como medio

de comunicación de la Tercera Revolución industrial y una visión o plan de acción por parte de la Unión Europea. En consecuencia, tras los inicios de la Tercera Revolución Industrial, la globalización está entrando en declive por lo que es vital reunificar los continentes, promover la cooperación de ayuda financiera o incluso promover el auge de las fuentes solares como los elementos de la transformación energética. En resumen, la era intercontinental debe adaptarse lentamente a nivel de relaciones internacionales, geopolítica y política de la biosfera (Rifkin, 2010: 584) y esto sólo se puede hacer con la ayuda de la tecnología y la IA. en relación con las universidades y el desarrollo de la tecnología docente aplicaciones o, incluso, servir para promover las energías renovables mediante el uso de la nanotecnología o la biotecnología.

1. LA CONVERGENCIA DE BIG DATA: APRENDIZAJE AUTOMÁTICO

La IA está siendo impulsada por la convergencia y la madurez industrial de tendencias tecnocientíficas. De hecho, el Big Data consiste en el procesamiento de enormes cantidades de los datos, así como el aprendizaje automático (la capacidad de las computadoras para conocer y comprender el campo del comercio) y el auge de los datos en la nube. Sin embargo, en el siglo XXI, la IA -como elemento de estudio- permite comprender la aceleración del aumento de la potencia informática, mientras que la reciente disponibilidad de los almacenes masivos y los flujos de datos digitales han permitido el despliegue de tecnologías o soluciones potentes basadas en el aprendizaje automático. Aunque uno podría verse tentado a pensar que la Inteligencia Artificial es una herramienta neutral, no está en un vacío carente de interacción social intereses (Miahe, 2018: 105). Acto seguido, con el Big Data, los sistemas son en realidad parte de un sistema sociotécnico complejo en el que participan los seres humanos, que han desempeñado y seguirán desempeñando -hasta que las máquinas dominen el mundo- un papel destacado en el mundo globalizado. Por lo tanto, no se trata realmente de la IA sino más bien de un trabajo colectivo inteligencia, que involucra cada vez más a comunidades globales, interdependientes y abiertas de actores con sus propias dinámicas de poder, mientras los equipos de ingeniería construyen grandes decorados de datos producidos por consumidores, vendedores, trabajadores, usuarios, ciudadanos, etc. En suma, hay algoritmos de diseño de macrodatos que interpretan algunos resultados y determinan la organización de nuestras sociedades a través del uso de teléfonos inteligentes y objetos, que son cada vez más interconectados con miles de millones de personas que utilizan Inteligencia Artificial (Gil, 2022) diariamente, y que contribuir a la formación y desarrollo de sus habilidades cognitivas y de sociabilidad. Los modelos de negocio específicos, a nivel industrial, son la base de la convergencia entre el Big Data, la potencia informática y el aprendizaje automático. Lo que distingue a los gigantes, como los Estados Unidos, es su inventiva y su contribución al desarrollo de la Investigación y la tecnología. En primer lugar, distinguimos los grupos que inicialmente

proponen la venta de productos y servicios de ventanas y aplicaciones telemáticas que recomiendan algunos servicios de venta online. En cuanto a los demás, se trata de explotar los datos de los usuarios para venderles diversos servicios comerciales, como ocurre con las empresas Facebook o Google (Hernández, 2018: 8). Por otro lado, el modelo económico se basa en la estrecha colaboración de sus actividades muy variadas, que están coherentemente vinculadas. Por ejemplo, el servicio de mensajería instantánea y los pagos como es el caso de algunas plataformas y aplicaciones a las que se les han facturado grandes sumas de dinero en China. Esto interfiere con la funcionalidad de varias aplicaciones. Sin embargo, algunos expertos en Big Data han denunciado los excesos de las principales plataformas. En casos en los que sus productos han beneficiado sin duda a usuarios de todo el mundo, estas empresas también participan en concursos de suma cero para llamar nuestra atención, que es algo que necesitan monetizar sus productos y servicios. Algunas empresas se ven constantemente obligadas a superar a sus competidores porque las diferentes plataformas dependen de los últimos avances en neurociencia para desplegar técnicas cada vez más persuasivas y adictivas. Su objetivo es mantener a los usuarios pegados a sus computadoras. Al hacer esto, influyen en nuestra percepción de la realidad, así como en nuestras decisiones o comportamientos. Por otro lado, la masificación de datos y las capacidades de procesamiento son necesarias para el aprendizaje automático, pero el ascenso al poder de la IA cuestiona casi inevitablemente una dinámica de mercado que es monopolística dentro de un área económica específica y oligopólica a nivel global (Rodríguez, 2018:16). De esta manera, somos testigos de una centralización del poder digital en manos de estos imperios informáticos. Tras esto, el desarrollo de la IA con las fuentes en la mano, corresponde a la dinámica de las economías de escala y alcance, así como a los efectos de las redes directas e indirectas, como las plataformas digitales que están en condiciones de recopilar y estructurar los datos de los consumidores, y atraer y financiar talentos capaces de dominar las funciones más avanzadas de la IA. Estos sistemas también proporcionan los talentos tienen acceso a potencia de procesamiento y a bases de usuarios lo suficientemente grandes como para desarrollar aún más sus capacidades de Inteligencia Artificial. En resumen, los algoritmos de aprendizaje automático utilizados en ciertas aplicaciones también son transferibles a otros en el marco de una disciplina llamada aprendizaje por transferencia, lo que favorece aún más a los actores estratégicos de juegos digitales. De hecho, Alpha Zero, un programa desarrollado por Google, se basa en el modelo Alpha Go, que se estructura a través de la dinámica del aprendizaje del proceso de transferencia. Hasta cierto punto, el predominio actual del mercado mundial surge naturalmente de los componentes necesarios para el desarrollo de la IA en el marco de la economía digital. Sin embargo, esta lógica, será inevitablemente modificado por el proceso continuo de convergencia entre hardware y software, lo que impulsa a los operadores a ofrecer soluciones diferentes.

2. LA CUESTIÓN DE LA SOBERANÍA DE LOS GOBIERNOS

Los actores nacionales e internacionales son cada vez más conscientes de las implicaciones estratégicas, los riesgos económicos y militares relacionados con las prácticas de inteligencia artificial. También anticipar su impacto en las elecciones, como lo demuestra la interferencia en las elecciones de 2016 y las elecciones presidenciales de Estados Unidos de 2020 o en el referéndum sobre el Brexit. Por lo tanto, Canadá, China, Corea del Sur, Dinamarca, la Comisión Europea, Finlandia, México, Japón, Italia, India, países bálticos y escandinavos, Emiratos Árabes Unidos, Taiwán, Reino Unido y Singapur, que en el último año y medio han presentado Estrategias para potenciar el uso y desarrollo de la IA. Estas estrategias consisten en abordar la educación, la investigación y el aprendizaje de diversas maneras con el desarrollo digital en las infraestructuras, los servicios públicos y la ética social. No todos los países pueden aspirar a liderazgo en este ámbito, sino que se trata, más bien, de identificar y desarrollar la capacidad comparativa ventajas y satisfacer las necesidades específicas de la nación. Algunos estados se centran en la ciencia o la investigación, otros sobre el cultivo del talento y la educación (Bruffee: 1999: 66); otros sobre la adopción de la IA en la administración o la ética y la inclusión. Los actores entre los países caracterizados por su aspiración al liderazgo global se encuentran no sólo Estados Unidos y China, sino también la Unión Europea. Sin embargo, países como Canadá, Francia y el Reino Unido también tiene una estrategia ambiciosa y está realizando inversiones significativas y avances en IA. Otros países se están especializando en aspectos concretos. Pero, siguiendo una estrategia de nicho de mercado exitosa, India, por ejemplo, aspira a convertirse en un lugar para el fomento y promoción de la IA especializado en aplicaciones específicas para países en desarrollo. Por otro lado, Polonia está explorando y desarrollar aspectos relacionados con la ciberseguridad y los usos militares, mientras que el gobierno de los Emiratos Árabes Unidos lanzó su estrategia digitalizada en el otoño de 2017, creándose de esta manera el primer ministerio de inteligencia artificial del mundo.

3. EL DUOPOLIO CHINO-ESTADOUNIDENSE: ESTADOS UNIDOS Y CHINA

Estados Unidos y China, constituyen actualmente una especie de Inteligencia Artificial duopolio basado en las dimensiones críticas de sus mercados y sus políticas de *laissez faire* (dejar hacer). En materia de protección de datos personales, su competencia es especialmente evidente en una guerra comercial que comenzó con la imposición por parte de la administración del presidente de Estados Unidos, Donald Trump, al aplicar un arancel del 25% a los productos chinos, lo que incluidos productos derivados de la Inteligencia Artificial, con un valor aproximado de Treinta y cuatro mil millones de dólares. China tomó represalias imponiendo barreras fiscales del 25% a cinco ciento cuarenta productos estadounidenses. Poco después, en agosto de 2017, Washington... inició una investigación

contra China (Chesney, 2019: 147), acusándola de prácticas injustas prácticas administrativas, financieras y comerciales en materia de propiedad intelectual propiedad, particularmente en el ámbito tecnológico. Por otro lado, el auge chino comenzó en marzo de 2016, cuando la empresa surcoreana Go, Lee Sewol, sufrió graves pérdidas. Derrota por el programa de Inteligencia Artificial de DeepMind o Alpha Go (Corella, 2017). En julio En 2017, la capital china publicó una estrategia nacional especialmente ambiciosa destinada a conseguir abastecer un mercado estimado de 15,7 billones de dólares para 2030. Todavía pretende convertirse en el centro privilegiado de la innovación en IA, dentro de nueve años. De acuerdo con esta política, que está dirigida específicamente a los niveles más altos del Estado. El ecosistema tecnológico chino seguirá una estrategia de la inversión ofensiva y refuerzo de las capacidades de Inteligencia Artificial. Por otro lado, por ejemplo, una de las empresas más solventes, como es el caso de Alibaba, contribuyó con quince mil millones de dólares a la investigación y desarrollo (I+D), una cifra comparable a la de los gigantes estadounidenses como como Amazon, que invirtió 16.100 millones de dólares en I+D en 2017. A raíz de esto, la IA avanzó rápidamente mientras los equipos chinos, por ejemplo, ganaron el concurso de reconocimiento visual ImageNet en 2016 ya también al año siguiente. Por otro lado, China ha integrado la inteligencia digital y artificial en su estrategia geopolítica. Sin embargo, desde 2016, su iniciativa titulada, Cinturón y Ruta Digital, que lleva a cabo la construcción de las infraestructuras para conectar Asia, África y Europa, han incluido un componente digital como apunta dicho programa. El último avance con respecto al programa fue la creación de un nuevo centro internacional de excelencia para las rutas de la seda digitales, en el caso de Tailandia, del invierno de 2018.

4. ¿GEOPOLÍTICA DE BLOQUES? LA SITUACIÓN DIGITAL DE LA UNIÓN EUROPEA

La Unión Europea se ha quedado atrás de las políticas de China y Estados Unidos en términos tecnológicos e industriales. En 2013, un informe emitido por el Senado francés expresó su preocupación al observar que el Viejo Continente se está convirtiendo en una especie de colonia digital. La situación no ha mejorado en los últimos años y tiene problemas reales en 2025. El enfoque europeo parece consiste en aprovechar su mercado de consumidores multimillonario para sentar las bases de un modelo industrial ético de la IA, al tiempo que se renegocian las relaciones estratégicas, de facto, asociadas con Estados Unidos. Así, en abril de 2018, cuando la Comisión de la Unión Europea publicó su programa de IA y acto seguido el ministro sueco de Digital Development apuntó que no se esperaba que China hiciera lo mismo. De hecho, con una democracia y un sistema jurídico que funcionen, Europa debe considerar que la i es fundamental para el futuro. Competir con China con Estados Unidos es, por supuesto, importante, pero si no creamos un marco legal y ético necesario, como continuamente estamos condenados al fracaso (Luveto; 2008: 5). Europa se queda atrás

en inversión privada en IA que ascendió aproximadamente a entre 2.400 y 3.200 millones de euros en 2016, y entre 5.000 y 7.000 millones en 2020, según en comparación con tres veces más millones en el continente asiático y hasta cinco veces más mucho en Estados Unidos, aunque, a pesar del tamaño de la convergencia de los llamados Big Data, impulsado por computadoras y aprendizaje. Para superar a sus competidores, las diferentes plataformas dependen de los últimos avances en neurociencia, que utiliza técnicas cada vez más persuasivas y adictivas para mantener usuarios pegados a sus pantallas. Al hacerlo, influyen en nuestra percepción de la realidad, nuestras elecciones y comportamientos en una forma poderosa y hasta ahora completamente desregulada en el software. De hecho, el desarrollo de la IA y su uso en todo el mundo es, por tanto, un tipo de dominación que, por medios no coercitivos, influye en el comportamiento de los actores o incluso define sus propios intereses. En este sentido, se puede hablar de un proyecto de imperios digitales (Attina, 2001: 190). Junto con las masas críticas de datos y procesamiento capacidades necesarias para el aprendizaje automático, el ascenso al poder de la inteligencia artificial. Casi inevitablemente pone en marcha una dinámica de mercado que es monopolística dentro de un determinado y un área económica oligopólica a nivel global. Por lo tanto, estamos presenciando una centralización del poder digital en manos de estos imperios digitales. El desarrollo de la tecnología sofisticada corresponde a la dinámica de las economías de escala y alcance, así como los efectos de las redes directas e indirectas: las grandes plataformas digitales pueden recopilar y estructurar más datos sobre los consumidores y atraer y financiar a los talentos que son capaces de dominar las funciones más avanzadas de la IA. De hecho, estas plataformas también pueden proporcionar a estos talentos el acceso necesario al procesamiento, conocidas como las bases de poder y de usuarios que sean lo suficientemente grandes como para seguir desarrollando sus capacidades de inteligencia artificial. Los algoritmos de aprendizaje automático utilizados en ciertas aplicaciones también son transferibles y han beneficiado aún más a los bloques políticos. En concreto, Alpha Go aprendió con videojuegos, jugando contra sí mismo, sin intervención humana. Expertos y gracias a un estricto entrenamiento logró vencer a los mejores ordenadores en ajedrez. Sin embargo, hasta cierto punto, el dominio actual del mercado mundial surge de los componentes necesarios para el desarrollo de la Inteligencia Artificial en el ámbito digital economía. En suma, esta lógica se verá naturalmente acentuada por el proceso continuo de convergencia entre hardware y software, lo que está empujando a los operadores a recurrir a desarrollando sus propias soluciones y componentes críticos ya que, la centralización del poder digital en torno a unos pocos actores es lo que caracteriza a un nuevo tipo de imperio digital. Como se ha dicho anteriormente, la Unión Europea se ha quedado muy por detrás de China y Estados Unidos, desde las perspectivas comercial, tecnológica y/o digital, provocando -durante más de una década- una gran preocupación por parte de la Unión Europea. El continente será controlado digitalmente por Estados Unidos o China y Rusia dentro de unos años (Koerner, 2020: 1). En otro orden de cosas, el enfoque europeo

parece consistir en aprovechar su mercado de 500 millones de consumidores para sentar las bases de un modelo industrial ético de la IA al tiempo que renegocia una asociación estratégica de facto con Estados Unidos. A pesar del tamaño de su mercado, que aún no se ha integrado digitalmente, con su reconocida excelencia científica y su vigor emprendedor y creativo. Hay pocos modelos europeos, por lo que su crecimiento se ve frenado por la falta de una plataforma digital. por el enfoque industrial y capitalista del Viejo Continente. Con. Estos mimbres, es difícil -bajo estas condiciones- evitar la compra de datos digitales europeos por parte de gigantes británicos, estadounidenses y chinos, como fue el caso de DeepMind, pionera en el campo de la IA que fue comprada por Google en 2014 por 500 millones de dólares, o en el caso de la compra de Cuka, una de las principales corporaciones alemanas de robótica, por parte de El gigante chino de electrodomésticos Midea que, en 2016, pagó alrededor de cuatro mil millones de dólares por la filial alemana. Como señaló el cofundador de Darty en un dictamen reciente Artículo, la revolución de la inteligencia artificial se percibe en la Europa tecno-fóbica, hasta cierto punto. mayor grado que en China o Estados Unidos debido a una ola extranjera que amenaza su modelo socioeconómico y de la cual debe protegerse. Por lo tanto, impulsados por actores industriales dominantes, que no han logrado llevar a cabo sus para su propia transformación digital, los europeos buscan regular y reducir la IA. La revolución de la IA hay que orientarla hacia ciertos propósitos. Además, hay una ambición universalista, que es más conservadora que progresista. El resultado es la búsqueda de un modelo europeo de la IA que vincule la afirmación de la soberanía y la búsqueda del poder con respeto a la dignidad humana. Sin embargo, equilibrar estos tres objetivos no será fácil, porque regular desde una posición de extrema debilidad y dependencia industrial en relación con los estadounidenses. Es probable que Europa, China, bloquee su propio ascenso a la dominación global. Este es el riesgo que, el Reglamento General de Protección de Datos probablemente suponga un desafío para las ambiciones de Europa en materia de protección de datos. Posteriormente, la Comisión Europea anunció un aumento de 1.500 millones de euros en sus inversiones en investigación e innovación en IA para 2018-2020 (aunque todavía no hay una financiación sostenible), en el marco del programa Horizonte 2020. La IA permite influir en los votantes en los países democráticos, Estados autoritarios o, también, puede reforzar el control de la población. Tras esto, las plataformas deben integrar estas preocupaciones éticas y políticas en su estrategia. En resumen, la inteligencia artificial, como cualquier revolución tecnológica, ofrece un potencial considerable oportunidades, pero también presenta una serie de riesgos económicos, sociales y políticos que se superponen exponencialmente con estos (Bale, 2019: 47).

5. EPÍLOGO: ¿HACIA DÓNDE NOS DIRIGIMOS?

Finalmente, es necesario plantear varias cuestiones, entre las que cabe mencionar las siguientes: Destacado: ¿Cómo debemos actuar y sobrevivir en la jungla de la

geopolítica? ¿Es artificial? ¿Es la inteligencia esencial para comprender la esfera pública? Para resolver estas preguntas, sería necesario conocer las intenciones y capacidades del nuevo integrante digital en geopolítica internacional. La IA se ha convertido en un elemento importante poder y muy competitivo. Esta es la realidad que enfrentan los Estados miembros de la Unión Europea y debe enfrentar. Por otro lado, la IA está preparada para influir en el mundo equilibrio de poder y la relación entre quince estados, así como la geopolítica, en general. En nuestro ámbito, la Unión Europea debe tomar en serio este desafío y comprometerse con los cambios que se están produciendo en el mundo. Por eso, es hora de que Europa invirtiera más tiempo, esfuerzo y dinero para garantizar que el continente europeo pueda beneficiarse de los desafíos internacionales que la Inteligencia Artificial está introduciendo. Para que la IA sea adoptada en cualquiera de estos tres conjuntos de funciones, debe demostrar la eficacia comparable o mayor que la de los humanos a un coste comparable. Por lo tanto, en la situación en la que los legisladores confían en la capacidad de los analistas humanos para llegar a un acuerdo. En lo que respecta al control de armas estratégicamente importantes, es poco probable que entren en vigor. De hecho, es relevante transferir dichos procesos a un sistema completamente nuevo y no probado. En cualquier caso, la implementación no será simplemente una cuestión de entregar las llaves o, incluso, de accionar un interruptor. Parece que no habrá agentes IA impuestos a asumir el reto roles humanos, sino más bien que la Inteligencia Artificial se combinará cada vez más con los humanos. Analistas para participar en actividades y proyectos. De esta manera, los agentes hombre-máquina ofrecen la mayor promesa en el futuro previsible y representan la máxima oportunidad para una transición gradual y alejada de la supervisión humana exclusiva en lo que respecta estas tareas. Teóricamente, los emparejamientos permiten combinar las mejores cualidades del ser humano. y la inteligencia de la máquina: la máquina puede procesar grandes cantidades de datos rápidamente mientras el ser humano puede detectar y corregir cuando sea necesario, así como comprender, enmarcar y responder a los resultados de maneras que interactúen con los mecanismos de políticas existentes. Algunos ya han hecho esfuerzos para comenzar a diseñar un marco ético o legal para sistemas autónomos. Más recientemente, en mayo de 2018, una organización no gubernamental, firmó una nueva declaración sobre la protección de los derechos, la igualdad y la no discriminación en los sistemas de aprendizaje automático. Por otro lado, en el marco del derecho internacional de los derechos humanos y haciendo hincapié en las responsabilidades de los derechos humanos actores de los sectores público y privado, como se describe en la Declaración de Toronto, que tiene como objetivo garantizar que las nuevas tecnologías de aprendizaje automático y la IA y los sistemas de datos relacionados, de manera más amplia, incorporan principios de respeto por la inclusión y no discriminación. Sin embargo, el tiempo de reacción, en el contexto humano, podría no ser adecuados para su acoplamiento o necesidad de mayor autonomía. Incluso si un cambio fundamental. En la guerra impulsada por la autonomía no es una perspectiva inmediata, sino que e produce por los cambios en las normas y estándares sobre cómo los responsables de las políticas ven y responden a las amenazas son probablemente sean radicales. De hecho, se previó que ciertas armas definirían la legalidad del desarrollo y uso de armas autónomas. Estos pasos son

bienvenidos, pero están en las etapas iniciales, pues todavía estamos en proceso de desarrollo de estas nuevas tecnologías. Sin embargo, este estudio no pretende cubrir todo el alcance o amplitud de las implicaciones de las tecnologías de la IA para la política, sino señalar, que hay áreas importantes, incluida la salud pública o el derecho, a través de la IA. Los sistemas de inteligencia pueden ser transformadores en formas que impactan directamente los procesos del sistema internacional en los próximos veinte años. En resumen, es necesario centrarse en la aplicación de la IA desde la seguridad militar, económica y ciudadana perspectivas, en el ámbito de la privacidad (Saura y Palacios, 2021), para evitar la caída de la Unión Europea en manos de los grandes imperios digitales.

6. CONCLUSIONES

El estudio es original porque profundiza en los mecanismos de la robótica, como el objetivo final de la Inteligencia Artificial, según lo explica un grupo de expertos. Sin embargo, a pesar de las afirmaciones post-canónicas, la IA no logra una mejora, debido a la secuencia de patrones determinados, en la vida de los seres humanos sino, más bien, lo que refleja y se explica en este capítulo es que las mentes brillantes de la disciplina, que han fomentado este campo de estudio lo han orientado principalmente a la economía campo y, por lo tanto, no se considera sentimientos humanos, en términos de robótica y artificial. Por otra parte, la Unión Europea debe ser competitiva (aumentando su financiación en Inteligencia Artificial) desarrollando un conjunto de técnicas, así como nuevas formas de trabajar, en la materia antes mencionada, para evitar ser absorbidos por los mercados chino y estadounidense. En resumen, para comprender que la IA tal como está funcionando, es una herramienta para unir los sentimientos híbridos entre seres humanos y las máquinas.

BIBLIOGRAFÍA

- Ayres, RU Ayres, Cruzando la brecha energética: pasando de los combustibles fósiles, Dependencia para un futuro de energía limpia, Upper Saddle River, (Nueva Jersey), Wharton Editorial, 2010.
- Azcona Pastor, José Manuel y Gil González, Fernando, “Geopolitics, Governance and Artificial Intelligence”, en Saura, José Ramón y Debasa, Felipe (Coords.), Artificial Intelligence in Government practices and process, Pensilvania, IGI Global, 2022, pp. 1-12.
- Bale N. “Inteligencia artificial: evaluación de riesgos y consideraciones para el futuro”, Revista Internacional de Aplicaciones Informáticas, 181 (43): pp.47-49 (marzo de 2019).
- Baños P. “Así se domina el mundo”, Ariel, 2017, Barcelona, págs. 468. Bárcena A. et al. “La nueva revolución digital desde la Internet del consumidor hasta la Internet industrial”, CEPAL y Naciones Unidas, 2016, Santiago, pp. 100.
- Bauman, Z. Modernidad líquida, Cambridge, 2000, Polity.
- Bruffee, KA Aprendizaje colaborativo: educación superior, interdependencia y la Autoridad del conocimiento, Baltimore, Baltimore University Press, 1999. Carton, S. “Por qué el ascenso de Donald Trump debería hacernos dudar del bombo publicitario sobre ¿Inteligencia artificial? 2016, págs. 13.
- Chomsky, N. Ocupar Wall Street, Ediciones Urano, 2012, Madrid.
- Corbella, J. Inteligencia Artificial Arpa Jo, Diario La Vanguardia, octubre del 2017 Deleuz G. “Posdatas de las sociedades de control”, octubre, 59, (invierno de 1992 Flanagan, K. Teología implícita de Bauman, 2010, Palgrave, pp. 92106.
- Informe Global Wind 2008 (Bélgica), 2009.
- Hernández García, C. “Big data o como los datos masivos están cambiando el mundo”, Ciudad de México”, Dirección General de Divulgación de la Ciencia, UNAM, n° 241, pp.8-13.
- Janssen, M., Brous, P., Estevez, E., Barbosa, LS y Janowski, T. (2020). Datos Gobernanza: Organización de datos para una Inteligencia Artificial confiable. Gobierno Información trimestral, 37(3), 101-493.
- Koerner, K. “¿Cómo se convertirá la UE en una super estrella de la IA?”, Economía digital y cambio estructural, Deutsche Bank, 18 de marzo de 2020.
- Lu, Lucy y Etzkowitz, H. “Desafíos estratégicos para la creación de conocimiento basado en Innovación en China: Transformando la triple hélice universidad gobierno industria relaciones”, Revista de Gestión Tecnológica en China, Vol. 3. N° 1, pp. 511.
- Lyon, D. “La frontera está en todas partes. Documentos de identidad, vigilancia y demás “en E Zureik, y MB Salter, Vigilancia y Policía Global, Collumpton, Willan, 2005, pp. 66-82.

- Mesa Escobar, E. “Esfera pública: entre lo público y la política en la construcción de la opinión política”, Foro. Revista Departamento de Ciencia Política, n° 5 / Enero junio de 2014, pp. 105-117.
- Mialhe, N. Hodes, CR; Buse, C., Lannquist, y Jeanmaire, C. “Geopolítica de la Inteligencia artificial. Política exterior, vol. 34, N° 193, 2020, págs. 56-69.
- Mialhe, N. “La geopolítica de la inteligencia artificial: ¿El regreso de los imperios? Política Étrangère, 3, 2018, pp. 107-118.
- Morales, E. (Coord.), México; Instituto de Investigaciones Bibliotecológicas y de la Información de la UNAM, 2018, pp. 271.
- Parker K. “¿Puede sobrevivir la ciudad en la colina?”, Washington Post, 5 de diciembre p. 23.
- Rhodes, L. “Intimidades panópticas”, Public Culture, 10, 2, 1998, p. 308.
- RibeiroNavarrete, S., Saura, JR, & PalaciosMarqués, D. (2021). Hacia una nueva era de recopilación masiva de datos: evaluación de las tecnologías de vigilancia de pandemias para preservar la seguridad de los usuarios Privacidad. Pronóstico tecnológico y cambio social, 167, 120681. <https://doi.org/10.1016/j.techfore.2021.120681>
- Rodríguez Manzano, A. “El uso de los datos masivos para salvar vidas Dirección General de Divulgación de la Ciencia, UNAM, n° 241, pp. 16-19.
- Saura, JR, RibeiroSoriano, D., & PalaciosMarqués, D. (2021b), “Configuración de privacidad por defecto en la IoT social: Teorizando los desafíos y direcciones en la investigación de Big Data”, Big Data Research, 25, 15 de julio de 2021
- Tester, K. El pensamiento social de Zygmunt Bauman, Londres, Palgrave 2004.
- VVAA. Transiciones de la guerra a la paz. Presentes y futuros de Irán, Politique Étrangère, 3, 2018, IFRI, pp. 216.
- Whitlock, C. “El sistema de vigilancia Gorgon Stare observa la zona de guerra afgana”, The Washington Post, 29 de abril de 2011.



CAPÍTULO 10.
INTELIGENCIA ARTIFICIAL, ROBOTIZACIÓN DE LA JUSTICIA: ESPECIAL REFERENCIA A COMO SE VEN AFECTADOS LOS DERECHOS HUMANOS Y LAS GARANTÍAS PROCESALES³⁸³

MARÍA CONCEPCIÓN RAYÓN BALLESTEROS³⁸⁴

Profesora Permanente Laboral de Derecho Procesal (acreditada para Profesor Titular)

Universidad Complutense de Madrid
mcrayon@ucm.es

Resumen

El desarrollo y despliegue de sistemas de inteligencia artificial y herramientas de robotización en la Administración de Justicia plantea oportunidades relevantes de eficiencia, celeridad y racionalización de recursos, pero también desafíos de primer orden para la protección de los derechos humanos y las garantías procesales. El artículo analiza, desde una perspectiva jurídico-dogmática y a la luz de los principales estándares internacionales y europeos, el impacto de la automatización de la toma de decisiones judiciales y de los sistemas de apoyo algorítmico al juzgador. En primer lugar, se examinan los conceptos de inteligencia artificial aplicada a la justicia y de “justicia robotizada”, así como sus principales manifestaciones prácticas (sistemas de reparto automatizado, herramientas de predicción de riesgo, generadores de borradores de resoluciones, asistentes virtuales procesales, etc.). En segundo lugar, se estudia cómo estas tecnologías inciden en derechos fundamentales como el derecho al juez imparcial, el derecho a un proceso con todas las garantías, la tutela judicial efectiva, el derecho de defensa, el principio de igualdad de armas y la protección de datos personales. Se presta especial atención a los riesgos de opacidad algorítmica, sesgos discriminatorios, falta de control humano significativo y despersonalización de la función jurisdiccional. Finalmente, se proponen criterios y salvaguardas normativas y prácticas para asegurar que la

³⁸³ El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

³⁸⁴ Profesora de Derecho Procesal (acreditada para Prof. Titular) Facultad de Derecho de la Universidad Complutense de Madrid, Abogada del Ilustre Colegio de la Abogacía de Madrid, Mediadora inscrita en el Ministerio de Justicia, Doctora en Derecho y Licenciada en Ciencias Políticas ORCID <https://orcid.org/0000-0002-7362-2885>

incorporación de la inteligencia artificial en la justicia sea compatible con un modelo de Estado de Derecho y con la centralidad de la persona, reforzando la transparencia, la explicabilidad, la rendición de cuentas y el control jurisdiccional sobre las decisiones automatizadas o asistidas por algoritmos.

Palabras clave

Inteligencia artificial, robotización de la justicia, Derechos humanos, garantías procesales, sesgo algorítmico, transparencia, explicabilidad.

Abstract

The development and deployment of artificial intelligence systems and justice robotization tools offer significant opportunities in terms of efficiency, speed, and resource optimization, but they also raise major challenges for the protection of human rights and procedural safeguards. This article provides a legal-dogmatic analysis, in light of the main international and European standards, of the impact of automated judicial decision-making and algorithmic decision-support systems for judges. First, it explores the concepts of artificial intelligence in the justice system and “robotized justice”, as well as their main practical manifestations (automated case allocation systems, risk-assessment tools, draft-judgment generators, and virtual procedural assistants, among others). Second, it examines how these technologies affect fundamental rights such as the right to an impartial tribunal, the right to a fair trial, effective judicial protection, the rights of the defence, the principle of equality of arms, and data protection. Special attention is paid to the risks posed by algorithmic opacity, discriminatory bias, lack of meaningful human control, and the depersonalization of the judicial function. Finally, the article proposes normative and practical safeguards to ensure that the integration of artificial intelligence into the justice system remains compatible with the rule of law and the centrality of the individual, by strengthening transparency, explainability, accountability, and judicial oversight over automated or algorithmically assisted decisions.

Key words

Artificial intelligence, robotized justice, Human Rights, procedural safeguards, algorithmic bias, transparency and explainability.

Sumario: INTRODUCCIÓN 1. CONCEPTOS FUNDAMENTALES: LA JUDICIAL Y ROBOTIZACIÓN DE LA JUSTICIA 1.1. *Inteligencia Artificial Judicial: definición y tipología* 1.2. *Diferencias conceptuales: LA vs. Robotización* 2. MARCO NORMATIVO INTERNACIONAL Y EUROPEO 2.1. *La Carta de Derechos Fundamentales de la Unión Europea y el Convenio Europeo de Derechos Humanos* 2.2. *Reglamento General de Protección de Datos (RGPD)* 2.3. *Reglamento (UE) 2024/1689 sobre Inteligencia Artificial (Ley de LA)* 2.4. *Recomendaciones del Consejo de Europa y la UNESCO* 2.5. *Derecho comparado: experiencias internacionales* 2.6. *Necesidad de una regulación procesal específica* 3. PRINCIPALES DERECHOS HUMANOS AFECTADOS 3.1. *Dignidad humana* 3.2. *Derecho a un juicio justo, con todas las garantías y al derecho de defensa* 3.3. *Igualdad y no discriminación* 3.4. *Derecho a la*

privacidad y protección de datos personales 3.5. *Libertad de información* 3.6. *Tutela judicial efectiva y acceso a los tribunales* 4. RIESGOS JURÍDICOS Y ÉTICOS 4.1. *Sesgos algorítmicos* 4.2. *Opacidad y el problema de la “caja negra”* 4.3. *Pérdida de supervisión humana y automatización de la autoridad* 4.4. *Ciberseguridad* 4.5. *Deshumanización de la justicia* 5. PROPUESTAS Y BUENAS PRÁCTICAS 5.1. *Principio “Human in the Loop”: supervisión humana significativa* 5.2. *Auditorías algorítmicas obligatorias y certificación* 5.3. *Explicabilidad y transparencia algorítmica* 5.4. *Regulación procesal* 5.5. *Alfabetización para operadores jurídicos* 5.6. *Principios de diseño ético: equidad, inclusión y accesibilidad* 5.7. *Estrategia de ciberseguridad integral* 5.8. *Gobernanza algorítmica y supervisión institucional con creación de entornos de prueba o sandbox* 6. CONCLUSIONES. BIBLIOGRAFÍA

INTRODUCCIÓN

La incorporación de tecnologías de inteligencia artificial (IA) en los sistemas de justicia ha dejado de ser una cuestión de ciencia ficción para convertirse en un fenómeno jurídico y social de alcance global. En la actualidad, los Estados exploran soluciones algorítmicas y sistemas robotizados que van desde el apoyo en la gestión de expedientes hasta la propuesta de resoluciones judiciales preliminares, con el objetivo declarado de optimizar recursos, reducir la sobrecarga de los tribunales y mejorar el acceso a la justicia³⁸⁵. Esta tendencia responde a una realidad innegable: los sistemas judiciales enfrentan un aumento exponencial en el volumen de litigios y un déficit de medios humanos y materiales que ralentiza los procedimientos, afectando el derecho fundamental a obtener una tutela judicial efectiva en un plazo razonable³⁸⁶.

Sin embargo, la adopción de herramientas de IA en el ámbito judicial no está exenta de riesgos. Si bien la tecnología tiene el potencial de incrementar la eficiencia y la objetividad, su uso plantea desafíos éticos y jurídicos relacionados con la protección de los derechos humanos. El núcleo del debate reside en determinar hasta qué punto puede automatizarse el proceso judicial sin sacrificar garantías esenciales como el derecho a un juez imparcial, la motivación de las resoluciones, la protección de datos personales y la dignidad humana. Estos principios, consagrados en instrumentos normativos como la Carta de Derechos Fundamentales de la Unión Europea (CDFUE) y el Convenio Europeo de Derechos Humanos (CEDH), constituyen el marco de referencia indispensable para cualquier reflexión sobre el futuro de la justicia en el siglo XXI³⁸⁷.

³⁸⁵ Susskind, R. (2019). *Online Courts and the Future of Justice*. Oxford University Press.

³⁸⁶ European Commission for the Efficiency of Justice (CEPEJ). (2020). *European judicial systems: Efficiency and quality of justice*. Council of Europe Publishing.

³⁸⁷ Carta de Derechos Fundamentales de la Unión Europea, DOUE C 326, 26.10.2012, p. 391-407; Convenio Europeo para la Protección de los Derechos Humanos y de las Libertades Fundamentales, Roma, 1950.

El impacto de la IA en el ámbito judicial no es homogéneo. Mientras algunos países experimentan con sistemas avanzados de justicia predictiva —como los algoritmos de evaluación de riesgo penal utilizados en Estados Unidos—, otros priorizan soluciones orientadas a la simplificación procesal, como los tribunales virtuales o plataformas de mediación automatizada. En Europa, el enfoque es más cauteloso: el Reglamento (UE) 2024/1689 sobre Inteligencia Artificial, conocido como Ley de IA, establece un marco de regulación basado en el riesgo que somete los sistemas aplicados en justicia a la categoría más estricta de “alto riesgo”³⁸⁸. Este enfoque refleja una visión garantista propia del modelo europeo, donde el desarrollo tecnológico debe subordinarse al respeto del Estado de Derecho³⁸⁹.

El presente artículo se propone analizar, desde una perspectiva jurídico-doctrinal y crítica, los retos y oportunidades que plantea la introducción de IA y sistemas robotizados en la administración de justicia, con especial énfasis en su compatibilidad con los derechos humanos fundamentales. Para ello, se estudiarán: los conceptos fundamentales de IA judicial y robotización procesal; el marco normativo internacional y europeo; los derechos humanos más directamente implicados; los riesgos jurídicos y éticos asociados; y las propuestas de buenas prácticas y regulación sectorial.

El análisis pretende contribuir al debate académico y legislativo con una visión integradora, que combine el rigor técnico-jurídico con una perspectiva crítica.

1. CONCEPTOS FUNDAMENTALES: IA JUDICIAL Y ROBOTIZACIÓN DE LA JUSTICIA

La comprensión de los términos “inteligencia artificial judicial” y “robotización de la justicia” es esencial para delimitar el alcance del debate jurídico sobre su impacto en los derechos humanos. Ambos conceptos se sitúan en el cruce entre tecnología, derecho y ética, y su correcta interpretación es clave para evaluar los riesgos y beneficios de su aplicación en los sistemas judiciales.

1. 1. Inteligencia Artificial Judicial: definición y tipología

El término inteligencia artificial judicial hace referencia al uso de sistemas computacionales dotados de capacidades de aprendizaje automático (*machine learning*), procesamiento de lenguaje natural y análisis predictivo, destinados a asistir, complementar o reemplazar determinadas funciones humanas en la administración de justicia. Estos sistemas no se limitan a tareas mecánicas de búsqueda documental; algunos son capaces de generar valoraciones probabilísticas sobre el desenlace de un litigio,

³⁸⁸ Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, sobre Inteligencia Artificial y por el que se modifican determinados actos legislativos de la Unión.

³⁸⁹ Floridi, L., & Cowls, J. (2021). *A Unified Framework of Five Principles for AI in Society*. *Harvard Data Science Review*, 3(1).

identificar patrones en la jurisprudencia o incluso recomendar resoluciones en casos concretos.

Generalmente se distinguen dos categorías principales:

- IA de apoyo: son sistemas orientados a facilitar el trabajo judicial sin desplazar el control decisorio. Ejemplos incluyen bases de datos jurisprudenciales avanzadas, herramientas de análisis de sentencias mediante algoritmos semánticos o asistentes virtuales para abogados y jueces³⁹⁰.
- IA de sustitución: se refiere a sistemas que asumen un rol decisorio o cuasidecisorio, como aquellos que determinan evaluaciones de riesgo de reincidencia o fijan cuantías de indemnización de forma automatizada. Su uso plantea mayores tensiones jurídicas, ya que la intervención humana puede quedar reducida a una mera formalidad³⁹¹.

El desarrollo de estos sistemas se ha visto impulsado por el auge del *big data* jurídico y el incremento de litigiosidad. En países como Estados Unidos, el algoritmo COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) ha sido ampliamente criticado por sesgos raciales, tras evidenciarse que arrojaba mayores índices de reincidencia para acusados afroamericanos. En contraste, algunos países europeos han optado por enfoques experimentales más prudentes, priorizando proyectos de digitalización documental y análisis estadístico antes que sistemas de predicción judicial³⁹².

El concepto de robotización de la justicia engloba no solo el uso de inteligencia artificial, sino también la automatización de procesos en los tribunales. Esta robotización comprende desde sistemas de asignación automática de casos y notificaciones electrónicas hasta el uso de algoritmos para elaborar borradores de sentencias o decisiones de menor cuantía³⁹³. El Consejo de Europa, a través de su Comisión Europea para la Eficiencia de la Justicia (CEPEJ), ha destacado que estos avances pueden agilizar procesos, pero solo si se aplican respetando principios de calidad, transparencia y control humano³⁹⁴.

La robotización judicial responde a la necesidad de gestionar grandes volúmenes de casos y reducir costos, pero plantea interrogantes sobre la calidad de la justicia, la

³⁹⁰ Ashley, K. D. (2017). *Artificial Intelligence and Legal Analytics: New Tools for Law Practice in the Digital Age*. Cambridge University Press.

³⁹¹ Dwork, C., & Mulligan, D. K. (2013). *It's Not Privacy, and It's Not Fair*. *Stanford Law Review Online*, 66, 35–40.

³⁹² European Commission for the Efficiency of Justice (CEPEJ). (2020). *European judicial systems: Efficiency and quality of justice*. Council of Europe Publishing

³⁹³ Susskind, R. (2019). *Online Courts and the Future of Justice*. Oxford University Press.

³⁹⁴ CEPEJ. (2018). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*. Council of Europe.

comprensión pública de las decisiones y el rol del juez como garante último de los derechos.

1.2. Diferencias conceptuales: IA vs. robotización

Aunque ambos términos se relacionan estrechamente, conviene diferenciarlos: la IA judicial alude a sistemas con capacidades cognitivas simuladas (aprendizaje, predicción, generación de contenido), mientras que la robotización enfatiza la automatización de tareas repetitivas y estructuradas. La combinación de ambas tecnologías abre la puerta a una “justicia algorítmica” que no solo procesa información más rápido, sino que también puede influir en la interpretación de la ley, lo que exige una regulación específica para salvaguardar la independencia judicial.

El debate académico y legislativo coincide en que la aplicación de estas tecnologías debe enmarcarse en principios éticos y jurídicos sólidos. La Carta Ética Europea sobre el uso de la IA en los sistemas judiciales (CEPEJ, 2018) establece cinco principios rectores: respeto a los derechos fundamentales, no discriminación, calidad y seguridad, transparencia y control por parte de los usuarios, y diseño “orientado al usuario”³⁹⁵. Estos principios constituyen la base para diseñar políticas públicas que armonicen innovación y garantías procesales.

2. MARCO NORMATIVO INTERNACIONAL Y EUROPEO

El desarrollo de sistemas de IA y robotización de la justicia exige un marco regulatorio sólido que equilibre la innovación tecnológica con el respeto a los derechos humanos y los principios del Estado de Derecho. En Europa, esta tarea se articula principalmente a través de tres instrumentos: la Carta de Derechos Fundamentales de la Unión Europea (CDFUE), el Convenio Europeo de Derechos Humanos (CEDH) y la normativa sectorial sobre protección de datos y tecnologías emergentes, especialmente el Reglamento General de Protección de Datos (RGPD) y el Reglamento (UE) 2024/1689 sobre Inteligencia Artificial (Ley de IA)³⁹⁶. A nivel internacional, documentos como la Recomendación de la UNESCO sobre la Ética de la Inteligencia Artificial (2021) y las directrices del Consejo de Europa completan el marco de referencia.

2.1. La Carta de Derechos Fundamentales de la Unión Europea y el Convenio Europeo de Derechos Humanos

La CDFUE, proclamada en el año 2000 y vinculante desde el Tratado de Lisboa (2009), establece los derechos y libertades fundamentales que deben respetar las

³⁹⁵ CEPEJ. (2018). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*. Council of Europe.

³⁹⁶ Carta de Derechos Fundamentales de la Unión Europea, DOUE C 326, 26.10.2012; Convenio Europeo de Derechos Humanos, Roma, 1950; Reglamento (UE) 2016/679; Reglamento (UE) 2024/1689.

instituciones y los Estados miembros al aplicar el Derecho de la Unión. Entre los artículos más relevantes para el análisis de la IA judicial se encuentran:

- Artículo 1: protección de la dignidad humana.
- Artículo 7 y 8: derecho al respeto de la vida privada y a la protección de datos personales.
- Artículo 21: prohibición de discriminación.
- Artículo 47: Derecho a la tutela judicial efectiva y a un juez imparcial.

El CEDH, por su parte, es aplicable a todos los Estados miembros del Consejo de Europa y ha sido interpretado extensamente por el Tribunal Europeo de Derechos Humanos (TEDH). Este Tribunal ha desarrollado estándares sobre independencia judicial, motivación de resoluciones y proporcionalidad que resultan esenciales para evaluar la legitimidad de las tecnologías aplicadas a la justicia.

2.2. Reglamento General de Protección de Datos (RGPD)

El Reglamento (UE) 2016/679, en adelante RGPD, es uno de los pilares regulatorios más importantes para el uso de IA en la justicia. Su artículo 22 titulado “Decisiones individuales automatizadas, incluida la elaboración de perfiles” otorga a los ciudadanos el derecho a no ser objeto de decisiones basadas únicamente en el tratamiento automatizado de datos, salvo excepciones expresamente previstas³⁹⁷. Este artículo es de aplicación directa a sistemas que emitan resoluciones automatizadas o recomendaciones vinculantes, ya que impone:

- Intervención humana significativa: las decisiones deben ser supervisadas por una persona con capacidad real de modificar el resultado.
- Derecho de explicación: las personas tienen derecho a recibir información clara sobre la lógica aplicada y las consecuencias de dichas decisiones.
- Derecho a expresar su punto de vista y a impugnar la decisión.

2.3. Reglamento (UE) 2024/1689 sobre Inteligencia Artificial (Ley de IA)

Adoptado en junio de 2024, el Reglamento de IA establece el primer marco jurídico vinculante para el desarrollo y uso de sistemas de inteligencia artificial en la Unión Europea³⁹⁸. Su enfoque se encuentra basado en el riesgo, clasificando los sistemas en cuatro categorías: riesgo mínimo, limitado, alto y prohibido. Los sistemas utilizados en administración de justicia se consideran de alto riesgo (art. 6), lo que conlleva:

- Requisitos de gobernanza y calidad de datos (arts. 10-11).
- Documentación técnica y trazabilidad (art. 12).

³⁹⁷ Reglamento (UE) 2016/679, art. 22.

³⁹⁸ Reglamento (UE) 2024/1689, arts. 6, 10-15.

- Explicabilidad y supervisión humana obligatoria (arts. 14-15).
- Evaluaciones de conformidad previas a su comercialización o uso (arts. 19-20).
- Registro obligatorio en una base de datos europea (art. 51).

El Reglamento también prohíbe prácticas de IA incompatibles con los derechos fundamentales, como la vigilancia masiva indiscriminada o la manipulación cognitiva (art. 5). Este enfoque refleja el modelo europeo de gobernanza tecnológica, que prioriza los derechos humanos frente a la eficiencia algorítmica, en contraste con modelos más permisivos como el estadounidense o el chino.

2.4. Recomendaciones del Consejo de Europa y la UNESCO

El Consejo de Europa ha sido pionero en establecer principios éticos vinculados a la IA en la justicia, a través de la Carta Ética Europea sobre el Uso de la Inteligencia Artificial en los Sistemas Judiciales (CEPEJ, 2018). Esta Carta articula cinco principios: respeto de derechos fundamentales, no discriminación, calidad y seguridad, transparencia, y control por parte de los usuarios³⁹⁹. Aunque no es jurídicamente vinculante, ha inspirado políticas nacionales, como las estrategias de IA adoptadas en Francia, Alemania y España⁴⁰⁰.

La Recomendación de la UNESCO sobre la Ética de la IA (2021) representa la primera recomendación mundial sobre ética de la IA con un consenso internacional en torno a valores universales como la dignidad, la inclusión y el desarrollo sostenible⁴⁰¹. Este documento propone una gobernanza global de la IA basada en derechos humanos, y subraya la necesidad de aplicar estándares éticos más estrictos en sectores sensibles como la justicia.

2.5. Derecho comparado: experiencias internacionales

Varios países han incorporado inteligencia artificial (IA) en sus sistemas de administración de justicia mediante implementaciones formales y proyectos piloto, con diferentes tipos de aplicaciones, grados de despliegue y áreas judiciales involucradas

China se destaca por la adopción integral de IA en su sistema judicial⁴⁰².

³⁹⁹ CEPEJ. (2018). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*. Council of Europe.

⁴⁰⁰ European Commission. (2021). *Coordinated Plan on Artificial Intelligence 2021 Review*. COM(2021) 205 final.

⁴⁰¹ UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO.

⁴⁰² Fuentes verificables: Informes del Supreme People's Court (SPC), estudios peer-reviewed en la revista "Computer Law & Security Review," y reportes de Reuters.

- Tipo de implementación: automatización de gestión de casos, asistentes virtuales para jueces, análisis predictivos de resultados judiciales, reconocimiento facial para acceso seguro, procesamiento de lenguaje natural para revisión documental.
- Año de implementación: desde 2016 (primeros tribunales basados en IA), despliegue nacional progresivo entre 2017 y 2023.
- Área judicial: tribunales civiles, comerciales y administrativos; sistema completo de e-justice.
- Limitaciones y controversias: críticas por la opacidad algorítmica, sesgos de datos, y falta de transparencia respecto a las garantías procesales.

Estonia es pionera en la digitalización judicial⁴⁰³.

- Tipo de implementación. juez virtual (“Robot Judge”) para reclamaciones civiles de bajo monto; sistemas automatizados de gestión de expedientes.
- Año de implementación: proyecto piloto presentado en 2019, con despliegue limitado a partir de 2021.
- Área judicial: pequeñas reclamaciones civiles y administrativas, digitalización integral de expedientes judiciales.
- Limitaciones: proyecto aún en fase piloto para el “Robot Judge”; cautela sobre la calidad de las resoluciones automáticas.

Brasil ha implementado IA a gran escala en la gestión y predicción judicial⁴⁰⁴.

- Tipo de implementación: sistema “Victor” en el Supremo Tribunal Federal para análisis predictivo de recursos extraordinarios; asistentes virtuales para gestión de expedientes y consulta jurisprudencial.
- Año de implementación: “Victor” desplegado en 2019 y consolidado en 2021.
- Área judicial: Corte Suprema (selección y tramitación de recursos), algunos tribunales regionales.
- Limitaciones: discusión sobre el posible error en valoración automática y la supervisión judicial final.

España ha desarrollado varios proyectos piloto y pruebas en IA aplicada a la justicia⁴⁰⁵.

- Tipo de implementación: asistentes virtuales y sistemas de filtrado automatizado en juzgados de instrucción, pilotaje de procesamiento automático de documentos judiciales y sistemas de recomendación doctrinal.
- Año de implementación: pilotos entre 2021 y 2025, inicio de despliegue parcial en juzgados de varios territorios.

⁴⁰³ Fuentes: Gobierno de Estonia, informes académicos en “AI & Society,” cobertura en Wired y BBC

⁴⁰⁴ Fuentes: Reportes oficiales del STF, artículos en Oxford Journal of Legal Studies, análisis de “The Economist”.

⁴⁰⁵ Fuentes: Ministerio de Justicia, artículos en “Revista Española de Derecho Digital,” medios como *El País*.

- Área judicial: juzgados de instrucción y contencioso-administrativo, gestión documental.
- Limitaciones: proyectos en fase experimental, de momento no hay despliegue nacional.

En Estados Unidos prevalece el uso de IA en evaluación de riesgo y apoyo al proceso judicial⁴⁰⁶.

- Tipo de implementación: herramientas como COMPAS para evaluación de riesgo en sentencias penales, sistemas de asistencia documental, filtrado de expedientes, IA para redacción y revisión legal.
- Año de implementación: uso extendido desde 2013, debate judicial y académico desde 2016.
- Área judicial: tribunales estatales de lo penal, asistentes en grandes despachos jurídicos.
- Limitaciones: controversia por sesgos algorítmicos, falta de transparencia en criterios de decisión.

Otros Ejemplos Relevantes

- Países Bajos: pruebas de IA en análisis de jurisprudencia en tribunales administrativos desde 2022; proyectos limitados a estudios piloto.
- Francia: implementaciones parciales de procesamiento de lenguaje natural para análisis jurisprudencial desde 2019 en el *Conseil d'État*.
- Chile: proyecto piloto de IA para gestión de expedientes en tribunales civiles en 2023, con fase de monitorización y revisión humana.

Hay que destacar que la mayoría de implementaciones exitosas combinan supervisión humana y automatización, sin sustituir fases de revisión o decisión judicial fundamental.

Los sistemas plenamente automatizados, como los experimentos de “Robot Judge” en Estonia y los algoritmos de predicción en China, enfrentan debates sobre la legitimidad democrática, sesgo inherente y control garantista de los procesos.

Los proyectos piloto y parciales son predominantes en Europa occidental y Latinoamérica, donde la cautela regulatoria y la exigencia de transparencia y explicabilidad técnica limitan la expansión de IA plena en el ámbito judicial.

⁴⁰⁶ Fuentes: Judicial Council reports, estudios en “Yale Journal of Law & Technology,” reportajes de ProPublica.

2.6. Necesidad de una regulación procesal específica

Pese al avance normativo, existe un vacío regulatorio en materia procesal. Ni el RGPD ni el Reglamento de IA establecen reglas detalladas sobre cómo deben integrarse los algoritmos en los procesos judiciales, por ejemplo, en la fase probatoria, en el sistema de recursos o en la valoración de pruebas digitales. Este vacío podría generar inseguridad jurídica y desigualdad en el acceso a la justicia. Se propone, por tanto, que los Estados miembros desarrollen leyes procesales sectoriales que adapten las garantías tradicionales (como la inmediación, la contradicción y la motivación) a un entorno algorítmico.

3. PRINCIPALES DERECHOS HUMANOS AFECTADOS

El despliegue de tecnologías de IA y robotización en la administración de justicia no puede analizarse únicamente desde una perspectiva técnica o de eficiencia procesal; es imprescindible un examen jurídico que valore su compatibilidad con los derechos humanos reconocidos en los instrumentos internacionales y europeos y las garantías procesales. La Carta de Derechos Fundamentales de la Unión Europea (CDFUE), el Convenio Europeo de Derechos Humanos (CEDH) y el Reglamento General de Protección de Datos (RGPD) proporcionan un marco normativo sólido que consagra principios como la dignidad humana, el derecho a un juicio justo, la igualdad, la protección de datos personales y la libertad de información⁴⁰⁷. A continuación, se analizan los derechos más directamente implicados en el contexto de la IA judicial.

3.1. Dignidad humana

La dignidad humana, proclamada en el artículo 1 de la CDFUE, constituye la piedra angular del ordenamiento jurídico europeo. Se trata de un derecho absoluto e inderogable, lo que implica que ninguna innovación tecnológica, por más eficiente que sea, puede justificar su vulneración. El Tribunal Europeo de Derechos Humanos (TEDH) ha reiterado que la dignidad es el fundamento del CEDH, sirviendo como criterio interpretativo de otros derechos.

Aplicado a la IA judicial, este principio exige que las personas sean tratadas como sujetos de derechos y no como meros datos estadísticos. Un sistema que “despersonaliza” a los justiciables, reduciendo sus circunstancias individuales a variables algorítmicas, socava la legitimidad de la decisión judicial. Como advierte Hildebrandt (2020)⁴⁰⁸, “el derecho debe proteger a las personas de ser reducidas a representaciones computacionales incompletas”. Esto implica que incluso en contextos donde la

⁴⁰⁷ Carta de Derechos Fundamentales de la Unión Europea, DOUE C 326, 26.10.2012, p. 391-407; Convenio Europeo de Derechos Humanos, Roma, 1950.

⁴⁰⁸ Hildebrandt, M. (2020). *Law for Computer Scientists and Other Folk*. Oxford University Press.

automatización sea eficiente, como en procedimientos masivos, debe preservarse la capacidad del juez para valorar la singularidad de cada caso.

3.2. Derecho a un juicio justo, con todas las garantías y al derecho de defensa

El artículo 47 de la CDFUE y el artículo 6 del CEDH consagran el derecho a un proceso equitativo y a la tutela judicial efectiva ante un juez independiente e imparcial. La justicia algorítmica plantea desafíos inéditos a esta garantía: si una resolución se basa en un algoritmo cuya lógica es opaca o incomprensible incluso para sus desarrolladores, ¿cómo puede el ciudadano ejercer un derecho de defensa efectivo?

El TEDH ha subrayado en múltiples ocasiones que la motivación de las sentencias es parte esencial del derecho a un juicio justo, pues permite a las partes comprender la decisión y preparar su recurso (*Hadi Anastassiou v. Grecia*, 1992)⁴⁰⁹. Un sistema de IA que actúe como “caja negra” sería incompatible con este estándar. Por ello, el artículo 22 del RGPD reconoce expresamente el derecho a no ser objeto de decisiones basadas únicamente en procesos automatizados que produzcan efectos jurídicos significativos, salvo que se garantice la intervención humana significativa⁴¹⁰.

La justicia no puede convertirse en un proceso tecnocrático donde la autoridad judicial se limite a refrendar mecánicamente la decisión que ha proporcionado un algoritmo. Este riesgo se conoce como *automation bias*: la tendencia de los operadores humanos a confiar excesivamente en sistemas automatizados, disminuyendo su propio juicio crítico.

3.3. Igualdad y no discriminación

El artículo 21 de la CDFUE prohíbe toda discriminación, directa o indirecta, por razones como raza, sexo, religión, edad u origen social. Los sistemas de IA, al entrenarse con datos históricos, reproducen patrones de discriminación preexistentes. El célebre caso del algoritmo COMPAS *Correctional Offender Management Profiling for Alternative Sanctions* o Administración de Perfiles de Criminales para Sanciones Alternativas del Sistema de Prisiones en Estados Unidos demostró cómo el sesgo racial se perpetuó en las predicciones de reincidencia, penalizando desproporcionadamente a acusados afroamericanos.

⁴⁰⁹ TEDH. *Hadi Anastassiou v. Grecia*, Demanda núm. 12945/87, Sentencia de 16 de diciembre de 1992.

⁴¹⁰ Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (RGPD).

En Europa, este tipo de sesgos resultan especialmente preocupantes en ámbitos sensibles como el derecho penal, la concesión de libertades condicionales o el acceso a beneficios sociales. El Consejo de Europa, en su Carta Ética Europea sobre el uso de IA en la Justicia (2018), alerta de que “la reproducción de prejuicios históricos a través de datos sesgados constituye una amenaza directa a la igualdad ante la ley”⁴¹¹.

La literatura académica enfatiza que la eliminación completa del sesgo algorítmico es prácticamente imposible aunque se puede ir corrigiendo poco a poco con el tiempo; por ello, es necesario un marco regulatorio que imponga auditorías periódicas de algoritmos para evaluación de los sistemas con implementación de medidas correctivas.

3.4. Derecho a la privacidad y protección de datos personales

El artículo 8 de la CDFUE establece el derecho fundamental a la protección de datos, reforzado por el RGPD. Los sistemas judiciales robotizados procesan ingentes volúmenes de datos personales, incluidos datos sensibles como antecedentes penales, información médica y datos financieros. El TJUE, en casos como *Digital Rights Ireland* (2014), ha declarado que cualquier tratamiento masivo de datos debe cumplir estrictamente con los principios de proporcionalidad y necesidad⁴¹².

La IA judicial plantea desafíos específicos:

- Anonimización insuficiente: los algoritmos pueden reidentificar personas a partir de datos supuestamente anonimizados.
- Decisiones automatizadas: el RGPD prohíbe que una persona sea objeto de decisiones basadas únicamente en tratamiento automatizado, salvo excepciones con garantías adecuadas.
- Transparencia en el tratamiento: los justiciables deben conocer qué datos se utilizan y con qué fines.
-

Estos principios no son meramente técnicos; son garantías sustantivas del Estado de Derecho que deben ser siempre respetadas.

3.5. Libertad de información

El artículo 11 de la CDFUE garantiza el derecho a recibir información plural e imparcial. Si los sistemas judiciales priorizan, filtran o limitan el acceso a información

⁴¹¹ CEPEJ. (2018). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*. Council of Europe.

⁴¹² TJUE. *Digital Rights Ireland Ltd c. Minister for Communications*, Asuntos acumulados C-293/12 y C-594/12, Sentencia de 8 de abril de 2014. <https://curia.europa.eu/juris/document/document.jsf?docid=150642&doclang=ES>

relevante mediante algoritmos opacos, pueden condicionar indebidamente la defensa de los ciudadanos. En el ámbito judicial, el acceso a bases de datos jurisprudenciales y procesales debe ser equitativo y transparente, evitando que el conocimiento del derecho dependa de algoritmos comerciales que prioricen ciertos resultados.

3.6. Tutela judicial efectiva y acceso a los tribunales

El uso de IA tiene el potencial de democratizar la justicia, reduciendo costes y eliminando barreras geográficas. Plataformas como el Civil Resolution Tribunal de Canadá han demostrado que los litigios menores pueden resolverse en línea con menor carga económica. Sin embargo, estas soluciones pueden excluir a personas en situación de vulnerabilidad tecnológica, como personas mayores, colectivos con bajos recursos o con escasas competencias digitales.

La Comisión Europea ha subrayado que toda innovación en justicia debe diseñarse bajo el principio de “accesibilidad universal” para evitar una brecha digital que genere desigualdades en el ejercicio de los derechos⁴¹³. En este sentido, el uso de IA debe considerarse una herramienta complementaria, no sustitutiva de la justicia presencial.

4. RIESGOS JURÍDICOS Y ÉTICOS

La implementación de sistemas de inteligencia artificial (IA) y robotización en el ámbito judicial presenta una serie de riesgos que trascienden lo meramente técnico y se insertan de lleno en el núcleo de las garantías procesales y los derechos fundamentales. El despliegue de algoritmos en funciones tradicionalmente reservadas a jueces y tribunales puede generar escenarios de deslegitimación de las decisiones judiciales, afectación de derechos básicos y pérdida de confianza ciudadana en el sistema de justicia.

Estos riesgos han sido ampliamente documentados tanto por la doctrina como por organismos internacionales, entre ellos el Consejo de Europa y la Comisión Europea, que han insistido en la necesidad de adoptar un enfoque precautorio en el desarrollo e implementación de estas tecnologías⁴¹⁴.

4.1. Sesgos algorítmicos

Uno de los principales riesgos de los sistemas de IA es su propensión a reproducir y amplificar sesgos preexistentes en los datos de entrenamiento. Como señalan Barocas

⁴¹³ Comisión Europea. (2021). *Digital Compass 2030: The European Way for the Digital Decade*. COM(2021) 118 final.

⁴¹⁴ CEPEJ. (2018). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*. Council of Europe.

y Selbst (2016)⁴¹⁵, “los algoritmos no son neutrales: son espejos matemáticos de nuestras estructuras sociales”. Este fenómeno es especialmente problemático en el ámbito judicial, donde la imparcialidad es una exigencia constitucional y convencional.

El caso del sistema COMPAS en Estados Unidos ejemplifica los peligros de un uso indiscriminado de IA: investigaciones periodísticas demostraron que este algoritmo sobreestimaba el riesgo de reincidencia en personas afroamericanas y lo subestimaba en personas blancas⁴. Este tipo de errores no solo vulnera el principio de igualdad ante la ley (art. 21 CDFUE), sino que erosiona la confianza pública en las decisiones judiciales.

En el contexto europeo, el Reglamento de IA (2024/1689) clasifica como “alto riesgo” cualquier sistema destinado a asistir o sustituir decisiones judiciales, precisamente por el impacto potencial en derechos fundamentales⁴¹⁶. Esta categorización implica obligaciones de auditorías periódicas, transparencia y evaluaciones de impacto, aunque persisten desafíos prácticos para garantizar su cumplimiento.

4.2. Opacidad y el problema de la “caja negra”

La opacidad algorítmica es otro de los grandes riesgos. Muchos sistemas basados en aprendizaje profundo (*deep learning*) son intrínsecamente opacos, ya que su lógica interna resulta ininteligible incluso para sus propios desarrolladores. Esta característica plantea una incompatibilidad estructural con el principio de motivación de las resoluciones judiciales, que exige que los jueces expliquen de manera clara las razones de sus decisiones.

El TEDH ha establecido en reiteradas ocasiones que la motivación es un elemento esencial del derecho a un juicio justo (*Hadjji Anastassiou c. Grecia*, 1992)⁴¹⁷. La introducción de sistemas “caja negra” en el proceso judicial podría, por tanto, vulnerar directamente el artículo 6 CEDH y el artículo 47 CDFUE.

La “explicabilidad” no debe ser una característica opcional de los sistemas de IA judicial, sino un requisito normativo básico, ya que sin ella es imposible ejercer un control efectivo. La opacidad tecnológica no puede convertirse en excusa para limitar derechos procesales.

⁴¹⁵ Barocas, S., & Selbst, A. D. (2016). Big data’s disparate impact. *California Law Review*, 104(3), 671–732. <https://doi.org/10.2139/ssrn.2477899>

⁴¹⁶ Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, sobre Inteligencia Artificial.

⁴¹⁷ TEDH. *Hadjianastassiou c. Grecia*, Demanda núm. 12945/87, Sentencia de 16 de diciembre de 1992.

4.3. Pérdida de supervisión humana y automatización de la autoridad

Otro riesgo significativo es la delegación excesiva de decisiones en sistemas automatizados, fenómeno conocido como *automation bias*. Los operadores humanos tienden a confiar más en las decisiones automatizadas que en su propio juicio, especialmente cuando los sistemas presentan una apariencia de objetividad científica.

En la justicia, este sesgo puede generar una automatización de la autoridad, donde los jueces se convierten en meros validadores de decisiones algorítmicas, debilitando su rol constitucional como garantes de los derechos fundamentales. El Reglamento de IA incorpora el derecho de intervención humana significativa, pero aún no detalla de manera precisa qué grado de supervisión es necesario para cumplir con este estándar⁴¹⁸.

4.4. Ciberseguridad

La robotización judicial implica una dependencia creciente de infraestructuras digitales. Esto aumenta la superficie de exposición a ataques cibernéticos, que podrían tener consecuencias devastadoras. Un ataque dirigido a sistemas automatizados de gestión procesal podría paralizar miles de procedimientos o filtrar datos sensibles, vulnerando el derecho a la intimidad y a la protección de datos (art. 7 y 8 CDFUE)⁴¹⁹.

Informes recientes del European Union Agency for Cybersecurity (ENISA)⁴²⁰ advierten que los sistemas de IA presentan vectores de ataque únicos, como la manipulación de datos de entrenamiento (*data poisoning*) o la generación de ejemplos adversariales que engañan a los modelos. La protección cibernética, por tanto, no es solo una cuestión técnica, sino una obligación jurídica derivada de los principios de seguridad y fiabilidad recogidos en el Reglamento de IA.

4.5. Deshumanización de la justicia

Finalmente, el riesgo más intangible, pero igualmente grave, es la deshumanización del proceso judicial. La justicia no es solo un mecanismo para resolver disputas, sino un pilar simbólico y social que garantiza la cohesión democrática. Como sostiene Susskind, “la justicia es un servicio público que debe inspirar confianza; reemplazar el contacto humano por interfaces digitales sin un marco ético adecuado puede erosionar esa confianza”⁴²¹.

El acceso a un juez humano no es únicamente una garantía procesal, sino una expresión del principio de dignidad humana. El TEDH ha subrayado que los procesos

⁴¹⁸ Reglamento (UE) 2024/1689, arts. 14-15.

⁴¹⁹ Carta de Derechos Fundamentales de la Unión Europea, DOUE C 326, 26.10.2012.

⁴²⁰ ENISA. (2021). *Artificial intelligence cybersecurity challenges: Threat landscape for AI*. Publications Office of the European Union. <https://doi.org/10.2824/179806>

⁴²¹ Susskind, R. (2019). *Online Courts and the Future of Justice*. Oxford University Press.

judiciales deben permitir a las partes ser escuchadas y tratadas como sujetos de derechos⁴²². El desafío consiste en aprovechar las ventajas de la automatización sin despersonalizar la justicia, manteniendo siempre el principio de *human in the loop* como pilar fundamental del diseño tecnológico⁴²³.

5. PROPUESTAS Y BUENAS PRÁCTICAS

La integración de sistemas de inteligencia artificial (IA) y procesos de robotización en la administración de justicia requiere un enfoque regulatorio y ético que priorice los derechos humanos, la transparencia y la confianza ciudadana en el sistema judicial. En este sentido, las recomendaciones que se exponen a continuación se inspiran en el marco normativo europeo, la jurisprudencia del Tribunal de Justicia de la Unión Europea (TJUE) y el Tribunal Europeo de Derechos Humanos (TEDH), así como en la literatura académica sobre gobernanza algorítmica.

Estas propuestas no solo buscan prevenir riesgos, sino también establecer un modelo de justicia algorítmica responsable que refuerce el Estado de Derecho.

5.1. Principio “Human in the Loop”: supervisión humana significativa

El primer pilar para garantizar una justicia respetuosa con los derechos fundamentales es el principio de intervención humana significativa, recogido en el artículo 22 del RGPD y reforzado en los artículos 14 y 15 del Reglamento de IA⁴²⁴. Este principio exige que las decisiones automatizadas estén siempre sujetas a una revisión humana efectiva, que no se limite a un mero trámite formal, sino que implique capacidad real de corrección y control.

El TEDH ha sostenido que el derecho a un juez independiente no se satisface con una revisión superficial⁴²⁵. En el contexto de la IA judicial, esto implica que los jueces no deben actuar como simples “notarios” de resoluciones algorítmicas, sino que deben ejercer un análisis crítico, con formación específica para comprender los sistemas que supervisan para tomar decisiones humanas y justas⁴²⁶.

5.2. Auditorías algorítmicas obligatorias y certificación

La auditoría algorítmica es una herramienta indispensable para detectar sesgos, errores y vulnerabilidades de los sistemas de IA. El Reglamento de IA ya establece

⁴²² TEDH. *Del Río Prada c. España*, Demanda núm. 42750/09, Sentencia de 21 de octubre de 2013.

⁴²³ CEPEJ. (2018). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*. Council of Europe.

⁴²⁴ Reglamento (UE) 2016/679, art. 22; Reglamento (UE) 2024/1689, arts. 14-15

⁴²⁵ TEDH. *H. c. Bélgica*, Demanda núm. 8950/80, Sentencia de 30 de noviembre de 1987.

⁴²⁶ Susskind, R. (2019). *Online Courts and the Future of Justice*. Oxford University Press.

obligaciones de evaluación de conformidad para sistemas de alto riesgo, pero sería recomendable complementarlas con auditorías periódicas independientes que incluyan:

1. Análisis de sesgos: Identificación de patrones discriminatorios.
2. Pruebas adversariales: Evaluación de la resistencia a ataques y manipulaciones.
3. Validación estadística: Verificación de precisión y fiabilidad.

Convendría implementar un sistema de auditoría multinivel para clasificar los sistemas en niveles de impacto y determinar requisitos proporcionales de revisión⁴²⁷ con certificaciones algorítmicas obligatorias para todo software utilizado en justicia.

5.3. Explicabilidad y transparencia algorítmica

La transparencia es una condición esencial del Estado de Derecho. En materia judicial, el derecho a una resolución motivada implica que los ciudadanos comprendan las razones de las decisiones que les afectan⁴²⁸. Por tanto, cualquier sistema algorítmico debe incorporar mecanismos de explicabilidad que permitan a jueces y usuarios conocer, al menos en términos generales, los criterios utilizados por el modelo.

La explicabilidad no es solo una exigencia técnica, sino una obligación democrática, ya que sin transparencia no hay rendición de cuentas. Este principio podría traducirse en:

- Publicación de documentación técnica accesible.
- Explicaciones adaptadas al público no especializado.
- Registros de decisiones algorítmicas para permitir auditorías ex post.

5.4. Regulación procesal

Aunque el Reglamento de IA y el RGPD constituyen marcos normativos sólidos, ninguno aborda de manera específica los desafíos procesales que plantea la justicia algorítmica. Es necesario desarrollar una regulación sectorial que:

- Establezca el grado de valor probatorio de las decisiones automatizadas.
- Detalle cómo integrar algoritmos en las fases de recurso y revisión.
- Establezca procedimientos claros para impugnar errores algorítmicos.

En este sentido el derecho procesal debe evolucionar hacia un marco *de dimensión computacional*, en el que las garantías clásicas se proyecten y adapten a entornos digitales.

5.5. Alfabetización para operadores jurídicos

El uso de IA en justicia exige alfabetización tecnológica para jueces, fiscales, abogados y personal judicial. La CEPEJ recomienda que los Estados miembros integren

⁴²⁷ Government of Canada. (2019). *Directive on Automated Decision-Making*.

⁴²⁸ TEDH. *Hadjianastassiou c. Grecia*, Demanda núm. 12945/87, Sentencia de 16 de diciembre de 1992.

contenidos sobre ética, gobernanza y funcionamiento básico de algoritmos en la formación inicial y continua⁴²⁹.

La alfabetización tecnológica jurídica es clave como destaca el Reglamento de IA en su artículo 4 y garantizar que los jueces mantengan su capacidad de supervisión crítica². Esta formación debería incluir nociones de estadística, algoritmos, modelos de IA, ciberseguridad y derechos digitales.

5.6. Principios de diseño ético: equidad, inclusión y accesibilidad

La justicia algorítmica debe diseñarse bajo el principio de inclusión digital. La Comisión Europea (2021) ha advertido que cualquier innovación tecnológica en servicios públicos debe garantizar la accesibilidad universal⁴³⁰. Esto implica:

- Interfaces adaptadas para personas con discapacidad.
- Lenguaje jurídico claro en plataformas digitales.
- Opciones presenciales para ciudadanos con brecha digital.

Asimismo, la UNESCO (2021) recomienda aplicar principios de equidad y diversidad desde la fase de diseño, incorporando datos representativos y equipos multidisciplinares⁴³¹.

5.7. Estrategia de ciberseguridad integral

El aumento de la robotización judicial requiere protocolos de ciberseguridad avanzados. ENISA la Agencia de la Unión Europea para la Ciberseguridad, no publica directamente "protocolos de ciberseguridad" en el sentido de un estándar técnico, sino que proporciona guías y recomendaciones, especialmente para implementar la Directiva NIS2, que sí establece requisitos de seguridad. Sus directrices se centran en la implementación de medidas de gestión de riesgos, respuesta a incidentes, y el uso de controles como el cifrado robusto y la autenticación. Igualmente identifica como riesgos prioritarios los ataques de *data poisoning* y *model inversion*, que podrían manipular resoluciones o filtrar datos sensibles⁴³². Las buenas prácticas incluyen cifrado de extremo a extremo en todas las comunicaciones judiciales, segmentación de redes judiciales para evitar propagación de ataques y elaboración de planes de contingencia para restaurar sistemas tras los incidentes que puedan producirse.

⁴²⁹ CEPEJ. (2018). *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their Environment*. Council of Europe.

⁴³⁰ Comisión Europea. (2021). *Digital Compass 2030: The European Way for the Digital Decade*. COM(2021) 118 final

⁴³¹ UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. Paris: UNESCO.

⁴³² ENISA. (2021). *Artificial Intelligence Cybersecurity Challenges: Threat Landscape for AI*. Publications Office of the European Union.

5.8. Gobernanza algorítmica y supervisión institucional con creación de entornos de prueba o *sandbox*

Se propone la creación de agencias independientes de supervisión algorítmica y de IA encargadas de evaluar y certificar todos los sistemas de IA utilizados en el ámbito judicial. Estas agencias podrían actuar como órganos técnicos de apoyo a los consejos generales del poder judicial de cada país, siguiendo el modelo de agencias europeas de protección de datos.⁴³³

La gobernanza y supervisión de la inteligencia artificial requiere la promoción de entornos de prueba o *sandboxes* regulatorios que faciliten una correcta adaptación de los sistemas, reforzando así la protección de los usuarios y minimizando riesgos de sesgos discriminatorios. Estos espacios de experimentación constituyen un instrumento clave no solo para que las empresas puedan innovar en un marco seguro, sino también para fomentar un contexto internacional de confianza en el desarrollo de la IA, en línea con iniciativas como pactos globales o códigos de buenas prácticas.

Asimismo, resulta esencial que se prevea el apoyo técnico y asesoramiento especializado a jueces y tribunales en aquellos procesos vinculados a la IA, junto con la elaboración de guías prácticas que orienten la implementación de sistemas éticos, transparentes y fiables.

De esta manera, se impulsa la difusión de información sobre los requisitos normativos y se acompaña a las organizaciones en la adopción de las medidas necesarias para garantizar el cumplimiento efectivo de los estándares de gobernanza responsable.

6. CONCLUSIONES

La incorporación de sistemas de inteligencia artificial (IA) y robotización en la administración de justicia constituye uno de los mayores desafíos contemporáneos para el Derecho, la ética y las políticas públicas a nivel mundial. Si bien estas tecnologías ofrecen un potencial transformador capaz de incrementar la eficiencia procesal, democratizar el acceso a la justicia y optimizar los recursos judiciales, su implementación sin garantías sólidas puede erosionar derechos fundamentales, socavar el principio de independencia judicial y debilitar la legitimidad democrática de los sistemas jurídicos.

El breve análisis desarrollado en este artículo evidencia que la justicia algorítmica no es neutra, sino que sus resultados reflejan tanto la calidad de los datos que la

⁴³³ En España contamos ya con la AESIA Agencia Española de Supervisión de IA, entidad de derecho público adscrita a la Secretaría de Estado de Digitalización e Inteligencia Artificial. Su función principal será la de llevar a cabo tareas de supervisión, asesoramiento, concienciación y formación, a empresas públicas y privadas para la adecuada implementación de la normativa entorno al adecuado uso de la IA y **concretamente de los algoritmos**.

alimentan como las prioridades normativas y políticas de quienes la diseñan. La ilusión de objetividad matemática puede ocultar sesgos históricos y desigualdades estructurales, generando una falsa sensación de neutralidad tecnológica.

Por ello, es esencial entender que el debate no se limita a cuestiones técnicas, sino que entraña una dimensión ética y jurídica profunda sobre el rol del juez, el valor de la deliberación humana y el significado mismo del Estado de Derecho.

Uno de los consensos más claros que emerge de los marcos normativos europeos es que la tecnología debe estar al servicio de la persona, y no al revés. Tanto la Carta de Derechos Fundamentales como el Reglamento de IA establecen que las aplicaciones de alto riesgo, como las empleadas en el sistema judicial, deben someterse a estrictos controles, transparencia y supervisión humana significativa⁴³⁴. Esta visión se enmarca en el modelo europeo de gobernanza tecnológica, que prioriza los derechos fundamentales frente a una lógica puramente utilitarista de eficiencia.

Europa lidera el escenario global con el Reglamento de IA, un marco pionero que clasifica los sistemas por niveles de riesgo y establece obligaciones proporcionadas. Sin embargo, esta normativa es solo un primer paso: su naturaleza generalista no aborda en detalle aspectos procesales clave, como el valor probatorio de las decisiones algorítmicas, los procedimientos de recurso ante errores automatizados o la integración de estas herramientas en la fase de instrucción penal. Se hace necesario, por tanto, complementar este marco con legislación procesal sectorial y directrices específicas para el ámbito judicial. Estas normas deben definir estándares claros sobre explicabilidad, auditorías obligatorias, formación interdisciplinar de operadores jurídicos y participación ciudadana en el diseño de tecnologías aplicadas a la justicia.

El análisis de los riesgos -sesgos algorítmicos, opacidad, ciberseguridad, pérdida de supervisión humana y deshumanización de la justicia- revela que las amenazas no son meramente hipotéticas. Casos como COMPAS en Estados Unidos o el creciente uso de tribunales inteligentes en China muestran que, sin un marco ético sólido, la IA puede convertirse en una herramienta de control y exclusión.

Para evitar estos escenarios, se requiere una cultura jurídica de derechos digitales, que impregne todas las fases del ciclo de vida de los sistemas de IA desde la planificación, el diseño y el entrenamiento hasta su implementación y evaluación con la instalación de las mejoras necesarias para su adecuado funcionamiento.

El futuro de la justicia pasa por una colaboración equilibrada entre humanos y algoritmos. La IA no debe reemplazar el juicio crítico del juez, sino complementarlo. Esto exige supervisión humana en todas las decisiones, transparencia y trazabilidad total de los sistemas utilizados, auditorías periódicas e independientes para evaluar

⁴³⁴ Reglamento (UE) 2024/1689, arts. 6, 14-15.

sesgos y riesgos, educación interdisciplinar que permita a los juristas comprender los fundamentos técnicos de estas tecnologías y garantías de accesibilidad para evitar nuevas formas de exclusión digital.

La legitimidad del sistema judicial en la era digital dependerá de su capacidad para preservar la humanidad en medio de la automatización. La justicia no es solo un servicio, es una institución simbólica que encarna valores democráticos que deben preservarse.

Como conclusión hay que destacar que se considera que la digitalización del poder judicial ofrece una oportunidad histórica para mejorar la eficiencia y el acceso a la justicia, pero también una advertencia: la justicia desprovista de humanidad deja de ser justicia. La regulación europea, con el RGPD y el Reglamento de IA, representa un marco pionero, pero debe reforzarse con políticas públicas de educación tecnológica, mecanismos de rendición de cuentas y participación ciudadana. En definitiva, el desafío central es garantizar que el algoritmo se someta a la persona y no la persona al algoritmo. Este principio rector debe guiar todas las reformas futuras, asegurando que los avances tecnológicos fortalezcan -y nunca debiliten- los valores esenciales del Estado de Derecho.

BIBLIOGRAFÍA

- Civil Resolution Tribunal. (n.d.). *About us*. Civil Resolution Tribunal. Retrieved from <https://civilresolutionbc.ca>
- Comisión Europea. (2021). *Digital compass 2030: The European way for the digital decade* (COM/2021/118 final). Publications Office of the European Union.
- ENISA. (2021). *Artificial intelligence cybersecurity challenges: Threat landscape for AI*. Publications Office of the European Union. <https://doi.org/10.2824/179806>
- European Commission. (2020). *The use of artificial intelligence in public administration*. Publications Office of the European Union.
- European Commission. (2021). *Coordinated plan on artificial intelligence 2021 review* (COM/2021/205 final). Publications Office of the European Union.
- European Commission for the Efficiency of Justice (CEPEJ). (2018). *European ethical charter on the use of artificial intelligence in judicial systems and their environment*. Council of Europe.
- European Commission for the Efficiency of Justice (CEPEJ). (2020). *European judicial systems: Efficiency and quality of justice*. Council of Europe Publishing.
- Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*. Oxford University Press.
- Floridi, L., & Cows, J. (2021). A unified framework of five principles for AI in society. *Harvard Data Science Review*, 3(1). <https://doi.org/10.1162/99608f92.8cd550d1>
- Government of Canada. (2019). *Directive on automated decision-making*. Government of Canada. Retrieved from <https://www.canada.ca>

- Hildebrandt, M. (2020). *Law for computer scientists and other folk*. Oxford University Press.
- Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (RGPD). DOUE L 119, 4.5.2016.
- Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, sobre inteligencia artificial y por el que se modifican determinados actos legislativos de la Unión.
- Susskind, R. (2019). *Online courts and the future of justice*. Oxford University Press.
- Tribunal de Justicia de la Unión Europea. (2014, April 8). *Digital Rights Ireland Ltd v. Minister for Communications* (C-293/12 y C-594/12).
- Tribunal Europeo de Derechos Humanos. (1992, December 16). *Hadjianastassiou v. Greece* (Application No. 12945/87).
- Tribunal Europeo de Derechos Humanos. (2002, April 29). *Pretty v. United Kingdom* (Application No. 2346/02).
- Tribunal Europeo de Derechos Humanos. (2013, October 21). *Del Río Prada v. Spain* (Application No. 42750/09).
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. UNESCO.



CAPÍTULO 11. DERECHOS FUNDAMENTALES DEL ART. 20. 1, a) y d) CE: IMPACTO DE LA INTELIGENCIA ARTIFICIAL⁴³⁵

Autor: Dr. **PEDRO RÓDENAS CORTÉS**

Profesor de Derecho Civil

Facultad de Derecho (Cáceres)

Universidad de Extremadura

rodenas@unex.es

ORCID: 0000-0001-7032-8077

Resumen

El presente trabajo lleva a cabo un análisis sistemático y práctico del modo en el que la IA afecta al ejercicio del derecho fundamental de la libertad de expresión y de información, reconocido en el artículo 20 de la Constitución Española. A lo largo de esta investigación, se estudian las principales aplicaciones de estos sistemas en los procesos de creación, difusión y consumo de contenidos, así como su impacto en el espacio público digital.

La exploración en este tema, evidencia que la IA opera como un nuevo intermediario informativo, cuya intervención puede considerarse legítima siempre que se respeten determinadas garantías vinculadas a la supervisión humana y la protección de los derechos fundamentales. Asimismo, se aborda la influencia que tienen estos sistemas en la formación de la opinión pública, especialmente entre adolescentes y jóvenes, que utilizan la IA como fuente de información y apoyo emocional.

El trabajo concluye destacando la relación entre IA y libertad de expresión, que exige un marco regulatorio mayor que combine innovación tecnológica y la salvaguarda de derechos, así como una participación activa tanto de las instituciones públicas como de la sociedad civil para garantizar que la IA se use de forma adecuada.

Palabras clave

Inteligencia artificial – derecho fundamental – libertad de expresión y de información – espacio público digital – opinión pública – innovación tecnológica

⁴³⁵ El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte (IAIURIS)* de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ca8b86cec8e8/content>

Abstract

This paper conducts a systematic and practical analysis of how artificial intelligence affects the exercise of the fundamental right to freedom of expression and information, recognised in article 20 of the Spanish Constitution. Throughout this research, the main applications of these systems in the processes of content creation, dissemination and consumption are studied, as well as their impact on the digital public space.

Exploration of this topic shows that artificial intelligence operates as a new information intermediary, whose intervention can be considered legitimate provided that certain guarantees related to the human supervision and the protection of fundamental rights are respected. It also addresses the influence these systems have on the formation of public opinion, especially among adolescents and young people, who use artificial intelligence as a source of information and emotional support.

The paper concludes by highlighting the relationship between artificial intelligence and freedom of expression, which requires a broader regulatory framework that combines technological innovation and the safeguarding of rights, as well as active participation by both public institutions and civil society to ensure that artificial intelligence is used properly.

Keywords

Artificial intelligence – fundamental right – freedom of expression and information – digital public space – public opinion – technological innovation

Sumario: BREVE PROEMIO 1.- CONCEPTO Y CARACTERÍSTICAS DE LA INTELIGENCIA ARTIFICIAL 1.1.- Antecedentes y fundamentos 1.2.- Concepto, características y tipología 2.- LIBERTAD DE EXPRESIÓN Y DE INFORMACIÓN 2.1.- Titularidad y contenido del derecho de libertad de expresión e información 2.2.- Límites y conflictos con otros derechos 2.3.- Requisitos de la libertad de información: veracidad y diligencia 2.4.- Garantías constitucionales 2.5.- Referencia a la dimensión internacional de la libertad de expresión 3.- IMPACTO DE LA INTELIGENCIA ARTIFICIAL 3.1.- Funcionalidad de la inteligencia artificial confrontada a la libertad de expresión e información 3.2.- Nuevas formas de participación y debate público condicionadas por la inteligencia artificial 3.3.- Amenazas emergentes de la inteligencia artificial sobre la libertad de expresión e información 3.4.- El dominio de la IA en el comportamiento e información recibida por jóvenes 4.- LA RESPONSABILIDAD CIVIL DERIVADA DEL MAL USO DE LA INTELIGENCIA ARTIFICIAL 5.- CONCLUSIONES BIBLIOGRÁFICA REPERTORIO JURISPRUDENCIAL

BREVE PROEMIO

El desarrollo tan acelerado de la inteligencia artificial (en adelante, IA)⁴³⁶, en las dos últimas décadas, ha conseguido transformar de forma muy pronunciada los

⁴³⁶ La Comisión Europea la define como sistemas de software, posiblemente también de hardware, diseñados por humanos que, ante un objetivo complejo, actúan en la dimensión física o digital:

procesos comunicativos y las dinámicas de acceso, producción y difusión de la información, generando un nuevo panorama en el que se replantean las condiciones materiales para el ejercicio de la libertad de expresión e información, previstas en el artículo 20.1. a) y d) de la Constitución Española.

La expansión de estos sistemas, capaces de analizar todo tipo de datos y de forma masiva, demuestran desafíos cuya relevancia excede lo tecnológico, ya que afecta directamente a la formación de la opinión pública, a la calidad del debate y al grado de autonomía con el que los ciudadanos participan en la esfera globalizada digital.

En este sentido, la IA) se ha convertido no solo en un instrumento complementario de la comunicación de las personas, sino que también se trata de un intermediario informativo cuyo funcionamiento y alcance requieren una revisión jurídico-constitucional rigurosa.

Los sistemas informáticos han generado un entorno dónde se discute, ordenan preferencias, moderan contenidos e influyen en la visibilidad de las ideas, la pregunta central ya no es únicamente cómo se expresa el ciudadano, sino quién controla las infraestructuras que determinan qué puede expresarse y qué llega efectivamente a la ciudadanía.

Por lo tanto, el presente trabajo tiene como objetivo analizar de forma sistemática la relación entre la IA y la libertad de expresión e información, atendiendo tanto a la dimensión conceptual y técnica, como a las implicaciones jurídicas derivadas de su uso en los procesos comunicativos de la actualidad.

A su vez, se aborda la configuración constitucional de la libertad de expresión, analizando su contenido, sus límites y sus garantías, así como su proyección internacional. A continuación, se examina el impacto concreto de la IA en estas libertades, identificando las aplicaciones tecnológicas más relevantes, las nuevas formas de participación y las posibles amenazas específicas (como la censura automatizada, los *deepfakes*, la desinformación o los sesgos algorítmicos) que comprometen la calidad en la comunicación y afecta de forma muy notable en el pluralismo democrático, además de generar responsabilidades civiles.

percibiendo su entorno, a través de la adquisición e interpretación de datos estructurados o no estructurados y, b) razonando sobre el conocimiento, procesando la información derivada de estos datos y decidiendo las mejores acciones para lograr el objetivo dado. Los sistemas de IA pueden usar reglas simbólicas o aprender un modelo numérico. También pueden adaptar su comportamiento al analizar cómo el medio ambiente se ve afectado por sus acciones previas. Dicho de otro modo: la IA es un campo de la informática que se enfoca en crear sistemas que puedan realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento y la percepción (consultado el día 9/12/2025, a las 19:08 horas, en <https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr>).

Finalmente, en su conjunto, se pretende ofrecer una reflexión y una investigación jurídica de forma integral y fundamentada a un fenómeno que, por su magnitud e impacto, constituye uno de los mayores retos contemporáneos para el Derecho en general, las libertades individuales de los ciudadanos y para la teoría democrática.

1.- CONCEPTO Y CARACTERÍSTICAS DE LA INTELIGENCIA ARTIFICIAL

1.1.- Antecedentes y fundamentos

Aparecen sus raíces en una combinación de avances matemáticos, lógicos y computaciones que se remontan a principios del siglo XX. En la década de 1940, *Norbert Wiener* y *John von Neumann*⁴³⁷ desarrollaron teorías sobre sistemas y programación que sentaron las bases para la futura IA, mientras que la literatura y la ciencia ficción comenzaron a explorar la idea de máquinas inteligentes, consolidando conceptos que hoy en día son parte de la cultura popular.

Por su parte, *Alan Mathison Turing*⁴³⁸ fue un matemático británico considerado el padre de la informática artificial. Su contribución más conocida fue la máquina de Turing, que constituye la base teórica de los ordenadores modernos y que se expande como objeto de estudio por todo el mundo. En el ámbito de la inteligencia artificial, *Turing* fue un visionario. Diseñó los primeros programas de ordenador capaces de realizar tareas que requerían razonamiento y proponiendo ideas clave como la “prueba de Turing” para analizar la inteligencia de las máquinas⁴³⁹.

Tras esta trayectoria, el nacimiento de la IA se sitúa en 1956, donde *John McCarthy*⁴⁴⁰ acuñó el término de IA durante la histórica conferencia de *Dartmouth*, en la que un grupo de investigadores discutió la posibilidad de crear máquinas capaces de pensar como seres humanos.

Entre los primeros objetivos estuvieron el reconocimiento de patrones y los juegos de ordenador, contribuyendo al nacimiento de avances posteriores en resolución de problemas y aprendizaje automatizado. De forma paralela, se iniciaron los primeros

⁴³⁷ Ambos son matemáticos que contribuyeron a fijar las bases de la configuración de la IA. *Norbert Wiener* (1894-1964) fue un matemático y filósofo estadounidense, pionero de la cibernética, mientras que *John von Neumann* (1903-1957) fue un matemático húngaro-estadounidense, clave en la teoría de juegos y la arquitectura de los computadores modernos.

⁴³⁸ *Alan Mathison Turing* (*Paddington, London*; 23 de junio de 1912-*Wilmslow, Cheshire*; 7 de junio de 1954) fue un matemático, lógico, informático teórico, criptógrafo, filósofo y biólogo teórico británico. Es considerado como uno de los padres de la ciencia de la computación y precursor de la informática moderna.

⁴³⁹ LÓPEZ DE MÁNTARAS BADIA, R., MESEGUER GONZÁLEZ, P., *Inteligencia artificial*, Editorial CSIC, Madrid, 2017.

⁴⁴⁰ Entre sus muchas reflexiones sobre la inteligencia artificial, una cita notable de *McCarthy* es: “La inteligencia artificial es la ciencia de hacer que las máquinas hagan cosas que requerirían inteligencia si las hicieran personas”. El legado de *John McCarthy* perdura en la inteligencia artificial moderna.

intentos de articular el lenguaje natural, motivados por la necesidad de traducciones automáticas durante la Guerra Fría. Estos programas enfrentaron dificultades por la ambigüedad del lenguaje y la falta de conocimiento de sentido común, un reto que sigue siendo relevante en la IA actual.

A lo largo de las últimas décadas, este sistema ha evolucionado gracias a avances en potencia informática y nuevos algoritmos, consolidándose como una de las tecnologías que más debate ha desencadenado y con mayor impacto en la sociedad.

1.2. Concepto, características y tipología

Tras la contextualización histórica y fundamento teórico, la IA se define como una disciplina de la informática y de la ingeniería que desarrolla sistemas capaces de realizar tareas que normalmente requieren inteligencia humana, como razonar, aprender, planificar, reconocer patrones o procesar lenguaje natural. No se trata únicamente de replicar habilidades humanas, sino de crear máquinas capaces de realizar comportamientos inteligentes, ya sea de forma especializada o general.

Todo esto en su conjunto, da lugar a que la IA se diferencie de manera sustancial de otros sistemas informáticos tradicionales, al integrar capacidades que imitan ciertos aspectos del razonamiento humano, pero con un alcance y una velocidad que los sistemas convencionales no poseen. La IA mejora su rendimiento a medida que procesa los datos, ya sea mediante aprendizaje supervisado, no supervisado o por refuerzo, va adaptándose a nuevas situaciones sin necesidad de intervención humana directa⁴⁴¹.

Además, es capaz de analizar información compleja e identificar patrones que le permiten proponer soluciones a problemas complejos, lo que constituye su alta capacidad de razonamiento. Estos sistemas también están diseñados para interactuar con personas y con otros procesos tecnológicos, procesando lenguaje natural, reconociendo la voz o generando contenido, lo que facilita, en parte, la comunicación y el acceso a la información.

Por último, destaca por su capacidad para procesar grandes volúmenes de datos de forma rápida y sistemática, filtrando y sintetizando información masiva, lo que resulta práctico en ámbitos como el periodismo digital o la educación, entre otros⁴⁴².

⁴⁴¹ Ver el informe de la Organización de las Naciones Unidas, para la Educación, la Ciencia y la Cultura, acerca de la “Recomendación de la ética de la inteligencia internacional, adoptada el 23 de noviembre de 2021”, consultado, el día 12-11-2025, a las 13:25 horas, en la web de la Asociación Española de Inteligencia Artificial, dentro de su sección “inteligencia artificial (Online)”, disponible en aepia.org/wp-content/uploads/2024/04/UNESCO_recomendaciones_etica.pdf

⁴⁴² NAVAS NAVARRO, S., GÓRRIZ, C., *Inteligencia artificial: tecnología, derecho*, 1ª edición, Tirant lo Blanch, Valencia, 2017.

Estas características permiten que la IA no solo automatice tareas rutinarias, sino que también influye en la toma de decisiones complejas, optimice procesos y contribuya a la producción de información.

Ahora bien, a raíz de las características, es conveniente hacer un análisis sobre los distintos enfoques que permiten entender mejor el funcionamiento de este sistema. Según su alcance, se distingue entre la IA débil o estrecha, que se centra en tareas específicas y limitadas, superando al ser humano en determinados ámbitos, pero sin alcanzar inteligencia general; y la IA fuerte o general, que aspira a simular la inteligencia humana en toda su amplitud, con capacidad de aprendizaje autónomo y con un razonamiento muy flexible.

Desde el punto de vista de la metodología, la IA puede ser simbólica, que funciona mediante reglas claras y explícitas, es decir, todo el conocimiento que maneja está organizado en forma de conceptos o normas que el sistema utiliza para razonar. Mientras que, la conexionista, se basa en redes neuronales y aprendizaje automático, que intenta imitar de forma muy simplificada la manera en la que funciona el cerebro humano. En este último caso, la máquina no sabe qué hacer, ella descubre cómo resolver un problema a partir de ejemplos.

En términos de aprendizaje, los sistemas pueden ser supervisados, siendo el más común, el sistema aprehende a partir de datos etiquetados, es decir, con la respuesta correcta indicada; los no supervisados, cuando identifican patrones sin guía externa, ya que la máquina analiza los datos por sí sola; o los basados en refuerzos, donde la IA aprende adquiriendo conocimiento mediante ensayo y error, recibiendo recompensas o penalizaciones según su desempeño⁴⁴³.

Finalmente, según su comportamiento y capacidades, se distinguen sistemas que piensan como humanos, simulando procesos de razonamiento; sistemas que actúan como humanos, imitando comportamientos específicos; sistemas que piensan racionalmente, tratando de encontrar la solución más adecuada utilizando la lógica a su disposición; y sistemas que actúan racionalmente, tomando decisiones óptimas, basadas en la información que tienen disponible.

Cabe señalar que estas clasificaciones no son excluyentes y, con frecuencia, un mismo sistema combina varias técnicas y enfoques dependiendo de los objetivos y la complejidad de las tareas que se les impongan. En el ámbito comunicativo e informativo, comprender estas distinciones resulta fundamental para analizar cómo la IA procesa grandes volúmenes de información, genera contenido y asiste en la toma de decisiones a las personas.

⁴⁴³ CORTÉS GARCÍA, U., BÉJAR ALONSO, J., MORENO RIBAS, A., *Inteligencia artificial*, Universidad Politécnica de Cataluña, Barcelona, 1993.

2. LIBERTAD DE EXPRESIÓN Y DE INFORMACIÓN

El artículo 20 de la Constitución Española reconoce y garantiza uno de los derechos fundamentales más trascendentes en una sociedad democrática: la libertad de expresión y de información⁴⁴⁴. Este derecho constituye principalmente la base del pluralismo ideológico y del libre intercambio de ideas e información, pilares sobre los que se construye todo sistema democrático. Por lo que, cabe destacar que, tras la dictadura franquista, el constituyente de 1978 otorgó relevancia a este artículo, estableciendo un conjunto de libertades que regulan la expresión de ideas y la información en España⁴⁴⁵.

En la misma línea interpretativa, encontramos referencias de carácter internacional otorgando una preponderancia esencial en el pluralismo democrático a la libertad de expresión, citándose a título ilustrativo el caso estadounidense *Texas v. Johnson*⁴⁴⁶, donde se recogió por su Corte Suprema que “la libertad de expresión es el corazón de la democracia”.

Retornando al ordenamiento español, se ha reiterado en numerosas ocasiones que la libertad de expresión constituye “una condición de existencia del pluralismo político”, como es en la sentencia del Tribunal Constitucional (STC) 6/1981⁴⁴⁷, determinando que:

⁴⁴⁴ Constitución Española, publicada en BOE núm. 311, de 29 de diciembre de 1978, vigente desde ese mismo día. Última revisión vigente desde el día 17 de Febrero de 2024 consagra las libertades fundamentales de expresión e información (párrafo 1, apartados a) y d), respectivamente), libertades no siempre fácilmente distinguibles, “pero que es necesario matizar para hacer plenamente operativos los mandatos constitucionales; de esta forma, la libertad de expresión hace referencia a la libertad para comunicar pensamientos, ideas, opiniones por cualquier medio de difusión ya sea de carácter general o más restringido (pasquines...), aunque se garantice una especial protección en el primer caso. Por su parte, la libertad de información se refiere a la comunicación de hechos mediante cualquier medio de difusión general, esto es la libertad de expresión conlleva un matiz subjetivo, mientras que libertad de la información contiene un significado que pretende ser objetivo. Evidentemente expresión e información con frecuencia no se dan separados, sino, por el contrario, unidos puesto que con las noticias es frecuente intercalar opiniones propias del informador” (consultado el día 7 de diciembre de 2025, a las 13:20 horas, en <https://app.congreso.es/consti/constitucion/indice/sinopsis/sinopsis.jsp?art=20&tipo=2>).

⁴⁴⁵ La Constitución de 1978 introdujo por primera vez en España un reconocimiento amplio y garantista de las libertades informativas, como reacción a décadas de censura y control estatal ejercidos durante el régimen franquista.

⁴⁴⁶ *Texas v. Johnson*, 491 U.S. 397 (1989), Corte Suprema de los Estados Unidos sostuvo que la quema de la bandera como protesta constituye “conducta expresiva” protegida por la Primera Enmienda, declarando inconstitucional la ley texana que penalizaba la profanación de símbolos nacionales.

⁴⁴⁷ STC, 16 de marzo de 1981, de la Sala Segunda (BOE, núm. 89, de 14 de abril de 1981), ECLI:ES:TC:1981:6, versa sobre una resolución de recurso de amparo en la que el Tribunal Constitucional protegió el derecho fundamental a la libertad de prensa. Concretamente, anuló el acuerdo

“La complejidad de este prolijo enunciado revela, según nos parece, que el legislador constituyente reconoce a la libertad de expresión y, específicamente, a la libertad de prensa, un aspecto institucional, además de su aspecto subjetivo; aspecto institucional que tiende a realizar en el plano de la información, el pluralismo político que el art. 1 proclama como uno de los valores superiores del ordenamiento jurídico y que requiere, en este campo concreto, la adopción de medidas correctoras de la desigualdad, en la confrontación de las fuerzas sociales, a tenor del mandato genérico del art. 9.2, lo que puede conseguirse mediante el uso adecuado de los medios de comunicación social del Estado”⁴⁴⁸.

Por tanto, no solo protege un derecho individual a comunicar ideas, sino que cumple una función institucional al garantizar que la sociedad pueda formarse una opinión propia libremente.

2.1. Contenido del derecho de libertad de expresión e información

La titularidad de la libertad de expresión es universal: pertenece a todas las personas por el simple hecho de serlo, sin distinción ninguna. Su ejercicio, por tanto, no depende de la nacionalidad ni de la condición profesional del emisor, sino de la capacidad que posea para comunicar ideas u opiniones.

Respecto al contenido del artículo 20 de la Constitución Española comprende diversas manifestaciones que, aunque relacionadas, poseen características y límites propios y diferentes: en primer lugar, haciendo referencia a la libertad de expresión en sentido estricto, que queda regulada en el art. 20.1. a), protege la emisión de opiniones, juicios de valor, creencias y pensamientos personales. Según la STC 51/1989, este derecho garantiza la manifestación de ideas subjetivas sin exigir veracidad ni objetividad, puesto que se trata de percepciones individuales del que las emite⁴⁴⁹. En este sentido, constituye la manifestación más amplia y esencial del precepto.

del Organismo Autónomo Medios de Comunicación Social del Estado que había suspendido la publicación de los diarios “La Voz de España y Unidad”, reafirmando que la censura previa sobre publicaciones vulnera los derechos fundamentales reconocidos en la CE.

⁴⁴⁸ STC, 16 de marzo de 1981, de la Sala Segunda (BOE, núm. 89, de 14 de abril de 1981), ECLI:ES:TC:1981:6. Trata de una resolución de un recurso de amparo en la que el Tribunal Constitucional protegió el derecho fundamental a la libertad de prensa. Concretamente, anuló el acuerdo del Organismo Autónomo Medios de Comunicación Social del Estado que había suspendido la publicación de los diarios “La Voz de España y Unidad”, reafirmando que la censura previa sobre publicaciones vulnera los derechos fundamentales reconocidos en la Constitución Española.

⁴⁴⁹ STC, 22 de febrero de 1989, de la Sala Segunda (BOE, núm. 62, de 14 de marzo de 1989), ECLI:ES:TC:1989:51, reafirma que el derecho a manifestar ideas subjetivas no exige veracidad ni objetividad.

Por otro lado, en el apartado 1. b) del citado artículo 20 CE, queda prevista la libertad de creación artística, literaria y científica, amparando la producción de obras de carácter creativo y contempla la llamada “excepción de ficción”, que reconoce que el receptor es consciente de que la obra no refleja hechos reales. Esta excepción, como se percibe, refuerza la protección frente a posibles limitaciones o censuras.

A su vez, la libertad de cátedra, establecida en el apartado 1. c) del precepto, que garantiza al profesorado la autonomía para exponer sus ideas y convicciones en el ámbito docente e investigador. La STC 217/1999 destaca que esta libertad se hace especialmente notoria en el ámbito universitario, donde el docente no solo imparte sus conocimientos sobre las materias, sino que también investiga, publica y participa activamente en la creación del pensamiento científico⁴⁵⁰.

No obstante, la libertad de cátedra no es absoluta, ya que debe ejercerse dentro de los límites constitucionales y legales, respetando otros derechos fundamentales y contribuyendo al orden público y al desarrollo responsable del proceso educativo.

Por último, el derecho a comunicar y recibir información veraz, regulado en el art. 20.1. d), donde se distingue del derecho de expresión porque protege la transmisión de hechos objetivos y no opiniones. Tiene una relevancia especial en el ámbito periodístico, ya que garantiza que los ciudadanos puedan acceder a información fiable y contrastada, lo que resulta esencial para la formación de una opinión pública informada y el correcto funcionamiento de la democracia, como se ha abordado anteriormente. El principal titular de esta prerrogativa es el periodista, quien debe cumplir con un deber de diligencia en la verificación de los hechos antes de difundirlos.

La jurisprudencia del TC, en particular la STC 240/1992, establece que la veracidad exigida no implica una obligación de transmitir la verdad absoluta, sino que se refiere a la “apariencia de verdad”⁴⁵¹, entendida como la credibilidad de la información con la protección de otros derechos fundamentales, como el honor o la reputación, al tiempo que reconoce la función social del periodismo como instrumento para garantizar la transparencia ante los ciudadanos.

⁴⁵⁰ STC, 29 de noviembre de 1999, de la Sala Primera, (BOE, núm. 310, de 28 de diciembre de 1999), ECLI:ES:TC:1999:217, además de resolver sobre derechos laborales, resalta la relevancia de la libertad de expresión en contextos profesionales, destacando su especial importancia en el ámbito universitario, donde el docente ejerce no solo la enseñanza como bien se ha puntualizado

⁴⁵¹ STC 240/1992, de 21 de diciembre, BOE núm. 17, de 20 de enero de 1993, ECLI:ES:TC:1992:240. La sentencia establece que la veracidad exigida a los medios de comunicación no implica transmitir la verdad absoluta, sino la “apariencia de verdad”, entendida como la credibilidad de la información, siempre respetando derechos fundamentales como el honor, la intimidad y la reputación, y reconociendo al periodismo una función social orientada a garantizar la transparencia.

2.2. Límites y conflictos con otros derechos

Ningún derecho es fundamental es ilimitado y, por tanto, la libertad de expresión e información encuentra sus fronteras en los derechos al honor, la intimidad personal y familiar y la propia imagen, previstos en el artículo 18 CE. Para resolver los conflictos entre ambos, el TC ha desarrollado la llamada doctrina de la posición preferente remitiéndonos a la STC 240/1992, al concretar que:

*“Según los criterios que se han ido perfilando en la jurisprudencia constitucional, esa confrontación de derechos ha de llevarse a cabo teniendo en cuenta la posición prevalente, que no jerárquica que sobre los derechos denominados de la personalidad del art. 18 CE ostenta el derecho a la libertad de información del art. 20.1 d) CE”*⁴⁵².

Como apreciamos, esta prevalencia matizada del artículo 20.1.d) CE, otorga una primacía inicial a dichas libertades consagradas, dada la función esencial en la formación de una opinión pública, sin olvidar que esta prioridad no goza de carácter absoluto, cuyas exigencias conllevan que la información transmitida sea “veraz” y esté referida a asuntos de “relevancia pública”, por ser de interés general derivado de los asuntos tratados en ellas o por la relevancia social de las personas intervinientes, “contribuyendo, en consecuencia, a la formación de la opinión pública”⁴⁵³.

El análisis de la doctrina mencionada requiere atender a tres elementos fundamentales: en primer lugar, el sujeto activo, ya que ciertas personas gozan de una protección reforzada al ejercer estas libertades, como los periodistas o los políticos, debido a su papel en la difusión de información y en el debate público; en segundo lugar, el sujeto pasivo, cuya protección del honor e intimidad varía según la condición pública

⁴⁵² Ver nota 16, en referencia a la STC 240/1992, los criterios que se han ido perfilando en la jurisprudencia constitucional, esa confrontación de derechos ha de llevarse a cabo teniendo en cuenta la posición prevalente, que no jerárquica que sobre los derechos denominados de la personalidad del art. 18 C.E. ostenta el derecho a la libertad de información del art. 20.1 d) C.E., en razón de su doble carácter de libertad individual y garantía institucional de una opinión pública indisolublemente unida al pluralismo político dentro de un Estado democrático, siempre que la información transmitida sea veraz y esté referida a asuntos de relevancia pública que son de interés general por las materias a que se refiere y por las personas que en ellos intervienen; contribuyendo, en consecuencia, a la formación de la opinión pública (SSTC 104/1986, fundamento jurídico 5º; 171 y 172/1990, fundamentos jurídicos 5º y 2º, respectivamente; STC 40/1992, fundamento jurídico 1º, 85/1992, fundamento jurídico 4º) y "alcanzando entonces su máximo nivel de eficacia justificadora frente al derecho al honor, el cual se debilita, proporcionalmente, como límite externo de las libertades de expresión e información" (STC 107/1988, fundamento jurídico 2º).

⁴⁵³ STC 30/1982, de 1 de junio, BOE núm. 153, de 28 de junio de 1982, ECLI:ES:TC:1982:30, declaró que las resoluciones impugnadas vulneraron los derechos fundamentales de los recurrentes y subrayó la doctrina de la posición preferente, que concede una primacía inicial a las libertades de expresión e información.

o privada de la persona afectada. De este modo, los personajes públicos o aquellos que han decidido exponer su vida de manera voluntaria, como políticos, artistas o celebridades, cuentan con una esfera privada mucho más limitada; y, por último, el contenido de la información, donde la relevancia pública y la proporcionalidad son criterios determinantes. Cuando la información contribuye a un debate de interés general, prevalece la libertad de información, mientras que, si versa sobre aspectos estrictamente privados sin transcendencia pública, se impone el derecho al honor o a la intimidad.

Un caso destacadísimo, por la repercusión política y mediática, fue la condena del Tribunal Europeo de Derechos Humanos a España por el asunto Otegi Mondragón c. España⁴⁵⁴, donde la Sentencia, trata sobre la condena penal impuesta a Arnaldo Otegi por el delito de injurias graves al Rey (art. 490.3 Código Penal) por unas declaraciones realizadas en 2003 en las que vinculaba al Rey con la responsabilidad sobre presuntos malos tratos cometidos por la Guardia Civil en la operación contra el diario “Egunkaria”. El TEDH resolvió que sí hubo violación a su derecho de libertad de expresión.

2.3. Requisitos de la libertad de información: veracidad y diligencia

La veracidad, como se ha señalado anteriormente, no exige corresponderse de forma absoluta con la realidad, sino que necesita de un soporte objetivo suficientemente contrastado. Por ese motivo, el TC defiende que quien informa debe actuar con diligencia profesional: comprobar y contrastar las fuentes, verificar los datos y actuar de buena fe. Básicamente, el informador debe tomar precauciones razonables para no difundir hechos falsos o insuficientemente contrastados. La STC 240/1992, antes mencionada, afirmaba que cuando este deber se incumple, el emisor pierde la cobertura constitucional del art. 20.1. d), lo que abre la puerta a la exigencia de responsabilidades civiles o penales en defensa de los derechos que han sido afectados, especialmente el del honor.

Junto a ello, la jurisprudencia ha desarrollado la denominada doctrina del “reportaje neutral”, que ofrece protección al profesional que se limita a reproducir de forma fiel y objetiva las declaraciones obtenidas de terceros. Esta protección o

⁴⁵⁴ STEDH caso Otegi Mondragón contra España, de 15 de marzo de 2011, ECLI:UE:TEDH:2011:2034, demanda número 2034/07, acordó que las injurias graves a la Corona son las únicas castigadas con una pena de prisión (de seis meses a dos años), mientras que para el derecho común y las otras Instituciones, las injurias graves se castigan con una pena de multa. Estas disposiciones confieren pues a la Corona “un privilegio exorbitante que no podría reconciliarse con la práctica y las concepciones políticas de hoy” (Colombani y demás, antes citado, § 68). El demandante se refiere a las evoluciones legislativas ocurridas en los Estados miembros del Consejo de Europa, donde la tendencia mayoritaria es someter al derecho común los ataques dirigidos contra el soberano. Así pues, la incriminación de las injurias al Rey no sería una medida necesaria en una sociedad democrática, tanto más que la incriminación de difamación e injuria bastaría a todo Jefe de Estado o monarca para hacer sancionar declaraciones que afecten a su honor”.

mecanismo opera siempre que se identifique claramente la fuente, no se manipule el contenido y la información tenga relevancia pública. Así, la STC 87/2025 señala que:

“La sentencia, con cita de jurisprudencia del Tribunal Supremo, enuncia los requisitos que deben darse para poder apreciar la doctrina del reportaje neutral: (i) el objeto de la noticia ha de hallarse constituido por declaraciones que imputan hechos lesivos del honor y que han de ser por sí mismas noticia y ponerse en boca de personas determinadas; (ii) el medio informativo ha de ser mero transmisor de tales declaraciones y (iii) la veracidad exigible se limita a la verdad objetiva de la existencia de la declaración [STS 231/2020, de 2 de junio (ECLI:ES:TS:2020:1525)]. También es exigible la ausencia de indicios racionales de falsedad evidente de lo transcrito para evitar que el reportaje neutro sirva indebidamente a la divulgación de simples rumores o insidias [SSTS 57/2012, de 13 de febrero (ECLI:ES:TS:2012:550), y 617/2016, de 10 de octubre (ECLI:ES:TS:2016:4544)]”⁴⁵⁵.

En estos supuestos, la responsabilidad recae sobre el autor original de las declaraciones y no sobre quien las difunde, incluso si esas resultan inexactas, polémicas o lesivas para un tercero. Esta doctrina responde a la función mediadora de los medios de comunicación en una sociedad democrática⁴⁵⁶.

2.4. Garantías constitucionales

Como se viene analizando, el mismo artículo 20 CE incorpora una serie de garantías institucionales destinadas a asegurar el ejercicio real y efectivo de las libertades de expresión e información.

El apartado 2 consagra la prohibición absoluta de la censura previa, impidiendo que cualquier autoridad o poder público pueda vetar la publicación o difusión de ideas antes de que lleguen al receptor. Esta garantía fue reforzada tras la experiencia autoritaria del franquismo y constituye uno de los pilares esenciales de la comunicación libre en España.

Junto a ello, el apartado 5 del precepto establece que el secuestro de publicaciones solo puede acordarse mediante una resolución judicial motivada. Se trata de una medida excepcional que busca preservar el equilibrio entre la libertad comunicativa y

⁴⁵⁵ STC 87/2025, de 7 de abril, publicada en el BOE núm. 117, de 15 de mayo, ECLI:ES:TC:2025:87, en relación con la sentencia de la Sala de lo Civil del Tribunal Supremo que estimó, en casación, una demanda por vulneración de los derechos al honor y a la propia imagen. Supuesta vulneración del derecho a la libertad de información: difusión de un vídeo en el que aparecen imágenes de quien no es sujeto de la noticia; doctrina del reportaje neutral (STC 62/2025).

⁴⁵⁶ BELMAR TALÓN, Á., *La definición legal de periodista en la historia de España: El artículo 20 de la Constitución española de 1978 como estatuto jurídico del profesional de la información*, tesis doctoral, 2017, publicado en Dehesa, Universidad de Extremadura, p. 196.

la protección de otros derechos o bienes constitucionales, evitando que el poder público pueda limitar de forma arbitraria acceder a cualquier tipo de información.

A estas garantías, se suma la protección de la cláusula de conciencia y del secreto profesional de los periodistas, previstas en el apartado 1. d), del mismo art. 20 CE, y desarrolladas por la Ley Orgánica 2/1997, consagrando en su artículo 1, sobre esta cláusula de conciencia un “derecho constitucional de los profesionales de la información que tiene por objeto garantizar la independencia en el desempeño de su función profesional”, cuyo desarrollo se afronta por la Ley Orgánica 2/1997 de forma más amplia⁴⁵⁷. Ambos mecanismos refuerzan la independencia ideológica de los profesionales de la información y les permiten negarse a actuaciones contrarias a sus propias convicciones o creencias, así como mantener en reserva la identidad de sus fuentes, incluso frente a requerimientos judiciales, tal y como ha reconocido la STC 15/1993:

*“Por el contrario ha quedado acreditado como hecho que se trata de una carta al director publicada en la sección "Vd. opina" del Bisemanario local "Igualada", y sobre la que el recurrente asume la responsabilidad únicamente en la medida en que como director de la publicación decide no revelar su autor, en aplicación de la cláusula del secreto profesional previsto igualmente en el art. 20 de la C.E. Es evidente que los directores de los medios informativos no adquieren la misma responsabilidad por la publicación de los escritos elaborados por los profesionales que en ellos trabajan que por los contenidos de los enviados por los lectores a las secciones destinadas a recoger opiniones e informaciones en principio ajenas a la línea editorial de los medios”*⁴⁵⁸.

Asimismo, es necesario destacar la Ley Orgánica 2/1984 reguladora del derecho de rectificación⁴⁵⁹, permitiendo a cualquier persona solicitar la corrección de informaciones inexactas o mal redactadas que puedan causarle un perjuicio. Esta cuestión

⁴⁵⁷ Ley Orgánica 2/1997, de 19 de junio, reguladora de la cláusula de conciencia de los profesionales de la información (BOE núm. 147, de 20 de junio de 1997, entrada en vigor el 21 de junio de 1997), en su exposición de motivos se consagra que: “La Constitución Española de 1978 ha introducido en su parte dogmática el reconocimiento del derecho de los profesionales de la información a la cláusula de conciencia. Y si bien es cierto que este derecho estaba ya reconocido por la legislación ordinaria, con diferentes grados de garantía, en diversos estados democráticos, la aportación del texto constitucional español ha sido la de integrarlo como elemento constitutivo del derecho fundamental a recibir y comunicar información”.

⁴⁵⁸ STC 15/1993, de 18 de enero, BOE núm. 37, de 12 de febrero de 1993, ECLI:ES:TC:1993:15, reafirmó en esta sentencia que la protección de la persona que informa incluye la facultad de no revelar las fuentes cuando ello resulte necesario para el ejercicio libre de la función informativa, especialmente en asuntos de relevancia pública.

⁴⁵⁹ Ley Orgánica 2/1984, de 26 de marzo, reguladora del derecho de rectificación, BOE núm. 74, de 27 de marzo, entrada en vigor el 16 de abril de 1984, previniendo sus art. 1, que tienen legitimación para el ejercicio de tal derecho “toda persona natural o jurídica” y, en su art. 2, que se “ejercitará mediante la remisión del escrito de rectificación al director del medio de comunicación”.

va dirigida principalmente al receptor de la información, que complementa al principio de veracidad.

Por último, regresando al art. 20, concretamente, su apartado 3, impone la existencia de medios de comunicación públicos sometidos a control parlamentario, lo que intenta asegurar el pluralismo y el servicio a la ciudadanía. Estos medios deben, además, reflejar la diversidad lingüística y cultural del Estado y, de este modo, ayudar a cohesionar la sociedad y asegurar que todas las personas tengan acceso a una información plural y equilibrada.

2.5. Referencia a la dimensión internacional de la libertad de expresión

Aunque se viene analizando el panorama español en cuanto a la libertad de expresión y de información, cabe hacer una mención a su sólido respaldo en el ámbito internacional. Por una parte, el Convenio Europeo de Derechos Humanos⁴⁶⁰ reconoce en su art. 10 el derecho a la libertad de expresión, incluyendo la libertad de opinión y de recibir o difundir información sin interferencia de autoridades ni fronteras, pero admite que los Estados puedan establecer ciertas autorizaciones previas para empresas de radiodifusión, televisión o cinematografía.

De manera complementaria, el art. 19 del Pacto Internacional de Derechos Civiles y Políticos⁴⁶¹ reafirma que toda persona tiene derechos a la libertad de expresión, abarcando la libertad de buscar, recibir y difundir información de todo tipo, por cualquier medio y sin restricciones de fronteras. Este art., al igual que el 10 del CEDH, establece que el ejercicio de dicha facultad conlleva, a su vez, deberes y responsabilidades, permitiendo así que los Estados impongan ciertas restricciones cuando sean necesarias para proteger la seguridad nacional, la integridad territorial, la seguridad pública o la autoridad del poder judicial, entre otras.

La ratificación en España tanto del Convenio Europeo de Derechos Humanos como del Pacto Internacional de Derechos Civiles y Políticos ha reforzado la protección nacional de las libertades abordadas y ha servido de guía jurisprudencial para

⁴⁶⁰ CONSEJO DE EUROPA (1950), *Convenio para la protección de los derechos humanos y de las libertades fundamentales (CEDH)*, redactado en Roma el 4 de noviembre de 1950. Ratificado por España mediante BOE núm. 243, de 10 de octubre de 1979. Se encarga de garantizar la libertad de expresión, con posibles restricciones para proteger derechos de terceros, seguridad, orden público, salud o moral, conteniendo su art. 10 que “toda persona tiene derecho a la libertad de expresión”, abarcando la libertad de opinión y de recibir o comunicar informaciones sin injerencias.

⁴⁶¹ NACIONES UNIDAS (1966). *Pacto Internacional de Derechos Civiles y Políticos (PIDCP)*, redactado en Nueva York el 19 de diciembre de 1966. Ratificado por España mediante BOE núm. 103, 30 de abril de 1977, reconoce la libertad de expresión y de opinión, con límites necesarios para proteger derechos de terceros y el orden público, cuyo art. 19 dispone que “nadie podrá ser molestado por causa de sus opiniones” y todas las personas tendrán derecho a la “libertad de expresión”.

delimitar sus límites, siempre en un equilibrio entre derechos individuales y el interés general. La doctrina internacional ha subrayado que estas restricciones deben ser precisas, necesarias y proporcionales, garantizando que la libertad de expresión cumpla su función esencial de promover el debate público, la transparencia y la formación de la opinión.

En este sentido, la dimensión internacional de la libertad de expresión complementa y refuerza el marco constitucional español, prestando criterios de interpretación para la aplicación de límites, proporcionando garantías en el ejercicio de la comunicación, incluyendo, como se desarrollará posteriormente su relación con las nuevas tecnologías y la IA en especial.

3. IMPACTO EN LA INTELIGENCIA ARTIFICIAL

3.1.- Funcionalidad de la inteligencia artificial confrontada a la libertad de expresión e información

Como se ha explicado anteriormente, la IA se ha convertido en una herramienta básica en los procesos comunicativos de la actualidad, puesto que transforma tanto la producción como la difusión de información o cualquier tipo de datos y, con ello, las condiciones en las que se ejerce la libertad de expresión en el entorno. Su incorporación a plataformas digitales y medios de comunicación hace más sencilla la gestión de grandes volúmenes de datos, agiliza procesos y, en muchas ocasiones, ofrece experiencias personalizadas a cada usuario⁴⁶².

Ahora bien, entre las aplicaciones más relevantes de este sistema en el ámbito comunicativo, destaca la automatización de tareas rutinarias, lo que consiste en que la tecnología sea capaz de generar contenidos básicos, clasificar mensajes, programar publicaciones o realizar interacciones a través de *chatbots* y asistentes virtuales. Todo esto en su conjunto, provoca o intenta que los profesionales puedan dedicar más tiempo a otras funciones, ya que la IA les proporciona respuestas rápidas sobre estrategias, situaciones creativas o de análisis de forma sencilla.

También, se cuenta con que este tipo de sistemas ha revolucionado la gestión de grandes volúmenes de datos, lo cual está totalmente ligado a la aplicación anterior, puesto que, al manejar tanta información en poco tiempo, puede dar resultados sobre tendencias, el comportamiento del público y anticipar dinámicas comunicativas. Con esto, se intenta dar pie a que los usuarios puedan adoptar decisiones más informadas, lo que desemboca en una participación pública mejor estructurada y con mayor capacidad de adaptación a los cambios del ámbito informativo.

⁴⁶² HERRERA ORTIZ, J. J., PEÑA AVILÉS, J. M., HERRERA VALDIVIESO, M. V., MORENO MORÁN, D. X. "La inteligencia artificial y su impacto en los medios de comunicación", *TELOS: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, volumen 26, número 1, 2024, págs. 278-296.

Además, otro elemento clave para la libertad de expresión digital en este aspecto, es la personalización de los contenidos. La capacidad que tiene la IA de ajustar recomendaciones, noticias o mensajes al perfil de cada persona permite que éstas accedan a información más relevante para sus propios intereses. Sin embargo, esta misma lógica influye en la agenda pública y condiciona el modo en el que los usuarios comparten y descubren contenidos, tanto de forma positiva como negativa.

Por otro lado, también cabe mencionar el reconocimiento de voz, el procesamiento del lenguaje humano o los sistemas de traducción automática, los cuales han supuesto una ampliación significativa en la accesibilidad comunicativa y de información. Estas cuestiones facilitan la participación de personas con barreras lingüísticas o funcionales y favorecen a su vez la integración de distintos grupos en el debate público. En este sentido, la IA actúa como un factor de expansión del derecho a la libertad de expresión y ser escuchado, especialmente en servicios públicos y entornos educativos.

Finalmente, es necesario añadir que, en el ámbito periodístico, la IA contribuye en la detección de información relevante, así como en la verificación de contenidos, siendo una herramienta capaz de rastrear miles de fuentes en tiempo real intentando detectar patrones asociados a bulos o manipulaciones que, en muchas ocasiones, no lo consigue. Por tanto, como se puede comprender, este sistema que ha evolucionado tan rápido y se ha incorporado de forma impactante en la vida real, no solo colabora en la realización de funciones, intentando simplificarlas lo máximo posible, sino que también transforma procesos comunicativos y reconfigura las condiciones que conlleva el debate público⁴⁶³.

3.2.- Nuevas formas de participación y debate público condicionadas por la inteligencia artificial

La aparición de estos sistemas ha provocado que se reduzcan las barreras de acceso a la creación y difusión de contenidos, permitiendo que cualquier persona produzca textos, imágenes, audios y vídeos de calidad profesional, prácticamente convirtiéndolos en realidad. Sin embargo, este fenómeno no está exento de ambigüedades: la misma tecnología que facilita la expresión puede introducir nuevas desigualdades y dependencias tecnológicas. El Programa de las Naciones Unidas para el Desarrollo (PNUD) señala que “la IA ha empoderado, y puede seguir empoderando, a las

⁴⁶³ SANAHUJA SANAHUJA, R., & LÓPEZ RABADÁN, P., “Ámbitos de aplicación periodística de la Inteligencia Artificial. Razón y Palabra. Mapa conceptual, funciones profesionales y tendencias en desarrollo en el contexto de la pandemia global de la Covid-19”, *Vol. 25, número 112. 2021, Comunicación Política y Tecnologías Aplicadas: desinformación, noticias falsas y verificación de hechos*, volumen 25, número 112, 2021, Proyectos de Investigación Científica y Desarrollo Tecnológico “Ética discursiva y Democracia ante los retos de la Inteligencia Artificial”, PID2019-109078RB-C21 y PID2019-109078RB-C22, financiados por el Ministerio de Ciencia, Innovación y Universidades, págs. 432-449.

personas, mejorando el avance de los derechos humanos y el derecho a buscar, recibir y difundir información”⁴⁶⁴, aunque este empoderamiento está condicionado por quién controla las infraestructuras técnicas y los datos que las alimentan.

Toda esta capacidad tecnológica ha ampliado la esfera pública digital, haciendo más fácil la participación y dando voz a grupos de personas que tradicionalmente eran marginados. No obstante, la ampliación del acceso no implica una ampliación real del pluralismo: la arquitectura algorítmica que posee la IA prioriza ciertos contenidos frente a otros, y puede reforzar dinámicas de atención que distorsionan el verdadero pluralismo del que se viene hablando y que se intenta conseguir.

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) en este caso, subraya que las nuevas tecnologías, en especial la IA, han revolucionado el discurso político al permitir que los ciudadanos actúen como creadores y distribuidores de información sin las limitaciones propias y tradicionales de los medios de comunicación. Pero esta liberación del discurso convive con mecanismos de filtraciones y recomendación que se escapan del control democrático y que pueden condicionar qué voces llegan efectivamente al debate público⁴⁶⁵. De este modo, la IA contribuye a generar nuevas formas de participación política, pero también introduce nuevas intermediaciones que no siempre son visibles por la población. Desde el punto de vista constitucional, esta evolución se conecta con la interpretación expansiva que el TC ha dado al art. 20 CE como “piedra angular del sistema democrático”, reconociendo:

*“El art. 20 de la Constitución, en sus distintos apartados, garantiza el mantenimiento de una comunicación pública libre, sin la cual quedarían vaciados de contenido real otros derechos que la Constitución consagra, reducidas a formas huecas las instituciones representativas y absolutamente falseado el principio de legitimidad democrática que enuncia el art. 1.2 de la Constitución, y que es la base de toda nuestra ordenación jurídico-política”*⁴⁶⁶.

⁴⁶⁴ Programa de Naciones Unidas, *Integridad de la información para las instituciones y los procesos electorales*, p. 11, 2023.

⁴⁶⁵ UNESCO, “Recomendación sobre la ética de la inteligencia artificial”, 2022, consultado el día 12-11-2025, a las 19:25 horas, en <https://www.unesco.org/es/articles/recomendacion-sobre-la-etica-de-la-inteligencia-artificial>

⁴⁶⁶ STC 6/1981, de 16 de marzo, BOE, número 89, de 14 de abril, ECLI:ES:TC:1981:6, en esta sentencia el TC subraya el papel estructural de las libertades del art. 20 CE dentro del sistema democrático procurando preservar “esta comunicación pública libre sin la cual no hay sociedad libre ni, por tanto, soberanía popular, exige la garantía de ciertos derechos fundamentales comunes a todos los ciudadanos, y la interdicción con carácter general de determinadas actuaciones del poder (*verbi gratia* las prohibidas en los apartados 2 y 5 del mismo art. 20), pero también una especial consideración a los medios que aseguran la comunicación social y, en razón de ello, a quienes profesionalmente los sirven”.

Las plataformas digitales y los sistemas algorítmicos actualmente son un epicentro de información, pero no se han acompañado de garantías. Como advierte *Balkin*, estas infraestructuras deben ser entendidas como “nuevos operadores estructurales de la libertad de expresión”⁴⁶⁷, cuyo diseño afecta directamente a la circulación del discurso público. En este sentido, no solo amplifican voces, sino que también pueden moldearlas, limitarlas o invisibilizarlas.

Paralelamente, *Zuboff* destaca que, en la era digital, el control de la información se convierte en una dimensión esencial del poder, especialmente cuando plataformas y la IA en sí influyen en la visibilidad y jerarquización de contenidos⁴⁶⁸. Así, el acceso a la expresión deja de depender exclusivamente del marco constitucional y pasa a depender, en gran medida, de decisiones privadas y algoritmos.

La IA a su vez, desempeña un papel creciente en los procesos electorales. Los organismos de gestión electoral utilizan herramientas para mejorar la difusión de información al votante e intentar combatir la desinformación⁴⁶⁹. Sin embargo, la introducción de sistemas como la IA también abre la puerta a una mayor diferencia entre actores políticos: quienes poseen más datos y capacidades tecnológicas, pueden influir de forma más precisa (y menos visible) en el comportamiento político del electorado. Asimismo, el trato personalizado que se analizó anteriormente puede fragmentar la experiencia democrática, sustituyendo un debate común por una multiplicidad de discursos dirigidos individualmente, lo que dificulta la formación de una opinión verdaderamente compartida y con diversidad de información.

En este contexto, resulta especialmente relevante la STS 1119/2025, se ha reconocido el derecho de la ciudadanía a acceder al código fuente de los sistemas algorítmicos utilizados por la Administración cuando estos inciden en derechos fundamentales o sociales:

“En la misma materia de reutilización, el artículo 157.2 de la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público contempla la oportunidad de declarar determinadas aplicaciones de las Administraciones como de fuentes abiertas para aumentar

⁴⁶⁷ Sobre el papel de las plataformas digitales como reguladores del discurso público y “operadores estructurales” de la libertad de expresión, véase BALKIN, Y., *Free Speech in the Algorithmic Society*, 2018, citado y analizado en UNESCO, “Libertad de expresión, inteligencia artificial y elecciones”, 2025.

⁴⁶⁸ ZUBOFF, S., *The Age of Surveillance Capitalism*, 2019, quien analiza cómo el control de la información y los algoritmos de las plataformas condicionan la visibilidad, jerarquización y acceso al discurso público; citado y analizado también en “Religación”, *Revista de Ciencias Sociales y Humanidades*, volumen 6, número 28), págs. 220-222, <https://doi.org/10.46652/rgn.v6i28.811>

⁴⁶⁹ Informe de Naciones Unidas sobre la integridad informativa y elecciones, “infórmate con responsabilidad, participa con integridad” sobre las elecciones generales en Chile, 2025, consultado el día 18-11-2025, a las 20:18 horas, en <https://www.undp.org/es/chile/ParticipaConIntegridadPNUD>

la transparencia en su funcionamiento o fomentar la incorporación de los ciudadanos a la Sociedad de la información”⁴⁷⁰.

Esta resolución supone un avance en materia de transparencia algorítmica y confirma que el funcionamiento interno de estos sistemas no puede quedar fuera del control democrático. En consecuencia, aunque la IA amplifica la expresión individual de los sujetos y transforma la esfera del espacio público democrático, lo hace dentro de un modelo comunicativo en el que el poder está crecientemente concentrado. Si estos sistemas se desarrollaran conforme a principios de transparencia, supervisión y respeto a los derechos fundamentales, pueden consolidar un debate público más plural e inclusivo.

Pero si la gobernanza de la IA continúa dependiendo de actores privados con intereses económicos propios, la libertad de expresión puede correr el riesgo de verse condicionada por cuestiones externas al interés de la democracia, con gran influencia de los algoritmos y de la desinformación.

3.3.- Amenazas emergentes de la inteligencia artificial sobre la libertad de expresión e información

A pesar de su enorme potencial transformador, la IA plantea riesgos muy significativos. Como señala Barrio Andrés, el Derecho está obligado a “elaborar una regulación avanzada que pueda impulsar el desenvolvimiento de la robótica y asegurarle un desarrollo congruente con los valores propios del ordenamiento jurídico”⁴⁷¹. La generalización de este tipo de tecnologías, las cuales están presentes tanto en motores de búsqueda como en redes sociales y plataformas, demuestra que este sistema ya forma parte de manera sustancial en la vida cotidiana de la mayoría de los seres humanos, con beneficios evidentes, pero con riesgos cada vez con mayor repercusión.

Por ello, esta disciplina no puede abordarse únicamente desde una perspectiva informática o desde la ingeniería, sino que exige un análisis jurídico profundo. La

⁴⁷⁰ STS 1119/2025, de 11 de septiembre, ECLI:ES:TS:2025:3826, concreta que la denegación a un particular del acceso al código fuente no puede resultar contraria a las previsiones del artículo 157 de la Ley 40/2015 porque la reutilización de sistemas y aplicaciones que regula este precepto se limita al ámbito de las relaciones entre Administraciones públicas, y porque la declaración como de fuentes abiertas se contempla como una mera posibilidad, condicionada a que de ello se derive una mayor transparencia en el funcionamiento de la Administración pública o se fomente con ello la incorporación de los ciudadanos a la sociedad de la información.

⁴⁷¹ BARRIO ANDRÉS, M., *Derecho de los robots*, 2018, citado en SÁNCHEZ ESPARZA, M., Y PALELLA STRACUZZI, S., *Inteligencia artificial y nuevas fronteras: la transformación de la comunicación, de los derechos humanos, la educación y la innovación organizacional*, Tirant lo Blanch, Valencia, 2024, p. 63.

Comisión Europea, en su Comunicación (COM, 2018, 237 final)⁴⁷², ya subrayaba la necesidad de adquirir directrices basadas en la Carta de Derechos Fundamentales de la Unión Europea⁴⁷³, centradas principalmente en la seguridad, la transparencia, la equidad, la protección de datos y el impacto en la dignidad humana. En todo caso, como recuerda la doctrina inspirada en Asimov, el control de estas tecnologías debe permanecer siempre en manos del ser humano⁴⁷⁴, pero desde una visión integrada en la democracia y no en sujetos privados con intereses económicos personales, como ya se abordó anteriormente.

La IA, por lo tanto, debe concebirse para mejorar la calidad de vida y no para subordinarnos a decisiones automatizadas que puedan coartar nuestra libertad o limitar nuestras capacidades de creatividad, pensamiento crítico e iniciativa personal.

Desde esta perspectiva, los avances tecnológicos no pueden comprometer los derechos y libertades de las personas. Sin embargo, este sistema puede desplegarse en ámbitos tan diversos como la industria, la educación, la salud, los servicios o la política, pero siempre bajo el principio rector de que la tecnología debe estar al servicio del ser humano, y no al revés. Precisamente, cuando esta tecnología se integra profundamente en el ecosistema informativo y comunicativo, surgen riesgos específicos para la libertad de expresión y para el pluralismo democrático protegidos por el artículo 20.1 CE.

Para comenzar a desarrollarlos, uno de los principales riesgos es la censura automatizada, entendida como la supresión o limitación de contenidos mediante decisiones automatizadas, adoptadas por sistemas de moderación sin supervisión humana adecuada. La UNESCO advierte que estos filtros pueden eliminar discursos legítimos (sobre todo de las minorías) por errores sistemáticos, lo que puede desembocar en la pérdida de diversidad informativa y autocensura (la inhibición de expresarse por temor a bloqueo o sanción). La falta de transparencia de los algoritmos dificulta su control público y plantea dudas sobre si cumplen con el requisito de previsibilidad establecido por el TEDH en el caso *Delfi AS v. Estonia*⁴⁷⁵.

⁴⁷² COMISIÓN EUROPEA, *Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones: Inteligencia artificial para Europa*, 2018, p. 237 final, Bruselas, 25 de abril de 2018.

⁴⁷³ Carta de los Derechos Fundamentales de la Unión Europea, Diario Oficial de las Comunidades Europeas, C 364/1, 18 de diciembre de 2000.

⁴⁷⁴ Sobre la necesidad de que el control sobre la inteligencia artificial permanezca siempre en manos del ser humano, y desde una visión integrada en la democracia y no subordinada a intereses privados, SÁNCHEZ ESPARZA, M., Y PALELLA STRACUZZI, S., *Inteligencia artificial y nuevas fronteras: la transformación de la comunicación, de los derechos humanos, la educación y la innovación organizacional*, Tirant lo Blanch, Valencia, 2024.

⁴⁷⁵ La sentencia *Delfi AS v. Estonia*, del TEDH, emitida el 16 de junio de 2015, ECLI:UE:TEDH:2015:6456, confirmó que un portal de noticias puede ser responsable por los comentarios ofensivos publicados por terceros en su sitio cuando no elimina rápidamente los mensajes

Por otro lado, la IA permite generar *deepfakes*, es decir, vídeos, audios o imágenes artificiales capaces de suplantar rostros y voces con gran realismo. El PNUD ha documentado que en muchas ocasiones se utiliza para manipular campañas políticas, desacreditar a candidatos y fomentar la violencia de género digital. Por la proliferación de contenidos falsos puede causar una crisis de confianza pública y dificultar que las personas distingan entre lo real y lo manipulado.

Como ejemplo reciente de los riesgos conexos con la manipulación de contenidos, la Agencia Española de Protección de Datos (AEPD) ha confirmado la primera sanción en Europa por crear y difundir *deepfakes* de contenido sexual, imponiendo una multa de 2.000 euros al responsable de generar imágenes falsas en las que se superponía el rostro de una menor sobre un cuerpo desnudo⁴⁷⁶.

A raíz de esto, se puede hablar de la facilidad con la que la IA puede difundir bulos o noticias falsas de forma masiva, contenidos que son creados principalmente para falsear, con técnicas como el *microtargeting*, *astroturting* o *fire-hosing*⁴⁷⁷. Estudios realizados por el Instituto Tecnológico de *Massachusetts*, han demostrado que actualmente la desinformación tiende a difundirse mucho más rápido que la información veraz. Aquí se encargan de analizar millones de tuits y publicaciones en redes sociales, identificando patrones de viralidad y confirmando que los contenidos falsos generan más interacciones y se propagan de mucho más rápida que los contenidos verdaderos.

Asimismo, es necesario hacer referencia a los sesgos algorítmicos, los cuales son distorsiones sistemáticas que surgen cuando los datos de entrenamiento de los sistemas de IA son incompletos, desactualizados o culturalmente sesgados, es decir, que los datos de los que se nutre la IA son antiguos, contienen prejuicios o no tienen la suficiente información⁴⁷⁸. Cuando ocurre esto, el sistema puede dar prioridad a unas informaciones y ocultar otras, o mostrar opiniones sesgadas sobre personas o grupos,

ilegales. El Tribunal consideró que esta responsabilidad no violaba la libertad de expresión, dada la necesidad de proteger el honor y derechos de las personas afectadas y la posición económica del portal para controlar los comentarios.

⁴⁷⁶ AEPD confirma la resolución sancionadora, relativa a la creación y difusión de imágenes sexuales falsas generadas mediante inteligencia artificial (*deepfake*), con multa de 2.000 euros. Información difundida por “El País” y otros medios a partir de la resolución pública, el expediente se refiere a los hechos ocurridos en septiembre de 2023 en Almendralejo (Badajoz).

⁴⁷⁷ Las tres técnicas comparten el objetivo de influir en la opinión pública y manipular la percepción social mediante el uso estratégico de información o mensajes. Todas recurren a medios digitales para segmentar o simular apoyo con contenido diseñado para lograr un impacto político o social, afectando así la libertad de información.

⁴⁷⁸ UNESCO, “Informe sobre el sesgo algorítmico y la inteligencia artificial”, p. 17, 2022, destaca que los sistemas de IA reproducen sesgos cuando se entrenan con datos incompletos o parciales.

repitiendo estereotipos de género, raza o nivel social. Por lo tanto, si la IA posee o consigue información parcial o injusta, dará soluciones o respuestas del mismo tipo.

Cabe destacar la alerta proveniente de diversos organismos internacionales sobre la concentración del poder informativo en manos de un reducido número de plataformas que controlan los algoritmos de recomendación y, en consecuencia, la visibilidad de los discursos. Esta situación permite moldear el debate público y afectar al pluralismo informativo que, como se ha venido analizando, queda protegido en el art. 20.1.d) CE. Estos entornos favorecen la formación de burbujas informativas o cámaras de eco que refuerzan prejuicios, limitan la exposición de otros puntos de vista diversos (al solo mostrarte o informarte sobre tu propio algoritmo) y empobrecen la manifestación de diferentes perspectivas para formar la propia ideología o creencia.

Pero, además de los riesgos informativos y democráticos, la IA también puede generar consecuencias graves cuando se utiliza en profesiones reguladas, como la abogacía, la medicina o la psicología. Un ejemplo reciente de ello lo constituye el Auto del Tribunal Superior de Justicia de Navarra (ATSJ Navarra núm. 2/2024)⁴⁷⁹, donde un abogado incorporó un artículo del Código Penal colombiano en una querrela tras usar ChatGPT. El tribunal en este caso destacó que, aunque la IA puede ser una herramienta útil para aumentar la eficiencia, la responsabilidad de verificación recae siempre en el profesional, subrayando los riesgos de las denominadas “alucinaciones legales” (información errónea o inventada por los sistemas generativos) y la necesidad de una supervisión humana estricta para garantizar la exactitud de los documentos jurídicos. Esta situación pone sobre la mesa que, incluso sin malas intenciones, el uso descuidado de la IA puede comprometer derechos, principios de buena fe y la integridad de las decisiones de los profesionales.

Con todo esto en su conjunto, resulta fundamental indagar sobre cómo estas dinámicas afectan a colectivos especialmente expuestos, donde destaca la juventud española, cuyo consumo intensivo tanto de redes sociales como de plataformas basados en la IA influye directamente en su comportamiento, en su relación con la información y en su formación de opinión, derivando de manera directa en su libertad y modo de expresión.

3.4.- El dominio de la inteligencia artificial en el comportamiento e información recibida por jóvenes

La expansión de la IA entre adolescentes y jóvenes está reconfigurando la manera en la que este grupo accede a la información, construye opiniones, gestiona

⁴⁷⁹ Auto TSJ Navarra núm. 2/2024, de fecha 4 de septiembre, ECLI:ES:TSJNA:2024:38, se abre la discusión sobre el papel de la IA en profesiones reguladas y recuerda que el componente humano en la verificación y supervisión sigue siendo esencial.

conflictos personales y expresa su identidad. Este fenómeno constituye una verdadera transformación del ecosistema comunicativo juvenil y plantea riesgos muy específicos en la formación de la personalidad⁴⁸⁰.

Aunque la mayoría de ellos piensen que este sistema no puede sustituir del todo a las personas, su uso creciente como apoyo emocional muestra una dependencia muy significativa.

El impacto psicológico del uso intensivo de este tipo de sistemas ha comenzado a manifestarse en casos graves que han llegado a los tribunales. A nivel internacional, se investigan diversas demandas que atribuyen a modelos generativos de IA parte de la responsabilidad en varios suicidios de jóvenes. Entre ellas, destaca el caso de *Adam Raine*, en Estado Unidos, un adolescente cuyos padres acusan a OpenAI de haber contribuido a su muerte tras ofrecerle el *chatbot* consejos explícitos sobre cómo suicidarse e, incluso, redactarle una nota de despedida⁴⁸¹.

Junto a esta, se han interpuesto otras acciones judiciales que denuncian daños severos en la salud mental, derivados del uso continuado del sistema. Aunque estos procedimientos siguen en fase de investigación, demuestran un riesgo muy elevado, observando la gran vulnerabilidad emocional de adolescentes y jóvenes.

A esta confianza se suma otro riesgo, lo que son las inexactitudes y errores sistemáticos de los modelos generativos, que con frecuencia ofrecen contenido inventado pero verosímil. La literatura científica ha demostrado que estas herramientas pueden proporcionar respuestas incorrectas en porcentajes muy elevados, mientras que una parte muy considerable del alumnado no es plenamente consciente de este problema. Esta cuestión se agrava si se considera que la desinformación viral y el consumo del algoritmo favorece a la difusión inmediata de discursos que polarizan y moldean la opinión juvenil sobre temas sensibles.

⁴⁸⁰ Según el informe “Así somos. El estado de la adolescencia en España” (*Plan International*), el 60% de los adolescentes considera que la IA actúa como “una segunda opinión experta e imparcial”, confiando en ella como fuente objetiva y neutra. PLAN INTERNATIONAL, “Así somos. El estado de la adolescencia en España”, que realizó una encuesta a 3.500 adolescentes de 12 a 17 años, 2025. *Plan International* es una Organización No Gubernamental fundada en 1937 que trabaja para defender y promover los derechos de la infancia y la adolescencia en todo el mundo. Del mismo modo, en el informe de “Así Hablamos. Las voces de la adolescencia”⁴⁸⁰ de la misma investigación, se revela un patrón similar: la IA funciona como consejera emocional y moral. Una de las participantes del proyecto de 15 años explicó que acudió a este sistema para resolver un conflicto con una amiga, mientras que otros chicos y chicas la utilizan para decidir compras, empezar dietas o comprender sus propias relaciones personales. Para este grupo, la IA es alguien “que lo sabe todo”, dotada de una autoridad percibida como estable y libre de juicios que las de sus entornos cercanos.

⁴⁸¹ KASSHMIR HILL, “Un adolescente tenía tendencias suicidas. ChatGPT3 fue el amigo en quien confió”, *The New York Times*, 27 de agosto de 2025, consultado el día 11-11-2025, a las 20:21 horas.

Frente a todo esto, deberá llamarse la atención de las administraciones públicas para enfatizar la urgencia de educación digital, que abarque pensamiento crítico, privacidad, ética tecnológica y ciberseguridad, ya que la IA se ha convertido en un recurso habitual para la solución de cuestiones de cualquier ámbito en la vida de los adolescentes. Su influencia no es neutra: condiciona la libertad de expresión, la formación de creencias y la vivencia de las relaciones afectivas.

4. LA RESPONSABILIDAD CIVIL DERIVADA DEL MAL USO DE LA INTELIGENCIA ARTIFICIAL

La incorporación de sistemas de IA en la creación y difusión de contenidos e información plantea importantes desafíos a los fundamentos tradicionales de la responsabilidad civil. Aunque la doctrina clásica se construyó pensando en daños causados por personas o por objetos carentes de autonomía⁴⁸², hoy existen sistemas capaces de generar textos, imágenes, vídeos y afirmaciones con apariencia verosímil, que pueden perjudicar la reputación, imagen o credibilidad de individuos concretos.

En el ámbito informativo y comunicativo las consecuencias se ven reflejadas directamente sobre los derechos fundamentales del art. 20 CE: el derecho a recibir información veraz; el derecho a emitir expresiones sin ser suplantado o manipulado; el derecho a no sufrir intromisiones ilegítimas en la imagen u honor mediante contenidos falsos; y el derecho a participar en un debate público libre de falsificaciones.

El hecho de distinguir entre el causante y el responsable del daño resulta especialmente complejo. Aunque el sistema tecnológico sea el agente material, el Derecho Civil rechaza atribuir responsabilidad a la máquina: “las cosas contribuyen al daño, pero la responsabilidad debe dirigirse siempre a una persona”⁴⁸³.

Por tanto, en este contexto de libertad de expresión e información, responderán civilmente quienes crean programas sin adoptar medidas para evitar contenidos falsos; utilizan la IA para generar información que perjudica a otros sin verificar su identidad; difunden contenidos sin comprobar sus efectos; integran algoritmos de recomendación que amplían la desinformación; o incluso profesionales que, por negligencia, reproducen información generada artificialmente sin verificarla.

⁴⁸² Recuérdese el encuadre jurídico de la responsabilidad extracontractual en el Código Civil, dentro del libro cuarto “de las obligaciones y contratos”; título XVI “de las obligaciones que se contraen sin convenio”, capítulo II, “de las obligaciones que nacen de culpa o negligencia”, en sus arts. 1902 a 1910.

⁴⁸³ GÓMEZ LIGÜERRE, C., “Responsabilidad por daños causados por la inteligencia artificial”, *InDret*, pp. 9-11, 2025, consultado el día 13-11-2025, a las 20:23 horas, en <https://indret.com/wp-content/uploads/2025/01/1924.pdf>

Con todo ello en su conjunto y como ya se ha mencionado, estos sistemas modifican sustancialmente los elementos tradicionales de la responsabilidad civil. La opacidad de los algoritmos dificulta identificar qué acción concreta produjo un contenido lesivo. A esto se le suma la rapidez y viralidad de la difusión digital, que multiplica el daño. En estos supuestos, la doctrina señala que la imputación debe basarse en el deber de control, supervisión y diligencia⁴⁸⁴, de acuerdo con el estándar del artículo 1902 CC.

Además, como ya se ha visto, el daño informativo puede adoptar diferentes formas, entre las que destacan los daños a causa de las *deepfakes* o la difusión viral de información falsa. Estos daños son plenamente indemnizables conforme a la Ley Orgánica 1/1982, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen⁴⁸⁵, que establece un sistema específico de protección frente a intromisiones ilegítimas y reconoce de forma expresa la reparación del daño moral. En concreto su artículo 9.3 establece que “la existencia de perjuicio se presumirá siempre que se acredite la intromisión ilegítima”.

La respectiva indemnización será extensiva al daño moral, cuya valoración atenderá a las “circunstancias del caso y a la gravedad de la lesión efectivamente producida”, teniendo presentes a efectos valorativos la difusión o audiencia del medio a través del que se haya producido⁴⁸⁶.

Asimismo, cuando los daños producidos por sistemas inteligentes son de naturaleza física, patrimonial o moral, estos pueden reclamarse al amparo del artículo 1902 CC, que configura la responsabilidad extracontractual por culpa o negligencia. Se trata de una cláusula general que obliga a reparar el daño causado por acción u omisión, incluso cuando la acción material provenga de un sistema inteligente cuyo funcionamiento autónomo dificulta la identificación del responsable.

⁴⁸⁴ AYO FERRÁNDIZ, C.; SEIJO BAR, Á.; GARRE ANGUERA DE SOJO, I.; Y GONZÁLEZ GUILLÉN, P., “Responsabilidad civil e inteligencia artificial”, *Actualidad Jurídica Uría Menéndez*, número 67, mayo 2025, pp. 29-56.

⁴⁸⁵ Ley Orgánica 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen, *Boletín Oficial del Estado*, número 115, de 14 de mayo, establece, en su artículo 9, el procedimiento legal para la defensa frente a las injerencias o intromisiones ilegítimas, así como las pretensiones que podrá deducir el perjudicado. En lo que respecta a la indemnización de perjuicios, se presume que éstos existen en todo caso de injerencias o intromisiones acreditadas, y comprenderán no sólo la de los perjuicios materiales, sino también la de los morales, de especial relevancia en este tipo de actos ilícitos.

⁴⁸⁶ STS 114/2022, de 15 de febrero, ECLI:ES:TS:2022:554, dónde se recoge este sistema de valoración del daño teniendo en cuenta, además, los criterios del TEDH: “la indemnización otorgada en estos casos no tenga el objetivo de producir un efecto disuasorio entre la abogacía en su conjunto, cuando consideren necesario criticar a un juez por actuaciones realizadas en el ejercicio de sus funciones (STEDH, de 12 de febrero de 2019, caso Pais Pires de Lima contra Portugal)”.

Finalmente, la relación de causalidad, entendida como el vínculo que conecta una acción u omisión con el daño producido, de manera que dicho daño pueda imputarse a un sujeto, es decir, que, sin esa acción u omisión, el daño no se habría producido. La IA hace que la cadena causal se vuelva fragmentada al intervenir múltiples actores, y opaca, ya que es difícil identificar qué módulo del sistema produjo el error. Por ello, determinar la acción u omisión en estos casos es sumamente complejo.

En este contexto, teniendo la finalidad de salvaguardar aleatorios inconvenientes que la IA puede acarrear derivados de las responsabilidades civiles suscitadas, esencialmente basados en la culpa⁴⁸⁷, “la Comisión Europea publicó una propuesta de Directiva para adaptación de las normas de responsabilidad civil extracontractual a la inteligencia artificial”⁴⁸⁸.

A pesar de estas dificultades, tanto la jurisprudencia como la doctrina civil han admitido ciertos criterios flexibles que permiten imputar el daño cuando la IA amplifica un riesgo que era previsible y evitable por parte del operador o fabricante. El más destacado es la *culpa in vigilando* y la *culpa in eligendo*, no mencionadas de forma expresa en el Código Civil, pero sí se recoge sus fundamentos en el art. 1903 CC, esa obligación “es exigible no sólo por los actos u omisiones propios, sino por los de aquellas personas de quienes se debe responder”⁴⁸⁹. La omisión del deber de vigilancia opera aquí como elemento de imputación de responsabilidad a la persona jurídica. En esencia, destaca una presunción refutable de causalidad entre la culpa del demandado y el daño ocasionado por sistemas de IA de alto riesgo en casos de incumplimiento por el demandado de “deberes de diligencia”. De este modo, la responsabilidad no deriva únicamente de la conducta del agente que ha causado el daño aparentemente, sino también del incumplimiento por parte del responsable superior de prevenir, supervisar o seleccionar adecuadamente a los que actúan bajo su dependencia.

5. CONCLUSIONES

La expansión de la IA está modificando notablemente el real ejercicio de las libertades de expresión e información reconocidas en el art. 20.1.a) y d) CE.

La IA se ha convertido en una herramienta del sistema comunicativo en la actualidad, influyendo tanto en la forma en la que se crean contenidos como en su

⁴⁸⁷ Véase el *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado hacia la excelencia y la confianza*, publicado por la Comisión Europea, 2020, 65 final, de 19 de febrero de 2020, p. 19, consultado el día 10-12-2025, a las 10:01 horas, en eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065

⁴⁸⁸ AYO FERRÁNDIZ, C.; SEIJO BAR, Á.; GARRE ANGUERA DE SOJO, I.; Y GONZÁLEZ GUILLÉN, P., “Responsabilidad civil e inteligencia artificial”, op. cit. p. 34.

⁴⁸⁹ Como ejemplo, son mencionados los padres o tutores respecto de sus hijos, los titulares de un centro docente respecto de los alumnos, o los empresarios respecto de los empleados, entre otros.

difusión, visibilidad y recepción. Pero esta influencia es muy ambigua, puesto que combina avances muy positivos con riesgos relevantes que requieren una reflexión jurídica y social.

Resulta innegable que la IA aporta beneficios claros, tales como facilitar la creación y acceso a la información, analizar grandes volúmenes de datos. Del mismo modo, favorece la participación de colectivos que tradicionalmente quedaban en ocasiones al margen del debate público, ampliando la pluralidad de voces presentes en la conversación digital. Sin embargo, este balance positivo convive, con efectos negativos que requieren atención y precaución.

Por el contrario, introduce riesgos graves relacionados con la desinformación, la difusión de contenidos falsos y manipulados, así como la creación de burbujas informativas que limitan el pluralismo y la diversidad de opiniones.

Ante este escenario, se vuelve imprescindible, y prácticamente necesario, promover el uso responsable y ético de la IA. Para ello, es fundamental garantizar mayores niveles de transparencia y control sobre los sistemas automatizados que intervienen en la recomendación y difusión de contenidos e información. También, resulta esencial reforzar la supervisión humana de la que se viene hablando en los procesos comunicativos y exigir que, tanto los desarrolladores como los usuarios, actúen con la debida diligencia para evitar daños. La IA debe entenderse como un instrumento al servicio de las personas y de la democracia, y no como un elemento que condicione la libertad individual de los sujetos.

Asimismo, el futuro del derecho de la libertad de expresión e información se ve condicionado a estas herramientas, por lo que requiere una adaptación a las nuevas realidades tecnológicas. La protección constitucional ya no puede limitarse a los medios de comunicación tradicionales, sino que debe abarcar también el ámbito digital, donde hoy en día se desarrolla la mayor parte del debate público. Esto implica reconocer y prever jurídicamente que los algoritmos de recomendación desempeñan un papel similar al de editores o directores de medios y su funcionamiento debe estar sujeto a límites, controles y garantías.

Igualmente, será necesario actualizar la regulación en materia de responsabilidad civil para responder de forma adecuada a los daños y perjuicios provocados por la creación y difusión de contenidos generados por la IA, conllevando una actualización y modernización del sistema de responsabilidad extracontractual previsto en nuestro Código Civil (*ex* artículos 1902 a 1910).

En todo este proceso, la sociedad civil y las instituciones públicas tienen un papel determinante, donde las autoridades deberían impulsar normativas claras, sistemas de supervisión eficaces y políticas que garanticen el respeto a los derechos fundamentales en el ámbito digital. Y por su parte, la ciudadanía necesita formación en competencias

digitales y pensamiento crítico para desenvolverse con seguridad en un entorno comunicativo cada vez más complejo.

En último término, la IA representa una oportunidad para ampliar el ejercicio de la libertad de expresión e información, pero también plantea desafíos que pueden ponerla en riesgo si no se actúa con responsabilidad. Solo mediante un control adecuado sobre ella y un compromiso público y general será posible un debate público, libre y respetuoso con los derechos humanos.

BIBLIOGRAFÍA

- AYO FERRÁNDIZ, C.; SEIJO BAR, Á.; GARRE ANGUERA DE SOJO, I.; Y GONZÁLEZ GUILLÉN, P., “Responsabilidad civil e inteligencia artificial”, *Actualidad Jurídica Uría Menéndez*, número 67, mayo 2025.
- BARRIO ANDRÉS, M., *Derecho de los robots*, 2018, citado en SÁNCHEZ ESPARZA, M., Y PALELLA STRACUZZI, S., *Inteligencia artificial y nuevas fronteras: la transformación de la comunicación, de los derechos humanos, la educación y la innovación organizacional*, Tirant lo Blanch, Valencia, 2024.
- BELMAR TALÓN, Á., *La definición legal de periodista en la historia de España: El artículo 20 de la Constitución española de 1978 como estatuto jurídico del profesional de la información*, tesis doctoral, 2017, publicado en Dehesa, Universidad de Extremadura.
- COMISIÓN EUROPEA, *Libro Blanco sobre la inteligencia artificial: un enfoque europeo orientado hacia la excelencia y la confianza*, publicado por la Comisión Europea, 2020, 65 final, de 19 de febrero de 2020.
- CORTÉS GARCÍA, U., BÉJAR ALONSO, J., MORENO RIBAS, A., *Inteligencia artificial*, Universidad Politécnica de Cataluña, Barcelona, 1993.
- GÓMEZ LIGÜERRE, C., “Responsabilidad por daños causados por la inteligencia artificial”, *InDret*, pp. 9-11, 2025, consultado el día 13-11-2025, a las 20:23 horas, en <https://indret.com/wp-content/uploads/2025/01/1924.pdf>
- HERRERA ORTIZ, J. J., PEÑA AVILÉS, J. M., HERRERA VALDIVIESO, M. V., MORENO MORÁN, D. X. “La inteligencia artificial y su impacto en los medios de comunicación”, *TELOS: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, volumen 26, número 1, 2024.
- KASSHMIR HILL, “Un adolescente tenía tendencias suicidas. ChatGPT3 fue el amigo en quien confió”, *The New York Times*, 27 de agosto de 2025, consultado el día 11-11-2025, a las 20:21 horas.
- LÓPEZ DE MÁNTARAS BADIA, R., MESEGUER GONZÁLEZ, P., *Inteligencia artificial*, Editorial CSIC, Madrid, 2017.
- NAVAS NAVARRO, S., GÓRRIZ, C., *Inteligencia artificial: tecnología, derecho*, 1ª edición, Tirant lo Blanch, Valencia, 2017.
- SANAHUJA SANAHUJA, R., & LÓPEZ RABADÁN, P., “Ámbitos de aplicación periodística de la Inteligencia Artificial. Razón y Palabra. Mapa conceptual, funciones profesionales y tendencias en desarrollo en el contexto de la pandemia global

de la Covid-19”, Vol. 25, número 112, 2021, *Comunicación Política y Tecnologías Aplicadas: desinformación, noticias falsas y verificación de hechos*, volumen 25, número 112, 2021, Proyectos de Investigación Científica y Desarrollo Tecnológico “Ética discursiva y Democracia ante los retos de la Inteligencia Artificial”, PID2019-109078RB-C21 y PID2019-109078RB-C22, financiados por el Ministerio de Ciencia, Innovación y Universidades.

ZUBOFF, S., *The Age of Surveillance Capitalism*, 2019, quien analiza cómo el control de la información y los algoritmos de las plataformas condicionan la visibilidad, jerarquización y acceso al discurso público; citado y analizado también en “Religación”, *Revista de Ciencias Sociales y Humanidades*, volumen 6, número 28), págs. 220-222, <https://doi.org/10.46652/rgn.v6i28.811>

Repertorio cronológico jurisprudencial

STS 1119/2025, de 11 de septiembre, ECLI:ES:TS:2025:3826

STC 87/2025, de 7 de abril, ECLI:ES:TC:2025:87

ATSJ Navarra núm. 2/2024, de fecha 4 de septiembre, ECLI:ES:TSJNA:2024:38

STS 114/2022, de 15 de febrero, ECLI:ES:TS:2022:554

STEDH, de 16 de junio de 2015, ECLI:UE:TEDH:2015:6456

STEDH caso Otegi Mondragón contra España, de 15 de marzo de 2011, ECLI:UE:TEDH:2011:2034

STC, 29 de noviembre de 1999, ECLI:ES:TC:1999:217

STC 15/1993, de 18 de enero, ECLI:ES:TC:1993:15

STC 240/1992, de 21 de diciembre, ECLI:ES:TC:1992:240

Corte Suprema de los Estados Unidos, *Texas v. Johnson*, número 88-155, de 21 de junio de 1989

STC 51/1989, de 22 de febrero de 1989, ECLI:ES:TC:1989:51

STC 30/1982, de 1 de junio, ECLI:ES:TC:1982:30

STC 6/1981, de 16 de marzo, ECLI:ES:TC:1981:6



CAPÍTULO 12.

IMPACTO DE LA INTELIGENCIA ARTIFICIAL EN EL ÁMBITO DE LA PROPIEDAD INTELECTUAL, TENSIONES Y DESAFÍOS⁴⁹⁰

ALEJANDRO ROSILLO FAIRÉN⁴⁹¹

Profesor Titular de Derecho
Civil

Departamento de Ciencias Jurídicas y Humanidades
Facultad de Ciencias Jurídicas, Educación y Humanidades
Universidad Europea de Ma-
drid

alejandro.rosillo@universidadeuropea.es

Resumen

El presente trabajo efectúa un análisis crítico sobre la evolución de la Inteligencia Artificial y su impacto en la propiedad intelectual, ya que esta institución tiene una máxima relevancia en nuestra sociedad y está llamada a incrementarla en los próximos años, aunque existen diversos retos e incertidumbres que deben ser salvados. Las

⁴⁹⁰ El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

⁴⁹¹ Alejandro Rosillo Fairén es Doctor en Derecho con premio extraordinario de Doctorado, Profesor Titular de Derecho Civil de la UEM, acreditado ACAP/ANECA para las figuras de Contratado Doctor y Universidad Privada, con sexenio de investigación vivo. Algunas de sus últimas publicaciones en los últimos años son las siguientes: “El jurado en España: una aproximación histórica. Repercusión en el ámbito penal y civil”. Incluido en “Nuevas realidades: un enfoque desde el derecho” M.^a Ángeles Cano Linares (dir.), Ana Isabel Carreras Presencio (dir.) Ed. Dykinson. 2025. Pág. 146-170. “Reflexiones sobre la posible transformación del concepto de patria potestad en el marco de la Ley 8/2021”. Incluido en “La protección jurídica de las personas con discapacidad: perspectiva interna e internacional”. Ed. Aranzadi La Ley. 2024. Pág. 129-157. “Conflictos familiares y gastos extraordinarios en la pensión de alimentos: algunas cuestiones prácticas” Coautoría con Castro Acosta. Vox Juris. Vol. 42, N.º. 2, 2024, Pág. 171-189. “Reflexiones sobre la posible modificación o extinción de la pensión de alimentos: Análisis jurisprudencial”. Coautoría con Castro Acosta. Vox Juris. Vol. 40. N.º 2. 2022. Pág. 90-114.. “Breves reflexiones en torno al derecho a la integridad física de los menores de edad en el ámbito familiar. Los derechos humanos en el siglo XXI. En la conmemoración del 70 aniversario de la declaración. Pág. 329-333. Madrid. Edisofer. Real Academia de Jurisprudencia y Legislación, 2020.

posibilidades y facilidades que se auguran para la humanidad son casi ilimitadas, pero se requiere tener presente los riesgos derivados de la misma, entre los que destaca la destrucción masiva y la falta de ética en la programación y uso de la misma.

Palabras clave

Inteligencia Artificial – Propiedad Intelectual - España - autoría – obras – ética – retos – desafíos - Derecho Civil

Abstract

This paper provides a critical analysis of the evolution of Artificial Intelligence and its impact on intellectual property, as this institution holds paramount importance in our society and is expected to increase its relevance in the coming years. However, there are various challenges and uncertainties that must be overcome. The possibilities offered by artificial intelligence and the benefits it promises for humanity are almost limitless; nevertheless, it is essential to keep in mind the risks associated with it, among which stand out massive destruction and the lack of ethics in its programming and use.

Keywords

Artificial Intelligence – Intellectual Property – Spain – Authorship – Works – Ethics – Challenges – Issues - Civil Law

Sumario: 1. Orígenes de la IA. aproximación a su definición. acercamiento a su dimensión ética. 2. Marco normativo básico. hacia unos principios universales de la IA. límites éticos. 3. Autoría de obras realizadas con IA. 4. Riesgos en el uso de la IA por entes privados y públicos. 4.1 Análisis práctico de conflictos con la IA. 4.1.1 La IA dentro del proceso judicial: herramienta no sustitutiva de la labor de las partes. 5. Conclusiones. 6. Bibliografía.

1. ORÍGENES DE LA IA. APROXIMACIÓN A SU DEFINICIÓN. ACERCAMIENTO A SU DIMENSIÓN ÉTICA.

Modernamente se atribuye la creación del concepto IA al científico MCCARTHY en 1956⁴⁹², en pleno inicio de la informática, definiéndose como “la ciencia e ingenio de hacer máquinas inteligentes”. Todo ello sin perjuicio de las valiosas aportaciones de autores anteriores, entre las que destaca especialmente el científico español TORRES QUEVEDO, verdadero pionero que a finales del XIX inventó una “máquina algebraica”, que estaba diseñada para resolver ecuaciones algebraicas hasta polinomios de grado 8, utilizaba una “escala logarítmica” y era capaz de calcular valores de funciones polinómicas de forma continua. Se presentó en la Academia de Ciencias de París en

⁴⁹² MCCARTHY, J.; MINSKY, M.; ROCHESTER, N.; SHANNON, C. Una propuesta para el Proyecto de investigación de verano de Dartmouth en Inteligencia Artificial 31.08.1955. Recuperado de: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> (última consulta, 19.8.2025)

1901 y es considerada una de las primeras máquinas automáticas de la historia de la humanidad⁴⁹³. Posteriormente, en 1913 desarrolló una máquina para jugar al ajedrez que podía predecir la siguiente jugada en función de las circunstancias, así como el “aritmómetro electromecánico” en 1920, el que puede ser considerado como primer ordenador en sentido actual de la historia⁴⁹⁴.

TURING describió su máquina (1936) otra precursora de los modernos ordenadores⁴⁹⁵, que fue desarrollada por su equipo en Bletchey Park con fines militares, logrando incluso adelantar el fin del conflicto bélico al descodificar los complejos sistemas de comunicación de la Alemania nazi, que tenían fama de ser invulnerables, amparados en el amplio desarrollo científico del país. Todo ello sin perjuicio de otros autores coetáneos como McCULLOCH y PITTS⁴⁹⁶. Y en 1959, SAMUEL, ingeniero empleado de IBM, acuñó un término que es perfectamente válido en la actualidad: “aprendizaje automático”⁴⁹⁷.

Avanzado el siglo XX, mediando todavía el conflicto antagónico entre las dos superpotencias, URSS y EEUU en plena “guerra fría” y planteándose el riesgo cierto de una guerra termonuclear a escala global, en 1983 el cine nos mostró una excelente e impactante visión del creciente e imparable desarrollo de la IA, aunque en aquel momento bien es cierto que no era conocida de esa forma, sino que había referencias genéricas a las “supercomputadoras”, que se presentaban ante el gran público como máquinas de un tamaño descomunal que ocupaban salas e incluso edificios completos y que eran capaces de procesar millones de operaciones por segundo.

Especialmente relevantes eran las aplicaciones militares informáticas al más alto nivel, en la conocida película “Juegos de guerra”. Se planteaba en la misma la incipiente comunicación en red entre ordenadores, así como los peligros de la adopción de esta tecnología cuando se situaba fuera del control humano y podía adoptar decisiones autónomas con el manejo de armas atómicas, máxima que se consolidaría en 1984 con el inicio de la saga “Terminator” donde una IA ficticia, llamada “skynet”, “toma conciencia de sí misma” y actúa de forma autónoma, impidiendo su propia desconexión cuando los humanos son conscientes del peligro que entraña para su supervivencia,

⁴⁹³ <https://photocalcul.com/Calcul/Machines/NoticesMachines/Torres%20Quevedo.pdf> (última consulta, 19.8.2025)

⁴⁹⁴ GONZÁLEZ DE POSADA, F. y GONZÁLEZ, REDONDO, F. A. “Leonardo Torres Quevedo (1852-1936) Primera parte. Las máquinas algébricas”. La gaceta de la RSME. Vol. 7.3 (2004), Págs. 787–810

⁴⁹⁵ TURING, A. M. Sobre los números computables, con una aplicación al Entscheidungsproblem. Actas de la Sociedad Matemática de Londres. 1936. Pág. 230 y ss.

⁴⁹⁶ McCULLOCH, W. & PITTS, W. ACálculo lógico de las ideas inmanentes en la actividad nerviosa. Boletín de Biofísica Matemática., Vol. 5, pp. 115-133 (1943).

⁴⁹⁷ SAMUEL, A. L. Algunos estudios sobre el aprendizaje automático utilizando el juego de damas. Revista IBM. Julio de 1959. Pág. 210 y ss.

propiciando la destrucción de la humanidad en un futuro terrorífico y distópico en el que las máquinas pasan a dominar el mundo y los seres humanos se encuentran al borde de la extinción tras una guerra apocalíptica y el posterior invierno nuclear. Acertadamente, ante tantos riesgos, diversos autores plantean el “problema del alineamiento”: ¿cómo conseguir que las máquinas inteligentes no se revuelvan contra sus creadores?”⁴⁹⁸ ¿El crecimiento informático ilimitado plantea riesgos ilimitados?

Otro hito en el desarrollo de la IA, verdadero punto de inflexión, acaeció en 1997, cuando el ordenador “deep blue” venció al mítico campeón de ajedrez ruso KASPAROV. En 2016 volvió a suceder lo mismo, cuando la IA “AlphaGo” derrotó al que fuera dieciocho veces campeón mundial de Go, el surcoreano LEE SE-DOL. Se ratificaba así la creciente opinión de que la humanidad había sido superada por la tecnología, lo que supuso un indudable impacto en la ciudadanía, que hoy día parece haber asumido esa circunstancia, habida cuenta de la potencia de los dispositivos electrónicos domésticos. Esa progresión tecnológica se encuentra en plena consonancia con las previsiones de la Ley de Moore, formulada en 1965 y en virtud de la cual el aumento de los transistores en circuitos integrados pasaría a doblarse cada dos años. Y aunque dicha Ley ha sido cuestionada, lo cierto es que desde entonces se ha producido un verdadero crecimiento exponencial. ¿Tendrá fin el mismo en algún momento? ¿Será un crecimiento verdaderamente ilimitado? ¿Deberá ponerse fin, aunque la tecnología tenga capacidad para avanzar, en aras de alcanzar el beneficio y seguridad de la humanidad y garantizar la supervivencia de las generaciones futuras?

Dicho todo lo anterior, la tarea de acotar la IA, con el ánimo de pretender definir la misma no resulta en modo alguno sencilla, sobre todo cuando es una tecnología que se desarrolla a una velocidad exponencial con aplicaciones en todos los ámbitos. De hecho, a lo largo de las últimas décadas, son múltiples las definiciones que sobre la misma se han ido dando en diversos ámbitos del conocimiento, que en muchos casos tienen un carácter similar o complementario.

Dentro del ámbito anglosajón se ha destacado que “La IA es un campo de la ciencia y la ingeniería que se ocupa de la comprensión, desde el punto de vista informático, de lo que se denomina comúnmente comportamiento inteligente. También se ocupa de la creación de artefactos que exhiben este comportamiento”⁴⁹⁹. E Igualmente se ha clasificado en dos subdivisiones, a saber, “IA como ciencia y la IA como

⁴⁹⁸ LLANO ALONSO, F. “Ética de la inteligencia artificial y derecho. Consideraciones a propósito de los límites y la contención del desarrollo tecnológico derechos y libertades”. ISSN: 1 133-0937. Número 51, Época II, junio 2024. Pág. 185.

⁴⁹⁹ SHAPIRO, S. C.: *Enciclopedia de la Inteligencia Artificial* (2nd Ed.). John Wiley & Sons, Inc., New York, 1992.

ingeniería del conocimiento”⁵⁰⁰ y “combinación de algoritmos planteados con el propósito de crear máquinas que presenten las mismas capacidades que el ser humano”⁵⁰¹.

Para la Real Academia Española, la IA se trataría de una clase de “disciplina 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”. Y desde un punto de vista netamente administrativo, para el Gobierno de España, la IA “es un campo de la informática que se enfoca en crear sistemas que puedan realizar tareas que normalmente requieren inteligencia humana, como el aprendizaje, el razonamiento y la percepción. Estos sistemas pueden percibir su entorno, razonar sobre el conocimiento, procesar la información derivada de los datos y tomar decisiones para lograr un objetivo dado”⁵⁰².

A nivel internacional, consciente de los riesgos de su implementación masiva sin límites ni cortapisas, La UNESCO publicó en 2022 su Recomendación sobre la ética de la IA, configura la misma de forma deliberadamente amplia: “Son aquellos sistemas con capacidad para procesar datos de forma similar a un comportamiento inteligente”⁵⁰³.

La importancia de la figura es tal que en 2022 según Fundéu, la Fundación del Español Urgente, la palabra del año fue “inteligencia artificial”, ya que se encuentra verdaderamente omnipresente en nuestra sociedad. Hasta tal punto es así que si nos preguntásemos si sería posible un mundo en 2025 carente de IA, muy probablemente la respuesta sea negativa, por la ingente variedad de asuntos en los que tiene aplicación, tanto civiles como militares. Desde el mundo jurídico a la inversión en el mercado bursátil, pasando por el diagnóstico de enfermedades graves o la realización de informes sobre una casi ilimitada variedad de temas. Y quizás el impacto real para la humanidad sea incluso superior al que supuso la irrupción masiva de internet hace un cuarto de siglo. De hecho, para muchos autores este va a ser el “gran cambio” de la humanidad⁵⁰⁴ en el que ya es un hombre “cyborg”⁵⁰⁵ y muy probablemente supondrá una

⁵⁰⁰ ROQUE, M. y PALMA, J. *Inteligencia Artificial. Técnicas, métodos y aplicaciones*. Madrid, 2008. McGraw-Hill. Pág. 5

⁵⁰¹ RUSSELL S. J., NORVIG, P. *Inteligencia artificial. Un enfoque moderno*. Englewood Cliffs, NJ: Prentice Hall, 1994.

⁵⁰² <https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr> (última consulta, 19.8.2025)

⁵⁰³ UNESCO, Recomendación sobre la ética de la IA. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. París (Francia). 2022. <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics> (última consulta, 19.8.2025)

⁵⁰⁴ <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (última consulta 28.11.2025)

⁵⁰⁵ Pérez Suárez, J. R. (2017) *Somos cíborgs: desarrollo de un modelo teórico para comprender el comportamiento delictivo en Internet*. Criminología y justicia.

nueva revolución industrial cuyo alcance está todavía por determinar. De hecho, ya en este año 2025 cobra fuerza un nuevo concepto en nuestra sociedad, el denominado “transhumanismo”, como transformación de la condición humana a través de tecnologías. De hecho ya se encuentran a la venta diversos robots para su uso doméstico a precios razonablemente asequibles, que superan y con creces el concepto clásico de electrodoméstico: “NEO Gamma, el robot minimalista que aspira a ser el mayordomo de tu casa: te escucha y limpia tu hogar”⁵⁰⁶.

Las mejoras que puedan lograrse en nuestra sociedad a través de la IA son uno de los grandes reclamos de la misma. Así por ejemplo pueden reseñarse: la automatización de tareas repetitivas, mejora en la toma de decisiones y en la productividad, reducción de errores humanos, disponibilidad permanente, avances en medicina o Derecho, predicción y prevención, impulso a la innovación, optimización en transporte y logística: mejora rutas, reduce costes y tiempos de entrega:

“Los sistemas inteligentes de la IA detectan ineficiencias y desperdicios en la cadena de suministro. Al abordar estas ineficiencias, las organizaciones pueden reducir significativamente el desperdicio en las fases de producción, almacenamiento y distribución. La IA evalúa la utilización de recursos en los procesos de producción, recomendando alternativas más sostenibles o formas de utilizar menos recursos”⁵⁰⁷.

2. MARCO NORMATIVO BÁSICO. HACIA UNOS PRINCIPIOS UNIVERSALES DE LA IA. LÍMITES ÉTICOS.

La IA se trata sin duda de una materia de máxima relevancia dentro de la agenda político-administrativa a nivel mundial, aunque bien es cierto que en no pocos países se ha acometido de forma ciertamente tardía. De hecho, la inactividad de la UE dio paso “a iniciativas como las de Francia, que el 12 de septiembre de 2023 anunciaba un proyecto de ley con la intención de ampliar el marco de protección imperativa de los titulares de derechos de autor frente a las amenazas que supone la IA para el futuro de la Propiedad Intelectual, con el consiguiente riesgo de fragmentación del Mercado Único Digital”⁵⁰⁸.

⁵⁰⁶ https://www.elespanol.com/omicrono/hardware/20250224/neo-gamma-robot-minimalista-aspira-mayordomo-casa-escucha-limpia-hogar/926657798_0.html (última consulta 28.11.2025)

⁵⁰⁷ <https://theofficer.es/que-es-la-inteligencia-artificial-responsable-y-como-impacta-en-la-estrategia-de-responsabilidad-social-corporativa-de-las-empresas/> (última consulta 28.11.2025)

⁵⁰⁸ MARTÍN ALÁEZ, F. La tutela civil de los derechos de propiedad intelectual en el mercado único digital (entre plataformas e inteligencia artificial) Tesis doctoral UDC / 2024. Pág. 671. <https://ruc.udc.es/rest/api/core/bitstreams/9bb06547-ef7a-44d3-9eb5-8a509bae0592/content> (última consulta, 19.8.2025)

En España el Consejo de Ministros, aprobó el 15 de mayo la Estrategia de Inteligencia Artificial 2024, que da continuidad y refuerza la Estrategia de Inteligencia Artificial (ENIA) publicada en 2020, en cumplimiento del Plan de Recuperación, Transformación y Resiliencia (PRTR), y la adapta a los nuevos avances tecnológicos. El impulso de la inteligencia artificial también está contemplado en la Agenda España Digital 2026 como un elemento clave de carácter transversal para transformar el modelo productivo e impulsar el crecimiento de la economía española⁵⁰⁹, especialmente necesitado tras encontrarnos en términos reales con la misma capacidad adquisitiva que hace tres décadas.

Todo ello se encuentra en un contexto de fuerte digitalización en todas las áreas, incluyendo la Administración de Justicia, en cual la IA está llamada a ostentar un importante papel, sobre todo en lo relativo a la creación de múltiples asistentes que colaboren en la misma⁵¹⁰, aunque todavía tiene un carácter inicial o incipiente⁵¹¹. Así cabe destacar que en 2024 el CGPJ constituyó un grupo de trabajo para el estudio y análisis de las implicaciones de la IA en el ejercicio de la función jurisdiccional⁵¹². Por su parte la Ley Orgánica 1/2025, de 2 de enero, de medidas en materia de eficiencia del Servicio Público de Justicia, es un nuevo y decidido impulso de las actuaciones telemáticas⁵¹³

⁵⁰⁹ <https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/transformacion-digital-y-funcion-publica/Paginas/2024/ia-inteligencia-artificial-estrategia-espana.aspx> (última consulta, 19.8.2025)

⁵¹⁰ BEIRO MAGÁN, J. M. “Automatización e Inteligencia Artificial: por qué debe generalizarse su implantación en el sector justicia español”<https://pensamientocritico.sisej.com/automatizacion-e-inteligencia-artificial-por-que-debe-generalizarse-su-implantacion-en-el-sector-justicia-espanol/> (última consulta, 19.8.2025)

⁵¹¹ LORENZO PÉREZ, C. “Inteligencia artificial en la Administración de Justicia: regulación española y marco europeo e internacional. Proyectos desarrollados por el Ministerio de Justicia de España”. Ed. Centro de Estudios Jurídicos. Madrid, 2022.

⁵¹² “Los objetivos iniciales del grupo de trabajo serán certificar y evaluar las aplicaciones potenciales de la IA respecto del Poder Judicial; desarrollar políticas y directrices específicas de aplicación en los procesos judiciales, asegurándose de que los estándares éticos, legales y procesales sean debidamente respetados; impulsar proyectos piloto y pruebas; Fomentar el desarrollo de actividades formativas en esta materia para la Carrera Judicial y garantizar que los mecanismos de auditoría y control que se implantan funcionen de forma adecuada, transparente y segura”. https://www.poderjudicial.es/portal/site/cgpi/menuitem.65d2c4456b6ddb628e635fc1dc432ea0/?vgnext-toid=76d00f8d8a822910VgnVCM1000004648ac0aRCRD&vgnextchannel=ea1732cd1ddaa210VgnVCM100000cb34e20aRCRD&vgnextfmt=default&vgnextlocale=es_ES (última consulta, 19.8.2025)

⁵¹³ Artículo 8: “Actuaciones desarrolladas por medios telemáticos. 1. Las partes podrán acordar que todas o alguna de las actuaciones de negociación en el marco de un medio adecuado de solución de controversias, se lleven a cabo por medios telemáticos, por videoconferencia u otro medio análogo de transmisión de la voz o la imagen, siempre que quede garantizada la

que ya se venían consolidando desde Real Decreto 84/2007, de 26 de enero, sobre implantación en la Administración de Justicia del sistema informático de telecomunicaciones Lexnet para la presentación de escritos y documentos, el traslado de copias y la realización de actos de comunicación procesal por medios telemáticos. A nivel administrativo, ello tuvo su reflejo en la Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas y la Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público, que dieron lugar a plataformas cibernéticas como “red sara”, superando el régimen clásico previsto en la Ley 30/1992, de 26 de noviembre, de Régimen Jurídico de las Administraciones Públicas y del Procedimiento Administrativo Común y más aún del Decreto de 26 julio de 1957 por el que se aprueba el texto refundido de la Ley de Régimen jurídico de la Administración del Estado.

En el ámbito europeo, cabe destacar el impacto sobre la IA del Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas sobre la inteligencia artificial y se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 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 (Ley de Inteligencia Artificial). Directiva (UE) 2019/790, del Parlamento Europeo y del Consejo, de 17 de abril de 2019, sobre los derechos de autor y derechos afines en el mercado único digital. Su artículo 1 señala que

“El objetivo del presente Reglamento es mejorar el funcionamiento del mercado interior y promover la adopción de una inteligencia artificial (IA) centrada en el ser humano y fiable, garantizando al mismo tiempo un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta, incluidos la democracia, el Estado de Derecho y la protección del medio ambiente, frente a los efectos perjudiciales de los sistemas de IA en la Unión así como prestar apoyo a la innovación”.

Resulta obvio (y necesario) que se reconozca expresamente a todos los niveles y muy especialmente en el ámbito legislativo el riesgo inherente a la utilización y desarrollo de esta tecnología, que sin ningún género de dudas se va a incrementar en los próximos años, conforme se vaya desarrollando la misma. ¿Quién sería capaz en este año 2025 de “poner puertas el campo” en este ámbito? Y partiendo de esa indiscutible base, que la IA no se encuentra exenta de riesgos, ya hemos anunciado que la

identidad de los intervinientes y el respeto a las normas previstas en este título y, en su caso, a la normativa de desarrollo específicamente contemplada para la mediación. 2. Cuando el objeto de controversia sea una reclamación de cantidad que no exceda de seiscientos euros se desarrollará preferentemente por medios telemáticos, salvo que el empleo de éstos no sea posible para alguna de las partes”.

dimensión ética en su implementación y desarrollo es absolutamente irrenunciable⁵¹⁴, ya que su uso “puede derivar en aplicaciones nocivas o beneficiosas para las personas”⁵¹⁵. De ahí que la UNESCO haya propuesto acertadamente “un marco universal de valores, principios y acciones para orientar a los Estados en la formulación de sus leyes, políticas u otros instrumentos relativos a la IA, de conformidad con el Derecho internacional”⁵¹⁶.

El desarrollo de unos principios universales que puedan inspirar el desarrollo tecnológico no es ni mucho menos un concepto novedoso. El célebre y mítico ASIMOV ya sentó las bases con sus famosas leyes, que deben rescatarse y aplicarse, pues su vigencia tiene más relevancia que nunca. No en vano se afirma una obviedad manifiesta: que el ser humano ha de estar por encima de las máquinas en todo momento. “Primera Ley. Un robot no hará daño a un ser humano, ni por inacción permitirá que un ser humano sufra daño. Segunda Ley. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley. Tercera Ley. 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”⁵¹⁷. ¿Estas leyes se pueden llegar a cumplir en su plenitud con una IA que obedezca ilimitadamente al ser humano sin ninguna clase de límites éticos? ¿Podemos imaginar un sistema totalitario en el que paradójicamente el humano esclavice a la IA en perjuicio de otros humanos?

Efectivamente, se plantea si ha de existir o no un verdadero binomio entre IA y ética -considerando desde estas líneas que es absolutamente indisoluble- especialmente en las grandes multinacionales que tienen la capacidad económica y técnica de desarrollar esta tecnología al más alto nivel:

“La ética es voluntaria, pero no opcional: es una exigencia de la excelencia. El directivo ético procura tener en cuenta todas las consecuencias relevantes de sus acciones, para los demás y para él mismo; sabe ponerse en la piel del otro, no omite la consideración de ninguna parte interesada relevante; puede llevar a cabo una adecuada gestión del riesgo

⁵¹⁴ MARÍN GARCÍA, S. Ética e inteligencia artificial. Cuadernos de la Cátedra CaixaBank de Responsabilidad Social Corporativa. N° 42 Septiembre de 2019. DOI: <https://dx.doi.org/10.15581/018.ST-522>

⁵¹⁵ ARGANDOÑA, A. Ética e inteligencia artificial (I). IESE Blog Network: Economía, Ética y RSE. 2019. Recuperado de blog.iese.edu/antonioargandona/2019/03/25/etica-e-inteligencia-artificial-i. (última consulta, 19.8.2025)

⁵¹⁶ UNESCO, Recomendación sobre la ética de la IA. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. París (Francia). 2022. <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics> (última consulta, 19.8.2025)

⁵¹⁷ ASIMOV, I. «Círculo vicioso». Los robots. trad. Domingo Santos. Barcelona: Ed. Martínez Roca. Barcelona, 1989.

por los impactos y prácticas de la empresa; desarrolla una cultura de diálogo; se anticipa a los problemas”⁵¹⁸.

Y partiendo de esa base, se han sintetizado por la doctrina los principios fundamentales que han de guiar la ética en el marco de la IA: 1. Transparencia, 2. Justicia, equidad e igualdad, 3. No maleficencia, 4. Responsabilidad, 5. Privacidad⁵¹⁹.

Dichos principios recogidos doctrinalmente están siendo recogidos por la jurisprudencia, con especial ímpetu y refuerzo en lograr la transparencia. Cabe destacar a estos efectos la Sentencia del TJUE (Sala Primera) de 27.2.2025, Caso Dun & Bradstreet Austria GmbH / CK. En este caso, una persona (CK) fue considerada no apta para un contrato telefónico porque una evaluación automatizada de su solvencia (por Dun & Bradstreet) la situó como riesgo. CK solicitó al proveedor información sobre la lógica del proceso (UE) 2016/679 (RGPD) y la directiva sobre secretos empresariales. Se determinó que el derecho de acceso del art. 15(1)(h) RGPD debe permitir al interesado “información significativa sobre la lógica implicada” en la decisión automatizada. Esa información debe presentarse de forma concisa, transparente, inteligible y fácilmente accesible, no basta con entregar una fórmula matemática compleja o todos los pasos técnicos del algoritmo. Donde se invoca secreto comercial o datos de terceros protegidos, el controlador puede remitir la información al órgano competente (autoridad o tribunal) para que éste equilibre derechos de forma adecuada.

Resulta igualmente llamativo que los poderes públicos no apliquen siempre estos principios en materia de IA. Así podemos destacar la STS 1119/2025, de 11 de septiembre sobre la aplicación informática BOSCO del Ministerio para la Transición Ecológica y el Reto Demográfico que gestiona el bono social eléctrico. La Fundación Ciudadana Civio reclamó el acceso al código fuente de la aplicación BOSCO, que decide de forma automatizada si un solicitante cumple los requisitos para la prestación eléctrica del bono social. El Tribunal Supremo falló que dicho código fuente puede considerarse información pública u objeto de derecho de acceso cuando la decisión que genera afecta derechos de los ciudadanos, y ordenó a la Administración facilitar dicho acceso, salvo partes justificadas de protección, propiedad intelectual o seguridad. Marca un precedente importante en España sobre “transparencia algorítmica”: los algoritmos que determinan derechos o prestaciones públicos no quedan en una “caja negra” impune. Afirma que la protección de la propiedad intelectual o la seguridad no puede invocarse de forma genérica para impedir el acceso ciudadano cuando están en juego decisiones automatizadas de la Administración. Refuerza que la administración

⁵¹⁸ <https://blog.iese.edu/antonioargandona/2019/03/25/etica-e-inteligencia-artificial-i/> (última consulta 28.11.2025)

⁵¹⁹ JOBIN, JENCA Y VAYENA, «The Global Landscape of AI Ethics Guidelines» https://www.researchgate.net/publication/335579286_The_global_landscape_of_AI_ethics_guidelines (última consulta 28.11.2025)
<https://blogs.uoc.edu/informatica/es/cinco-principios-inteligencia-artificial-etica/> (última consulta 28.11.2025)

electrónica y el uso creciente de sistemas automatizados implican deberes de información, auditoría y control público.

Esa anunciada transparencia tiene también, paradójicamente, otra dimensión: la información real sobre el uso de la IA. Así el Auto de 6.6.2023 de la Audiencia Nacional “Caso Arbistar”. Los investigados publicitaban programas automáticos (“algoritmos”) como generadores de beneficios garantizados, pero la investigación judicial demostró que el supuesto algoritmo de inteligencia artificial nunca existió, tratándose de una operación de fraude piramidal. La Audiencia Nacional evidencia que la invocación de la IA o de algoritmos puede usarse como elemento de engaño en fraudes tecnológicos. El tribunal destaca la importancia probatoria de verificar la existencia y funcionamiento real del algoritmo alegado

3. AUTORÍA DE OBRAS REALIZADAS CON IA

En materia de Propiedad Intelectual, el Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual (LPI) regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia.

“Artículo 1. Hecho generador. La propiedad intelectual de una obra literaria, artística o científica corresponde al autor por el solo hecho de su creación”. “Artículo 2. Contenido. La propiedad intelectual está integrada por derechos de carácter personal y patrimonial, que atribuyen al autor la plena disposición y el derecho exclusivo a la explotación de la obra, sin más limitaciones que las establecidas en la Ley”.

Lo cual nos lleva necesariamente a la siguiente pregunta: ¿Quién es el verdadero autor de una obra creada por la IA? El Art. 5 LPI señala que “Se considera autor a la persona natural que crea alguna obra literaria, artística o científica”. Debe recordarse que la propia IA carece de personalidad jurídica, tanto en España como en el resto del mundo, aunque en ciertos ámbitos se planteó la posibilidad contraria⁵²⁰. No es un ente reconocido jurídicamente por parte del ordenamiento jurídico. Por tanto, aún pese a su sofisticación, no es sino una es mera herramienta, en ocasiones con menos

⁵²⁰Así 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)). “Pide a la Comisión que, cuando realice una evaluación de impacto de su futuro instrumento legislativo, explore, analice y considere las implicaciones de todas las posibles soluciones jurídicas, tales como: (...) f) crear a largo plazo 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”.

capacidades que las que se le atribuyen, como la falsa creación de idiomas cibernéticos singulares totalmente indescifrables para el ser humano que obligaron a la desconexión de los sistemas: una “leyenda urbana” que (recuerda peligrosamente al relato de “terminator” y “skynet”) persiste en nuestra sociedad⁵²¹. Y como tal, no es capaz de ostentar derechos de autor sobre tales creaciones, por lo que debe reputarse verdadero autor al humano que requiere los servicios de la IA, ordenando y matizando a la misma cual debe ser el contenido concreto que se desea, por lo que resulta indiscutible la participación humana. O en su defecto al programador, generando derechos conexos⁵²². Puede traerse a colación en este punto la resolución del Parlamento Europeo de 20.10.2020 sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial (punto 15) que manifiesta:

“las obras producidas de manera autónoma por agentes artificiales y robots no deben poder acogerse a la protección mediante derechos de autor, a fin de respetar el principio de originalidad, que está unido a una persona física”.

En esta misma línea dentro del ámbito europeo, se sitúa la sentencia de la Corte municipal de Praga (referencia: No 10 C 13/2023-16) que destaca que la elaboración de imágenes a través de la IA (tarea absolutamente vital en nuestra sociedad actual) no resulta merecedora de protección en los términos previstos legalmente:

“Más allá de eso, el tribunal considera que la imagen creada por la inteligencia artificial no constituye una obra de autoría según la Sección 2 de la CA, ya que no cumple con las características conceptuales de una obra de autoría. O no es un resultado único de la actividad creativa de una persona física: el autor. El solicitante no creó personalmente la obra, sino que esta se creó con la ayuda de inteligencia artificial, y no se ha establecido en el procedimiento sobre la base de qué encargo específico. En cuanto a la tarea en sí, que supuestamente iba a ser la base de la imagen posterior creada por la inteligencia artificial, se puede hablar del tema de la obra o, posiblemente, de la idea, pero estos no son en sí mismos obras de autoría según el artículo 2, apartado 6, de la Ley de derechos de autor.”

Por tanto, y pese a que puedan existir criterios contrapuestos, consideramos que los usuarios de la IA podemos y debemos ser considerados autores de las obras realizadas con esta tecnología, en tanto en cuanto no somos meros “espectadores” en la creación de las obras. Carecemos del papel de “terceros de buena fe”, parafraseando la máxima del ordenamiento jurídico. Más bien al contrario, hemos tenido un papel fundamental e indisoluble en la misma a través de nuestras concretas instrucciones, que en no pocos casos habrán sido repetidas en múltiples ocasiones con diversas matizaciones para poder perfilar los resultados de la misma. O incluso combinarse,

⁵²¹<https://www.cnbc.com/2017/08/01/facebook-ai-experiment-did-not-end-because-bots-invented-own-language.html> (última consulta, 29.11.2025)

⁵²² ANGUIANO, J. M. (2023) Inteligencia artificial y copyright. Del dilema del thaler a la doctrina “the right to read is the right to mine”. Instituto de Derechos de Autor. Madrid, 2023.

utilizando el “producto” final para integrarlo en una diferente IA que pueda ampliar o maximizar los resultados obtenidos. En esta línea se sitúan diversos ordenamientos a nivel mundial, desde la británica Copyright, Designs and Patents Act de 1988 al ordenamiento mexicano:

“Por tanto, aunque sea una persona física quien proporciona los lineamientos e instrucciones a la inteligencia artificial, la creación no deja de ser producto de la ejecución de los algoritmos que ésta emplea generándose un contenido artificial. Por ello, en términos de la Ley Federal del Derecho de Autor, no puede ser registrable una obra que no sea creada por una persona física, por el hecho de que se encuentra expresamente señalado de esa forma porque solamente ésta es la persona capaz de crear una obra original (requisito exigido por ley) porque para tal situación es necesaria la creatividad humana”⁵²³.

Resulta palmario, tal y como venimos expresando, que en ningún caso la IA y la base algorítmica que la sustenta puede ser legalmente la titular de la obra creada, pero tampoco podría serlo un animal. Esta manifestación podría resultar sorprendente e incluso causar sonrojo en el lector, pero el 23 de Abril de 2018, el Tribunal de Apelación del Noveno Circuito de los EE.UU dictaminó que el mono que se tomó un “selfie” usando la cámara de un fotógrafo no puede demandar por derechos de autor. El macaco “naruto” tomó una fotografía con la cámara de un profesional y PETA solicitó el reconocimiento de los derechos del animal, frustradamente. Téngase en cuenta que no puede resultar tan extraña esta reflexión cuando en España se ha avanzado hacia una creciente “humanización” de los animales, de la cual es un perfecto ejemplo la Ley 17/2021, de 15 de diciembre, de modificación del Código Civil, la Ley Hipotecaria y la Ley de Enjuiciamiento Civil, sobre el régimen jurídico de los animales, que señala expresamente:

“sigue las líneas que marcan otros ordenamientos jurídicos próximos, que han modificado sus Códigos Civiles para adaptarlos a la mayor sensibilidad social hacia los animales existente en nuestros días, y también para reconocer su cualidad de seres vivos dotados de sensibilidad: la reforma austriaca de 10 de marzo de 1986; la reforma alemana de 20 de agosto de 1990, seguida de la elevación de la protección de los animales a rango constitucional en 2002 al introducir en su Ley Fundamental el artículo 20 a); la regulación en Suiza, país que también incluye en su Constitución la protección de los animales y que modificó el Código Civil y el Código de las Obligaciones a este objeto; la reforma belga de 19 de mayo de 2009; y las dos más recientes: la reforma francesa de 16 de febrero de 2015 y, de manera muy especial por la proximidad con esta que ahora se presenta, la Ley portuguesa de 3 de marzo de 2017, que estableció un estatuto jurídico de los animales”.

⁵²³ Juicio Contencioso Administrativo Núm. 788/24-EPI-01-2.- Resuelto por la Sala Especializada en Materia de Propiedad Intelectual del Tribunal Federal de Justicia Administrativa, el 30 de agosto de 2024, por unanimidad de votos.- Magistrada Instructora: Elizabeth Ortiz Guzmán.- Secretaria: Lic. Dania Karely Espinoza Perea. R.T.F.J.A. Novena Época. Año III. No. 36. Diciembre 2024. p. 388 <https://www.tfja.gob.mx/cesmdfa/sctj/tesis-pdf-detalle/47734/> (última consulta, 29.11.2025).

4. RIESGOS EN EL USO DE LA IA POR ENTES PRIVADOS Y PÚBLICOS

En la actualidad nos encontramos en una verdadera eclosión de la IA, con un crecimiento exponencial que ha superado todas las previsiones que se habían realizado en los últimos años. Aunque paradójicamente nos encontremos todavía, a día de hoy, en una fase incipiente o “embrionaria” de la misma⁵²⁴. Según las condiciones generales de servicio de ChatGPT:

“La inteligencia artificial y el aprendizaje automático son campos de estudio en rápida evolución. Trabajamos constantemente para mejorar nuestros Servicios y hacerlos más exactos, fiables, seguros y beneficiosos. Dada la naturaleza probabilística del aprendizaje automático, en algunas situaciones el uso de nuestros Servicios puede generar Output que no refleje con exactitud personas, lugares o hechos.”⁵²⁵.

Por todo ello el crecimiento de la IA parece imparable, debe destacarse que realmente está todavía por desarrollarse, debiendo todavía múltiples desafíos, entre los que destacan la posible desinformación a los ciudadanos⁵²⁶ (las famosas “fake news” elevadas a una nueva categoría de verosimilitud) y la destrucción de empleos en diversos sectores de la actividad económica, lo cual resulta altamente alarmante en buena parte de la población que se plantea un futuro hartado deshumanizado con robots camareros y limpiadores del hogar, asistentes de personas mayores, vehículos autónomos de todo tipo y hasta robots que fabriquen y reparen nuevos robots.

“Según el análisis de Randstad, la inteligencia artificial podría poner en riesgo cerca del 9,8 % del empleo actual en España, lo que equivale a alrededor de dos millones de puestos de trabajo si se considera un escenario de transformación tecnológico profundo. Aunque algunos roles se verán automatizados, otros percibirán un aumento de productividad y podrían evolucionar para asumir funciones más analíticas o creativas, lo que plantea tanto una oportunidad como un desafío para el mercado laboral.”⁵²⁷

Otras fuentes rebajan sustancialmente estas cifras:

“La creciente sofisticación de la inteligencia artificial está cambiando radicalmente el mercado laboral, al automatizar tareas rutinarias que hasta ahora ocupaban una parte importante del empleo. Según varios analistas, esto podría causar la pérdida de hasta 400.000 empleos netos en España durante los próximos nueve años. Las profesiones más

⁵²⁴ MUÑOZ VELA, J. M. “Inteligencia artificial generativa. Desafíos para la propiedad intelectual. Desafíos de la inteligencia artificial generativa para la propiedad intelectual”. Revista de Derecho UNED, núm. 33, 2024. Pág. 19.

⁵²⁵ <https://openai.com/es-ES/policies/row-terms-of-use/> (última consulta 28.11.2025)

⁵²⁶ DÍAZ-NOCI, J. “Todos los caminos llevan a Roma: propiedad intelectual, inteligencia artificial y medios de comunicación”. Anuario ThinkEPI, 2024. v. 18, e18e15. Pág. 3.

⁵²⁷ https://www.infobae.com/tecno/2025/07/02/como-la-inteligencia-artificial-transformara-el-empleo-en-espana-2-millones-de-puestos-en-riesgo/?utm_source=chatgpt.com (última consulta 28.11.2025)

vulnerables incluyen aquellas que dependen de la repetitividad: administración, traducción, atención al cliente, televenta y call centers están entre las que podrían sufrir una fuerte transformación o incluso una eliminación significativa”⁵²⁸.

A todo ello se podía unir la profunda e insalvable crisis que atraviesan profesiones clásicas como la de los traductores, salvo aquellos que prestan sus servicios de forma oficial y por el momento no pueden ser sustituidos por una IA⁵²⁹. Pero no todo son sombras:

“Es cierto que también se plantea una sorprendente dualidad, que nos refiere la transformación del mercado de trabajo y la necesidad de acometer un constante reciclaje en los trabajadores, que tendría que ser impulsada por empresas y poderes públicos:

“La llegada de la inteligencia artificial al mundo empresarial está generando una dicotomía: mientras algunas empresas están recortando personal de forma agresiva para reducir costes y adoptar la automatización, otras están contratando activamente para impulsar sus proyectos de IA. Esta tendencia refleja la compleja naturaleza del impacto de la IA en el empleo: no se trata solo de destrucción, sino también de transformación y reestructuración de las plantillas”⁵³⁰

Tampoco podemos olvidar en este punto los riesgos ambientales y energéticos derivados de la IA⁵³¹. De hecho, aunque la IA tiene la capacidad de optimizar energía en redes⁵³², lo cierto es que el consumo de electricidad para satisfacer sus necesidades tiene una dimensión creciente, con previsiones de un astronómico 0,5% del consumo

⁵²⁸ https://cincodias.elpais.com/fortunas/2025-05-11/que-trabajos-estan-mas-en-riesgo-por-culpa-de-la-ia.html?utm_source=chatgpt.com (última consulta 28.11.2025)

⁵²⁹ Ley 2/2014, de 25 de marzo, de la Acción y del Servicio Exterior del Estado. “Disposición adicional décima sexta. Traducciones e interpretaciones de carácter oficial. Reglamentariamente se determinarán los requisitos para que las traducciones e interpretaciones de una lengua extranjera al castellano y viceversa tengan carácter oficial. En todo caso, tendrán este carácter las certificadas por la Oficina de Interpretación de Lenguas del Ministerio de Asuntos Exteriores y de Cooperación, así como las realizadas por quien se encuentre en posesión del título de traductor-intérprete jurado que otorga el Ministerio de Asuntos Exteriores y de Cooperación. Los requisitos para el otorgamiento de este título, así como el resto de elementos que conforman su régimen jurídico, se desarrollarán reglamentariamente. El traductor-intérprete jurado certificará con su firma y sello la fidelidad y exactitud de la traducción e interpretación”. E igualmente la Orden AEC/2125/2014, de 6 de noviembre, por la que se dictan normas sobre los exámenes para la obtención del título de Traductor-Intérprete Jurado.

⁵³⁰ https://www.economiamundial.es/noticias/las-empresas-estan-divididas-sobre-el-papel-de-la-ia-mientras-algunas-eliminan-puestos-de-trabajo-otras-contratan-personal/?utm_source=chatgpt.com (última consulta 28.11.2025)

⁵³¹ MARTÍNEZ, L., PÉREZ, S., y TORRES, A. Desafíos Ambientales y Energéticos. *Ecoenergy Journal*, 2022. 14(1), 23-36

⁵³² SILVA, M., & COSTA, A. Incremento de la eficiencia de plantas solares mediante redes neuronales. *Journal of Solar Energy*, 2020, 25(2), 145-160

mundial para 2027⁵³³, lo que necesariamente impulsa el desarrollo de la energía nuclear para satisfacer la demanda y los acuerdos energéticos que se están adoptando por las empresas de IA. Por otro lado, también se ha destacado la posible conversión de la ciencia en pseudociencia⁵³⁴, así como en el ámbito de los consumidores⁵³⁵ entre los que cabe destacar los desequilibrios adicionales entre las partes, uso manipulador de “big data”, limitación de la independencia real de los consumidores y decisiones automatizadas que pueden perjudicar a ciertas personas y grupos sin justificación aceptable⁵³⁶, con riesgo de posible discriminación. Las incertidumbres son muchas y crecientes.

En ámbitos como la informática y el periodismo el riesgo de destrucción masiva de empleos es evidente en el seno doctrinal. “Los diseñadores web y de interfaces digitales también deberían temer la implementación de herramientas de inteligencia artificial en su trabajo diario. Los analistas de noticias, reporteros y periodistas vienen después”⁵³⁷. Estas conclusiones se ratifican en un informe de la Universidad de Indiana que manifiesta que la IA generativa no solamente tendrá efectos en el empleo, sino que incluso aumentará las diferencias sociales⁵³⁸. ¿Se irá hacia una mayor polarización de la sociedad y una progresiva extinción de la clase media occidental? Otro autores ratifican este criterio, pero incidiendo de forma más intensa en los riesgos derivados de la misma:

“La IA generativa puede tener un impacto social y económico muy positivo, pero también contribuir a amenazar la democracia, aumentar la polarización, el sesgo, los

⁵³³ <https://www.nytimes.com/2023/10/10/climate/ai-could-soon-need-as-much-electricity-as-an-entire-country.html> (última consulta, 19.8.2025)

⁵³⁴ ETXEBARRÍA ECENAVO, V. “El negocio de la inteligencia artificial amenaza con convertir la ciencia en pseudociencia”. 27.5.2024. <https://theconversation.com/el-negocio-de-la-inteligencia-artificial-amenaza-con-convertir-la-ciencia-en-pseudociencia-230341> (última consulta, 19.8.2025)

⁵³⁵ SARTOR, G. Nuevos aspectos y retos en materia de protección de los consumidores. Servicios digitales e inteligencia artificial. ESTUDIO Solicitado por la comisión IMCO. Departamento de Políticas Económicas, Científicas y de Calidad de Vida Dirección General de Políticas Internas. Abril de 2020. [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/648790/IPOL_STU\(2020\)648790_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/648790/IPOL_STU(2020)648790_EN.pdf) (última consulta, 19.8.2025)

⁵³⁶ <https://www.abogacia.es/publicaciones/blogs/blog-de-derecho-de-los-los-consumidores/inteligencia-artificial-y-proteccion-a-los-consumidores-ultimo-informe-del-parlamento-europeo/> (última consulta, 19.8.2025)

⁵³⁷ HUI, X. RESHEF, O. ZHOU, L. Los efectos a corto plazo de la inteligencia artificial generativa en el empleo: Evidencia de un mercado laboral en línea. 2023. DÍAZ-NOCI, J. “Todos los caminos llevan a Roma: propiedad intelectual, inteligencia artificial y medios de comunicación”. Anuario ThinkEPI, 2024. v. 18, e18e15. Pág. 3.

⁵³⁸ Informe del Grupo de Trabajo sobre Inteligencia Artificial Generativa (GenAI) de la Universidad de Indiana. Marzo 2024. <https://ufc.iu.edu/committees/IU-Gen-AI-Taskforce-Final-Report-3.27.24.pdf> (última consulta, 19.8.2025)

prejuicios y las desigualdades sociales, atentar contra la privacidad y otros derechos fundamentales, contra la información confidencial empresarial, la PI o incluso contra la vida humana, si no se establecen las salvaguardas necesarias”⁵³⁹.

En lo referido a la propia actuación de los entes públicos, la IA tampoco está exenta de dudas, ya que la misma necesariamente tendrá un impacto en el control ejercido sobre los ciudadanos, lo que pone en riesgo el ejercicio de los derechos del individuo:

“Esta nueva era tecnológica propicia la aparición de nuevos métodos de investigación que vienen caracterizados por una mayor eficacia pero que, al mismo tiempo, pueden llegar a provocar un grave quebranto de los derechos fundamentales de las personas”⁵⁴⁰.

Este riesgo se pudo llegar a ver intensificado por la (temporalmente frustrada) propuesta de “Chat-Control 2.0”, “Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas para prevenir y combatir el abuso sexual infantil.”, que pasó sorprendentemente desapercibida para el gran público y los medios de comunicación. Con el fin de prevenir el abuso infantil, la misma quería implantar un sistema de vigilancia y escaneo general de todos los sistemas de mensajería instantánea antes de proceder a su cifrado, con lo que finalizaría así la confidencialidad de las conversaciones y asistiríamos a nuevo y desconocido escenario en este siglo XXI, que violentaba claramente los Derechos Fundamentales de rango constitucional y eliminaba “ab initio” la supervisión judicial de la intervención, por lo que el principio de las cartas magnas de la inviolabilidad de las telecomunicaciones pasaba a ser de facto un molesto “corpore insepulto”. ¿El fin justifica siempre los medios utilizados? Según el Dictamen del Comité Económico y Social Europeo sobre la «Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas para prevenir y combatir el abuso sexual de los menores» [COM(2022) 209 final — 2022/0155 (COD)]

“La responsabilidad del rastreo recaerá sobre las plataformas y los medios sociales a los que, por lo tanto, se les podría solicitar que localizaran los materiales de abuso de menores, so pena de multas correspondientes al 6 % de su volumen de negocios mundial. Deberían utilizar la inteligencia artificial y detectar modelos de lenguaje para poder bloquear las interacciones en las que los adultos practican la captación de menores con fines sexuales. Por lo tanto, el uso adecuado de los exploradores algorítmicos es esencial para evitar

⁵³⁹ MUÑOZ VELA, J. M. “Inteligencia artificial generativa. Desafíos para la propiedad intelectual. Desafíos de la inteligencia artificial generativa para la propiedad intelectual. Revista de Derecho UNED, núm. 33, 2024. Pág. 28.

⁵⁴⁰ LLORENTE SÁNCHEZ-ARJONA, M. “Inteligencia Artificial, valoración del riesgo y derecho al debido proceso”. *Inteligencia artificial legal y Administración de Justicia*. Ed. Thomson. Navarra, 2022. Pág. 377. BORGES BLÁZQUEZ, R., *Inteligencia Artificial y proceso penal*, Ed. Aranzadi, Pamplona, 2021, p. 111.

errores que den lugar a acusaciones infundadas, ya que dichos exploradores no son infalibles”⁵⁴¹.

Pareciera haberse olvidado por parte del legislador europeo lo acaecido en España con la Ley Orgánica 1/1992, de 21 de febrero, sobre Protección de la Seguridad Ciudadana, en la que para la prevención de hechos delictivos (tarea de máxima prioridad) se pretendió combatir la luchar contra las drogas ilícitas vulnerando otros derechos, concretamente la inviolabilidad del domicilio, declarándose posteriormente inconstitucional por la STC 341/1993, del 18 de noviembre⁵⁴². Y mientras se escriben estas líneas (noviembre de 2025) Dinamarca procedió a retirar esta propuesta para combatir el abuso sexual, desconociéndose cuál puede ser su futuro a corto o medio plazo.

En materia de Propiedad Intelectual, la implantación de la tecnología siempre tiene un cierto riesgo, como ya sucedió en décadas pasadas con el fotocopiado de libros (recurso tan económico como ética y legalmente cuestionable) o más recientemente, con el escaneado de los mismos. En áreas como la música el impacto de la piratería ha sido muy elevado. Sin embargo, con la IA se da un salto cualitativo, porque el centro del conflicto ya no obra solamente en la mera reproducción ilícita de las obras (problema típico desde el desarrollo tecnológico de la segunda mitad del siglo XX) sino también en la “creación” de las mismas. Doctrinalmente se ha alertado de “la posible vulneración de derechos de autor sobre los contenidos de los que se nutren y entrenan (inputs), la protección de los resultados generados con/por los mismos (outputs)”⁵⁴³.

En la octava sesión del Diálogo de la OMPI sobre la IA generativa y la PI⁵⁴⁴:

“Gracias a su capacidad para crear contenido novedoso y realista como imágenes, música y texto, la IA generativa tiene el potencial de revolucionar múltiples industrias, incluida la creación de música, imágenes y otras formas de contenido. La IA generativa repercute en

⁵⁴¹ <https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX:52022AE2804> (última consulta 28.11.2025)

⁵⁴² “Artículo 21. 2. A los efectos de lo dispuesto en el párrafo anterior, será causa legítima para la entrada y registro en domicilio por delito flagrante el conocimiento fundado por parte de las Fuerzas y Cuerpos de Seguridad que les lleve a la constancia de que se está cometiendo o se acaba de cometer alguno de los delitos que, en materia de drogas tóxicas, estupefacientes o sustancias psicotrópicas, castiga el Código Penal, siempre que la urgente intervención de los agentes sea necesaria para impedir la consumación del delito, la huida del delincuente o la desaparición de los efectos o instrumentos del delito”.

⁵⁴³ MUÑOZ VELA, J. M. “Inteligencia artificial generativa. Desafíos para la propiedad intelectual. Desafíos de la inteligencia artificial generativa para la propiedad intelectual. Revista de Derecho UNED, núm. 33, 2024. Pág. 45.

⁵⁴⁴ <https://www.wipo.int/es/web/frontier-technologies/artificial-intelligence/index> (última consulta, 19.8.2025)

la innovación y la creación y, por consiguiente, plantea diversas cuestiones de PI, como la forma en que la PI puede proteger los modelos de IA, la interrelación entre las aportaciones de datos y las cuestiones en torno a la protección de la PI respecto de los productos de la IA”. En relación con las antedichas conclusiones de la OMPI, totalmente certeras, cabe reseñar que las polémicas derivadas del uso de esta tecnología han sido constantes en los últimos años. De hecho, la propia IA analiza certeramente los riesgos relativos a la propiedad intelectual: “La IA aprende a programar ¿El fin de la propiedad intelectual?”⁵⁴⁵

5. ANÁLISIS PRÁCTICO DE CONFLICTOS CON LA IA

La anunciada destrucción de puestos de trabajo por la implementación de la IA parece que no es un simple mito derivado de los creadores de “fake news”. De hecho, en 2023 se produjo una huelga de cinco meses del sindicato de guionistas de Estados Unidos (WGA) debido al uso de la IA en la industria cinematográfica, que con un seguimiento masivo logró paralizar totalmente la misma. Fruto de esta relevante acción, se produjeron una serie de acuerdos que propiciaron que la IA se haya configurado como elemento de apoyo, pero no fundamental para la realización de obras artísticas, así como la correlativa obligación de transparencia de informar al público del uso de la misma (lo cual, dicho sea de paso, constituye a nuestro juicio una obligación insoslayable de las empresas frente a los consumidores y el público en general para evitar fraudes). No en vano, la IA no puede ser reputada a efectos legales como “escritor”⁵⁴⁶, tal y como venimos desarrollando en estas líneas. Y en 2025, hemos asistido a hechos que parecían impensable hace pocos años, como son las sorprendentes y polémicas campañas de llamativas modelos por parte de revistas punteras anunciando prendas de vestir, realizadas íntegramente con mujeres “virtuales”, fruto de la invención de la IA, aunque bien es cierto que ello se informaba a los consumidores⁵⁴⁷. ¿Acaso las modelos ficticias creadas por IA suplantarán a las reales en un futuro próximo? ¿Existirán supermodelos virtuales con fans de carne y hueso?

“Desde la propia protección de estos sistemas y sus elementos conforme a los marcos reguladores de la PI a nivel internacional, la posible vulneración de derechos de autor

⁵⁴⁵ “Este artículo ha sido primeramente redactado por la IA Gemini Advance 1.5 Pro, según las directrices e indicaciones del ¿autor?, y posteriormente modificado de forma manual” <https://www.piconyasociados.es/noticias/la-ia-aprende-a-programar-el-fin-de-la-propiedad-intelectual/> (última consulta, 19.8.2025)

⁵⁴⁶ <https://www.wgacontract2023.org/the-campaign/summary-of-the-2023-wga-mba> (última consulta 28.11.2025)

⁵⁴⁷ <https://www.infobae.com/tecnologia/2025/08/06/modelos-hechas-con-ia-la-polemica-campana-de-vogue-que-divide-a-la-industria-de-la-moda/> (última consulta, 19.8.2025)

sobre los contenidos de los que se nutren y entrenan (inputs), la protección de los resultados generados con/por los mismos (outputs)”⁵⁴⁸

Durante año 2024, se dio a conocer una importante sentencia en el caso de “Thomson Reuters” contra “Ross Intelligence”, dictada por la Corte del Distrito de Delaware el 11 de febrero, por el uso masivo de esta última de 25.000 documentos de la editorial jurídica, infringiendo los derechos de copyright y rechazando que se haya realizado un uso justo (“fair use”) por parte de la empresa de IA, constituyendo un hito en la jurisprudencia norteamericana⁵⁴⁹. Sin embargo, en EEUU hay otros precedentes de signo contrario, como *Authors Guild v. Google* (2015) en la que se concluye que “La digitalización de libros para permitir búsquedas de texto completo es transformativa y constituye uso legítimo”. En *Authors Guild v. HathiTrust* (2014) se determina que “La creación de una base de datos de búsqueda de texto completo es transformativa”.⁵⁵⁰

Igualmente, desde el punto de vista del Derecho comparado, puede destacarse la reciente resolución de la Corte Regional de Hamburgo, Alemania [2024]: *Robert Kneschke v. LAION e.V.*, Case No. 310 O 227/23, en el que el Tribunal Regional desestimó la demanda. La única cuestión que se planteó ante la Sala fue la admisibilidad de la descarga de la imagen controvertida, que el demandado llevó a cabo con el fin de comparar el contenido de la imagen con la descripción de la imagen preexistente y crear un nuevo conjunto de datos. La Sala consideró que la descarga de la imagen en este contexto estaba cubierta por la excepción de los derechos de autor para la extracción de texto y datos con fines de investigación científica realizada por organizaciones de investigación sin ánimo de lucro (artículo 60d de la Ley alemana de derechos de autor). El demandante no logró demostrar que la excepción no era aplicable.

Entrando ya en el ámbito de lo puramente ilícito, hay posibles repercusiones civiles, derivadas del ámbito de la Ley Orgánica 1/1982, de 5 de mayo, de protección civil del derecho al honor, a la intimidad personal y familiar y a la propia imagen. Igualmente, administrativas: Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. Esta norma regula el tratamiento de datos personales, entre ellos la imagen de una persona cuando permita su identificación. El tratamiento o difusión de imágenes sin consentimiento, especialmente si se trata de menores, vulnera esta norma y puede dar lugar a sanciones por parte de la

⁵⁴⁸ MUÑOZ VELA, J. M. “Inteligencia artificial generativa. Desafíos para la propiedad intelectual. Desafíos de la inteligencia artificial generativa para la propiedad intelectual. Revista de Derecho UNED, núm. 33, 2024. Pág. 45.

⁵⁴⁹ <https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2025/02/reuters-ross-court-ruling-ai-copyright-fair-use> (última consulta, 29.11.2025)

⁵⁵⁰ RIDDER, C. y TUSHNET, R. Escrito de amicus curiae de profesores de derecho de propiedad intelectual. Caso 3:23-cv-03417-VC Documento 509-2. 2025. <https://codigo-civil.es/el-uso-legitimo-fair-use-en-la-era-de-la-ia-escrito-amicus-curiae-en-kadrey-v-meta/> (última consulta, 19.8.2025)

Agencia Española de Protección de Datos. Esta institución ha comenzado a intervenir en supuestos de generación y difusión de imágenes manipuladas con IA, imponiendo sanciones incluso a menores de edad o a sus representantes legales.

No deben olvidarse las consecuencias incluso penales (Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal, con múltiples modificaciones posteriores⁵⁵¹) ya que nuestra sociedad padece de forma recurrente la creación y difusión de vídeos de famosas desnudas creadas con IA⁵⁵² e incluso, lo que resulta todavía más inquietante, la alteración y difusión de imágenes de menores, convirtiéndolas en escenas de carácter pornográfico:

“Utilizaron aplicaciones virtuales de inteligencia artificial para obtener imágenes manipuladas de diversas menores, de tal manera que, a los rostros originales de las chicas, que obtenían de fotografías de los perfiles sociales de las mismas, les superponían las imágenes de otros cuerpos femeninos desnudos, en poses que enfatizaban los órganos sexuales, aparentando de modo realista corresponder a las propias menores. Posteriormente, las fotografías alteradas las compartieron a través de dos grupos de WhatsApp (...) todo lo cual ha ocasionado humillación y desasosiego en las afectadas, que cuentan entre 11 y 15 años de edad” (Sentencia del Juzgado de Menores de Badajoz 86/2024, 20.6.2024)⁵⁵³.

⁵⁵¹ “Artículo 189. 1. Será castigado con la pena de prisión de uno a cinco años (...) A los efectos de este Título se considera pornografía infantil o en cuya elaboración hayan sido utilizadas personas con discapacidad necesitadas de especial protección (...) a) Todo material que represente de manera visual a un menor o una persona con discapacidad necesitada de especial protección participando en una conducta sexualmente explícita, real o simulada. b) Toda representación de los órganos sexuales de un menor o persona con discapacidad necesitada de especial protección con fines principalmente sexuales. c) Todo material que represente de forma visual a una persona que parezca ser un menor participando en una conducta sexualmente explícita, real o simulada, o cualquier representación de los órganos sexuales de una persona que parezca ser un menor, con fines principalmente sexuales, salvo que la persona que parezca ser un menor resulte tener en realidad dieciocho años o más en el momento de obtenerse las imágenes. d) Imágenes realistas de un menor participando en una conducta sexualmente explícita o imágenes realistas de los órganos sexuales de un menor, con fines principalmente sexuales”.

⁵⁵² https://www.lespanol.com/omicrono/software/20250806/nueva-polemica-grok-ia-elon-musk-genera-videos-taylor-swift-desnuda-nueva-funcion-pi-cante/1003743877123_0.html (última consulta 28.11.2025)

⁵⁵³ “Imponen la medida de libertad vigilada durante un año a los 15 menores acusados de manipular y difundir imágenes de menores desnudas en Badajoz”. Sentencia del Juzgado de Menores de Badajoz de 20.7.2024. <https://www.poderjudicial.es/cgpj/es/Poder-Judicial/Noticias-Judiciales/Imponen-la-medida-de-libertad-vigilada-durante-un-ano-a-los-15-menores-acusados-de-manipular-y-difundir-imagenes-de-menores-desnudas-en-Badajoz> (última consulta 28.11.2025)

Por su parte, la STS 409/2022, de 5 de abril declara en sus hechos probados que el acusado creó un montaje digital superponiendo el rostro de una actriz conocida sobre un cuerpo desnudo femenino y lo publicó en internet con comentarios de carácter sexual manifiesto y explícito. Se consideró que esta conducta constituía una intromisión ilegítima en el derecho a la propia imagen y al honor, al amparo de la Ley Orgánica 1/1982 y el artículo 18 CE, subrayando que la manipulación digital que hace reconocible a una persona en un contexto degradante supone una lesión a su dignidad y reputación, incluso cuando el cuerpo o la escena no sean reales. La mera fantasía de esos contenidos no excluye en modo alguno la responsabilidad del autor. Y condena a una indemnización de 30.000 euros por daños morales, eliminación de las imágenes y publicación del fallo en dos medios de comunicación de ámbito nacional. Aunque dicho todo lo anterior, se plantea otro problema. ¿Realmente es posible proceder a un borrado definitivo y absoluto de imágenes creadas a través de la IA en internet? ¿O aquella imagen que se subió a internet siempre permanecerá en la red de una u otra forma?

Igualmente la SAP de Almería 202/2025, de 5.5.2025:

“Una vez le fue intervenido al acusado su teléfono (...) tras la correspondiente autorización judicial se procedió al volcado y análisis del dispositivo, comprobándose que en el terminal había 4130 archivos de pornografía, algunos de ellos editados con inteligencia artificial, siendo mayoritariamente de naturaleza pedófila, puesto que aparecían menores de edad sin ropa posando de forma sugerente o participando entre sí o con adultos en conductas explícitas de contenido sexual (...) y además hacer fotomontajes a través de la inteligencia artificial, editándolas para que esas personas cercanas al acusado apareciesen junto a cuerpos ajenos desnudos y participando de explícitos actos de naturaleza sexual”.

Imágenes censurables que debe remarcarse han sido creadas en muchos casos por amistades o compañeros igualmente menores de edad, habida cuenta la inmensa facilidad para la creación de las mismas, sobre todo cuando la inexistencia de límites éticos brilla por su ausencia. Y dicho sea de paso, el hecho de encontrarnos en un contexto “embrionario” y con escaso o nulo control, no favorece la limitación de estas conductas. Todo ello derivado de la máxima “versatilidad delictiva” propia del medio, de la cual la doctrina más autorizada se venía haciendo eco hace décadas, antes de la eclosión total de la era de las telecomunicaciones⁵⁵⁴. En este contexto, las IA programadas sin

<https://www.interior.gob.es/opencms/eu/detalle/articulo/Investigado-un-menor-por-modificar-y-difundir-imagenes-utilizando-IA-de-sus-companeras-de-instituto-en-Valencia/> (última consulta, 19.8.2025)

⁵⁵⁴ MARCHENA GÓMEZ, M. (2000) “Prevención de la delincuencia tecnológica”. Derecho de Internet. Contratación Electrónica y Firma Digital. (Coord. Mateu de Ros, R. y Cendoya Méndez de Vigo, J.M.) Aranzadi. Navarra. Pág. 447.

un elevado contenido ético convierten en potenciales víctimas a todos los miembros de la sociedad, pasados, presentes y futuros.

5.1.1. LA IA DENTRO DEL PROCESO JUDICIAL: HERRAMIENTA NO SUSTITUTIVA DE LA LABOR DE LAS PARTES

El uso de la IA requiere ciertas habilidades por parte del usuario, para poder recibir un resultado lo más certero posible, de ahí que las instrucciones genéricas generen no solamente un riesgo cierto de confusión, sino incluso una posible responsabilidad profesional cuando se utiliza la IA con un exceso de confianza y sin revisar debidamente sus resultados o conclusiones.

En 2025 la IA tiene una función complementaria (que no sustitutiva) de la labor profesional, con limitaciones técnicas y riesgo de uso de fuentes incorrectas, lo cual se anuncia oportunamente por parte de las plataformas. Se abre aquí nuevamente el debate sobre la diligencia profesional de las profesiones jurídicas, lo que ha propiciado que desde los Colegios Profesionales:

“La IA debe ser vista como una aliada tecnológica sujeta a límites, por lo que el profesional de la abogacía debe mantener en todo momento la supervisión, el juicio crítico y una diligencia reforzada en su selección, integración y uso. Falta de comprensión jurídica real: Aunque la IA puede imitar la redacción jurídica, no entiende los conceptos legales ni el contexto de cada caso. Esto significa que puede producir errores o interpretaciones inexactas en materia legal. Estas limitaciones obligan al abogado/a a realizar siempre una revisión crítica y sistemática de los resultados generados. Propensión a errores y sesgos: Los modelos pueden replicar sesgos o errores presentes en los datos con los que fueron entrenados. Además, la IA generativa puede producir “alucinaciones”, es decir, información inexacta o inventada que aparenta ser correcta”⁵⁵⁵.

El Tribunal Superior de Justicia de Navarra (TSJNA) emitió en 2024 una resolución pionera en España, analizando las implicaciones legales y éticas del uso de la IA generativa en el ámbito judicial. El auto del TSJ de Navarra 2/2024, de fecha 4.9.2024 (Rec. núm. 17/2024), aborda un caso en el que un letrado uso información errónea (citando normativa de otro país) proporcionada por ChatGPT para la redacción de una querrela. Eso nos lleva a la rápida conclusión de que el uso en 2025 de la IA sin posterior y profunda revisión humana, resulta una práctica altamente temeraria, contraria al estatuto profesional. No en vano, la IA resulta un complemento, no un sustitutivo de

⁵⁵⁵ Guía ICAM de buenas prácticas para el uso de la Inteligencia Artificial (IA) en la Abogacía https://web.icam.es/wp-content/uploads/2025/10/Guia-ICAM_IA_2025_2.pdf (última consulta, 29.11.2025)

la actividad de los profesionales⁵⁵⁶. Y se corre siempre el riesgo de que al solicitar el usuario respuestas a la IA, la misma proceda a darlas sin mayores planteamientos, ya que supuestamente desconoce su error, incorrección o inadecuación al caso planteado. Probablemente debería reforzarse en este punto la información al usuario sobre esta circunstancia, con unas condiciones generales de la contratación por las que se regule el servicio bien visibles y accesibles para el ciudadano.

“El letrado tiene el deber de cotejar celosamente las fuentes que proporciona al tribunal, máxime si recurrió a un sistema de inteligencia artificial generativa, herramienta experimental aún en desarrollo. Y subraya que es ampliamente conocida la circunstancia de que, por esa misma razón, los resultados que ofrecen incluyen aún y en muchos casos las denominadas ‘alucinaciones’”⁵⁵⁷. “La firma de un escrito judicial implica siempre una asunción de responsabilidad: quien lo presenta debe haber verificado la corrección fáctica y normativa de su contenido. Si un letrado incorpora en su trabajo un texto generado por inteligencia artificial que contiene errores o citas inexistentes, la responsabilidad no puede trasladarse a la herramienta. Sigue siendo indelegable y recae en el profesional”⁵⁵⁸

En el Tribunal Superior de Inglaterra y Gales, el caso *Ayinde v. Haringey* reveló cómo un abogado junior incluyó cinco precedentes inexistentes en su escrito judicial. De forma similar, en *Al-Haroun v. Qatar National Bank* se detectaron dieciocho citas falsas, entre ellas una supuesta resolución atribuida al propio juez del proceso. Estos errores fueron consecuencia directa del uso acrítico de herramientas de inteligencia artificial generativa sin la debida verificación humana, lo que vulnera los principios básicos de diligencia profesional y responsabilidad en la práctica jurídica.

El fallo, dictado por la jueza Victoria Sharp y el juez Jeremy Johnson, constituye un precedente relevante en el ámbito anglosajón. La resolución establece que las herramientas de IA no pueden sustituir el criterio profesional ni la obligación de contrastar la información antes de presentarla ante el tribunal. En consecuencia, se ordenaron medidas correctivas inmediatas para evitar la repetición de estas prácticas, reforzando la idea de que la IA debe ser utilizada únicamente como apoyo y nunca como reemplazo del análisis jurídico humano⁵⁵⁹.

Cabe recordar aquí la Ley Orgánica 5/2024 de Derecho de Defensa, la cual señala expresamente en su Art. 13:

⁵⁵⁶ MANSO, P. ¿Puede una solución de IA* ejercer como abogado en España? <https://www.patriciamanso.com/post/puede-una-soluci%C3%B3n-de-ia-ejercer-como-abogado-en-espa%C3%B1a> (última consulta, 19.8.2025)

⁵⁵⁷ <https://documento.errepar.com/actualidad/ia-en-escritos-judiciales-responsabilidad-profesional-del-abogado-20250825192349644> (última consulta, 4.12.2015)

⁵⁵⁸ <https://revistaquorum.com.ar/2025/11/04/la-responsabilidad-profesional-de-los-abogados-frente-al-uso-de-la-inteligencia-artificial/> (última consulta, 4.12.2015)

⁵⁵⁹ <https://elartesano.digital/uso-ia-derecho-riesgos-etica-fallos/> (última consulta, 4.12.2015)

“La asistencia letrada será prestada por los profesionales de la abogacía, que son aquellas personas que, por cuenta propia o ajena, estando en posesión del título profesional regulado en la normativa sobre el acceso a las profesiones de la abogacía y la procura, están incorporadas a un colegio de la abogacía como ejercientes y se dedican de forma profesional al asesoramiento jurídico, a la solución de conflictos y a la defensa de derechos e intereses ajenos, tanto públicos como privados, en la vía judicial o extrajudicial”.

Ello se remarca con la firma digital de los correspondientes escritos presentados en sede judicial o administrativa, puesto que como resulta obvio, la IA carece (o debería carecer en todo caso) de la posibilidad de firmar escritos.

En Argentina resulta especialmente relevante la Sentencia de la Cámara Penal de Esquel, Provincia de Chubut. Causa: Provincia de Chubut c/ P.R.A., carpeta judicial n.º 6209 El juez de primera instancia condenó a R.A.P. por robo simple. La Cámara Penal advirtió que el juez había utilizado un asistente de inteligencia artificial generativa sin informar a las partes ni dejar constancia en el expediente, hallando frases generadas por IA (“Aquí tienes el punto IV reeditado, sin citas y listo para copiar y pegar”). La Cámara Penal declaró la nulidad de la sentencia condenatoria, al considerar que el uso no declarado de IA vulneró las garantías procesales básicas y el principio del juez natural. El fallo establece un precedente inédito en América Latina: el uso no transparente de inteligencia artificial en la redacción judicial puede provocar la nulidad de la sentencia. Se reafirma que la IA puede ser una herramienta auxiliar, pero no puede sustituir ni compartir la función decisoria del juez.

La Corte Suprema de Colombia analizó otro caso análogo al anterior: “El Tribunal de Sincelejo había revocado una decisión de primera instancia que negó la terminación del proceso por desistimiento tácito. Para justificar la revocatoria, citó apartes de las sentencias STC13560-2023 y STC4734-2025, supuestamente emitidas por la Sala de Casación Civil. Sin embargo, al ser revisadas por la Corte Suprema, se comprobó que “los párrafos transcritos por el Tribunal accionado son inexistentes, pues no forman parte de la jurisprudencia que utilizó para resolver el problema jurídico”.

6. CONCLUSIONES

PRIMERA.- Ante esta situación, se plantea la necesidad de establecer controles al desarrollo de la tecnología, tal y como se planteó con los inicios de la robótica en 1942. Bajo nuestro punto de vista, los límites que se puedan configurar sobre la IA no han tener la consideración de “censura previa” ni “autocensura” y se ha de fomentar un uso responsable por parte del gran público. No todo vale y la IA se ha de encuadrar en el marco normativo existente, que por supuesto puede mutar pero no someterse a las necesidades de aquélla.

SEGUNDA.- Paradójicamente nos encontramos con IA que utilizan como reclamo publicitario la absoluta falta de cortapisas, que pueden generar fuertes disfunciones e incluso provocar con cierta facilidad ilícitos de índole penal: “IA sin censura. Venecia ofrece los modelos más avanzados, precisos y sin censura para una experiencia de inteligencia artificial verdaderamente sin restricciones”⁵⁶⁰. “Lo Nuevo: ¡Ahora con Imágenes Sin Censura! La última y más avanzada IA Sin Censura (2024). No se requiere iniciar sesión, 100% privado, Velocidad Turbo. Confiado por más de 10,000 usuarios en todo el mundo”⁵⁶¹. En este punto la pregunta del millón es si realmente un usuario necesita una IA ilimitada para disfrutar de su “experiencia”. ¿Iremos hacia un horizonte digital con una “deep web” con IA clandestinas que permitan la comisión de toda clase de ilícitos?

TERCERA.- La responsabilidad social corporativa no puede ser ajena a esta nueva realidad y se han de exigir la implementación de Códigos de Buenas prácticas en las que se incluya expresamente y sin demora la IA, con claros protocolos de actuación, tanto para creadores de IA como para usuarios profesionales de la misma: “La industria de la IA debe operar de manera respetuosa con la propiedad intelectual e industrial, y garantizar los derechos de los autores, por lo que debemos ir hacia un modelo de gestión donde todas las partes involucradas ganen, esto es, las empresas de IA, editores y autores”⁵⁶². Aunque bien es cierto que la máxima de que la realidad social siempre se anticipa al Derecho, todos los agentes implicados deberían realizar un especialísimo esfuerzo de adecuación.

BIBLIOGRAFÍA

- ANGUIANO, J. M. (2023) Inteligencia artificial y copyright. Del dilema del thaler a la doctrina “the right to read is the right to mine”. Instituto de Derechos de Autor. Madrid, 2023.
- ARGANDOÑA, A. Ética e inteligencia artificial (I). IESE Blog Network: Economía, Ética y RSE. 2019. Recuperado de blog.iese.edu/antonioargandoña/2019/03/25/etica-e-inteligencia-artificial-i.
- ASIMOV, I. «Círculo vicioso». Los robots. trad. Domingo Santos. Barcelona: Ed. Martínez Roca. Barcelona, 1989.
- BEIRO MAGÁN, J. M. “Automatización e Inteligencia Artificial: por qué debe generalizarse su implantación en el sector justicia español”<https://pensamientocritico.sisej.com/automatizacion-e-inteligencia-artificial-por-que-debe-generalizarse-su-implantacion-en-el-sector-justicia-espanol/>

⁵⁶⁰ <https://venice.ai/es> (última consulta, 19.8.2025)

⁵⁶¹ <https://bai.tools/es/tools/chat-uncensored-ai> (última consulta 28.11.2025)

⁵⁶² MUÑOZ VELA, J. M. “Inteligencia artificial generativa. Desafíos para la propiedad intelectual. Generative artificial intelligent challenges for intellectual property”. Revista de Derecho UNED, núm. 33, 2024. Pág. 45.

- BORGES BLÁZQUEZ, R., *Inteligencia Artificial y proceso penal*, Ed. Aranzadi, Pamplona, 2021.
- DÍAZ-NOCI, J. “Todos los caminos llevan a Roma: propiedad intelectual, inteligencia artificial y medios de comunicación”. *Anuario ThinkEPI*, 2024. v. 18, e18e15. Pág. 3.
- ETXEBARRÍA ECENAVO, V. “El negocio de la inteligencia artificial amenaza con convertir la ciencia en pseudociencia”. 27.5.2024. <https://theconversation.com/el-negocio-de-la-inteligencia-artificial-amenaza-con-convertir-la-ciencia-en-pseudociencia-230341>
- GONZÁLEZ DE POSADA, F. y GONZÁLEZ, REDONDO, F. A. Leonardo Torres Quevedo (1852-1936) 1a Parte. Las máquinas algébricas. *LA GACETA DE LA RSME*, Vol. 7.3 (2004), Págs. 787–810
- HUI, X. RESHEF, O. ZHOU, L. Los efectos a corto plazo de la inteligencia artificial generativa en el empleo: Evidencia de un mercado laboral en línea. 2023.
- INFORME DEL GRUPO DE TRABAJO SOBRE INTELIGENCIA ARTIFICIAL GENERATIVA (GenAI) DE LA UNIVERSIDAD DE INDIANA. Marzo 2024. <https://ufc.iu.edu/committees/IU-Gen-AI-Taskforce-Final-Report-3.27.24.pdf>
- LORENZO PÉREZ, C. “Inteligencia artificial en la Administración de Justicia: regulación española y marco europeo e internacional. Proyectos desarrollados por el Ministerio de Justicia de España”. Ed. Centro de Estudios Jurídicos. Madrid, 2022.
- LLORENTE SÁNCHEZ-ARJONA, M. “Inteligencia Artificial, valoración del riesgo y derecho al debido proceso”. *Inteligencia artificial legal y Administración de Justicia*. Ed. Thomson. Navarra, 2022.
- MANSO, P. ¿Puede una solución de IA* ejercer como abogado en España? <https://www.patriciamanso.com/post/puede-una-soluci%C3%B3n-de-ia-ejercer-como-abogado-en-espa%C3%B1a>
- MARCHENA GÓMEZ, M. (2000) “Prevención de la delincuencia tecnológica”. *Derecho de Internet. Contratación Electrónica y Firma Digital*. (Coord. Mateu de Ros, R. y Cendoya Méndez de Vigo, J.M.) Aranzadi. Navarra. Pág. 447.
- MARÍN GARCÍA, S. Ética e inteligencia artificial. *Cuadernos de la Cátedra CaixaBank de Responsabilidad Social Corporativa*. N° 42 Septiembre de 2019. DOI: <https://dx.doi.org/10.15581/018.ST-522>
- MARTÍN ALÁEZ, F. La tutela civil de los derechos de propiedad intelectual en el mercado único digital (entre plataformas e inteligencia artificial) Tesis doctoral UDC / 2024. Pág. 671. <https://ruc.udc.es/rest/api/core/bitstreams/9bb06547-ef7a-44d3-9eb5-8a509bae0592/content>
- MARTÍNEZ, L., PÉREZ, S., y TORRES, A. Desafíos Ambientales y Energéticos. *Ecoenergy Journal*, 2022. 14(1), 23-36
- MCCARTHY, J.; MINSKY, M.; ROCHESTER, N.; SHANNON, C. Propuesta para el Proyecto de Investigación de Verano de Dartmouth sobre Inteligencia

- Artificial. 31.08.1955. Recuperado de: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- McCULLOCH, W. & PITTS, W. Cálculo lógico de las ideas inmanentes en la actividad nerviosa. Boletín de Biofísica Matemática, vol. 5, pp. 115-133 (1943).
- MUÑOZ VELA, J. M. “Inteligencia artificial generativa. Desafíos para la propiedad intelectual. Desafíos de la inteligencia artificial generativa para la propiedad intelectual”. Revista de Derecho UNED, núm. 33, 2024.
- RIDDER, C. y TUSHNET, R. Escrito de amicus curiae de profesores de derecho de propiedad intelectual. Caso 3:23-cv-03417-VC Documento 509-2. 2025. <https://codigo-civil.es/el-uso-legitimo-fair-use-en-la-era-de-la-ia-escrito-amicus-curiae-en-kadrey-v-meta/>
- ROQUE, M. y PALMA, J. Inteligencia Artificial. Técnicas, métodos y aplicaciones. Madrid, 2008. McGraw-Hill.
- RUSSELL S. J., NORVIG, P. Inteligencia artificial. Un enfoque moderno. Englewood Cliffs, NJ: Prentice Hall, 1994.
- SAMUEL, A. L. Algunos estudios sobre el aprendizaje automático utilizando el juego de damas. Revista IBM. Julio de 1959. Pág. 210 y ss.
- SARTOR, G. Nuevos aspectos y retos en materia de protección del consumidor. Servicios digitales e inteligencia artificial. ESTUDIO Solicitado por la comisión IMCO. Departamento de Políticas Económicas, Científicas y de Calidad de Vida Dirección General de Políticas Internas. Abril 2020. [https://www.euro-parl.europa.eu/RegData/etudes/STUD/2020/648790/IPOL_STU\(2020\)648790_EN.pdf](https://www.euro-parl.europa.eu/RegData/etudes/STUD/2020/648790/IPOL_STU(2020)648790_EN.pdf)
- SHAPIRO, S. C. Enciclopedia de inteligencia artificial (2.^a ed.). John Wiley & Sons, Inc., Nueva York, 1992.
- SILVA, M., & COSTA, A. Incremento de la eficiencia de plantas solares mediante redes neuronales. Journal of Solar Energy, 2020, 25(2), 145-160
- TURING, A. M. Sobre los números computables, con una aplicación al Entscheidungsproblem. Actas de la Sociedad Matemática de Londres. 1936.
- UNESCO, Recomendación sobre la ética de la IA. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura. París (Francia). 2022. <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>

CAPÍTULO 13.

CUESTIONES DE PROPIEDAD INTELECTUAL E INDUSTRIAL VINCULADAS A LA INTELIGENCIA ARTIFICIAL. ESPECIAL MENCIÓN AL CAMPO DE LA SALUD⁵⁶³

JUAN E. SUÑÉ CANO

Profesor asociado de Derecho Internacional Público⁵⁶⁴

(Universidad Complutense de Madrid)

jsune@ucm.es

Resumen

La aparición de la Inteligencia Artificial (IA), especialmente la IA generativa, ha planteado retos jurídicos al confrontar un sistema de Propiedad Intelectual centrado en la autoría humana con tecnologías capaces de generar contenidos autónomamente. Los contenidos creados íntegramente por IA, sin intervención creativa sustancial de personas naturales, carecen de autoría legal y entran en el dominio público. La IA desafía nociones clave como originalidad y autoría y plantea riesgos por el uso de datos protegidos en su entrenamiento (*data scraping*). En Propiedad Industrial, la participación de la IA en soluciones técnicas choca con la exigencia legal de que el inventor sea una persona natural. El sector sanitario evidencia los beneficios de la IA en eficiencia y calidad de atención, pero también riesgos de sesgo, privacidad y responsabilidad. La regulación europea, como el AI

⁵⁶³ El presente trabajo se realiza en el marco de las actividades propias del *Grupo de Investigación, Derechos Humanos ante el reto de la Inteligencia Artificial: creatividad y arte* (IAIURIS) de la Universidad Rey Juan Carlos, y en el transcurso de las Jornadas “Derechos Humanos e Inteligencia Artificial: una aproximación jurídica multidisciplinar”, mediante las Ayudas a Congresos, Jornadas y Seminarios 2025 del Vicerrectorado de Comunidad, Campus, Cultura y Deporte de la Universidad Rey Juan Carlos cuyas Actas se encuentran publicadas en <https://burjcdigital.urjc.es/server/api/core/bitstreams/293e9da5-3b08-4459-850b-ea8b86cec8e8/content>

⁵⁶⁴ Abogado especialista en asesoramiento internacional. Es Doctor en derecho, cuya tesis doctoral obtuvo la Calificación de *Cum Laude*, por la Universidad Complutense de Madrid donde se licenció. Cursó la primera edición del Máster de Analista de Inteligencia impartido conjuntamente por la Universidad Rey Juan Carlos y Carlos III, ubicado en diversos rankings internacionales como uno de los 50 masters más prestigiosos a escala global, así como el Máster en Relaciones Internacionales Iberoamericanas en la Universidad Rey Juan Carlos. En el campo profesional además de abogado ejerciente desde 2010, trabajó 9 años como Asesor Legal de la Agregaduría Legal para Europa de la Procuraduría General de la República de México en la persecución de delitos internacionales en más de 45 países y es docente en diversos centros universitarios en España y en el extranjero.

Act (Reglamento UE 2024/1689), busca garantizar transparencia, trazabilidad y cumplimiento de los derechos de Propiedad Intelectual.

Palabras clave

Inteligencia Artificial, Propiedad Intelectual, Propiedad Industrial, IA generativa, autoría, originalidad, salud, *data scraping*, AI Act, patentes.

Abstract

The emergence of Artificial Intelligence (AI), particularly generative AI, has posed legal challenges by confronting a traditionally human-centered Intellectual Property system with technologies capable of autonomously generating content. Content created entirely by AI, without substantial creative input from natural persons, lacks legal authorship and enters the public domain. AI challenges key concepts such as originality and authorship and raises risks related to the use of protected data in its training (*data scraping*). In Industrial Property, AI's participation in technical solutions conflicts with the legal requirement that inventors be natural persons. The healthcare sector illustrates AI's benefits in efficiency and quality of care, but also its risks, including bias, privacy issues, and accountability gaps. European regulation, such as the AI Act (Regulation (EU) 2024/1689), aims to ensure transparency, traceability, and compliance with Intellectual Property rights.

Key words

Artificial Intelligence, Intellectual Property, Industrial Property, Generative AI, Authorship, Originality, Health, Data Scraping, AI Act, Patents.

Sumario: INTRODUCCIÓN 1. APROXIMACIÓN AL CONCEPTO DE PROPIEDAD INTELECTUAL E INDUSTRIAL 2. LA INTELIGENCIA ARTIFICIAL, CONCEPTOS Y ESPECIFICACIONES DE LA INTELIGENCIA ARTIFICIAL GENERATIVA 2.1. Tipos de inteligencia artificial 2.2. El entrenamiento y alimentación de la Inteligencia Artificial 2.3. Redes Generativas Antagónicas 3. INTELIGENCIA ARTIFICIAL GENERATIVA Y EL RÉGIMEN JURÍDICO DE LA PROPIEDAD INTELECTUAL (DERECHO DE AUTOR) 3.1. El principio antropocéntrico de la autoría: núcleo dogmático del sistema 3.2. La naturaleza jurídica de los prompts expresión, idea o instrucción técnica 3.3. Obras generadas por IA: originalidad, reproducción y riesgos jurídicos 3.4. Autoría y titularidad de derechos en obras generadas o asistidas por IA 3.5. La Respuesta de las legislaciones de Propiedad Intelectual al Desarrollo de la IA 4. ALGUNOS ASPECTOS DE PROPIEDAD INDUSTRIAL E INTELIGENCIA ARTIFICIAL 5. INTELIGENCIA ARTIFICIAL EN EL ÁMBITO DE LA SALUD, LA BALANZA DE RIESGOS Y BENEFICIOS 5.1. Oportunidades Clave en el Sistema Sanitario 5.2. Aceleración de la Investigación y el Desarrollo de Vacunas 5.3. Riesgos y Desafíos Críticos en la Implementación de la IA 5.4. IA para la Aceptación Pública y la Comunicación 5.5. IA en el ámbito radiológico 6. CONCLUSIONES. APÉNDICE BIBLIOGRÁFICO

Abreviatura	Significado Completo
AI Act	Reglamento (UE) 2024/1689 (Legislación integral sobre inteligencia artificial de la Unión Europea)
AI Board	Órgano de coordinación creado bajo el AI Act
ARNm	(Plataformas de) ARNm (mencionado en el contexto de las vacunas)
CNN	Redes neuronales convolucionales
CPC	Clasificación para solicitudes
DL	Deep Learning (Aprendizaje Profundo)
European AI Office	Oficina Europea de Inteligencia Artificial
GAN	Redes Generativas Antagónicas (Generative Adversarial Networks)
IA	Inteligencia Artificial
IAD	Inteligencia Artificial Débil
IAG	Inteligencia Artificial General
IPC	Clasificación para solicitudes
LLM	(Grandes) Modelos de Lenguaje
LPI	Ley de Propiedad Intelectual (o normativa española de Propiedad Intelectual)
ML	Machine Learning (Aprendizaje Automático)
OCDE	Organización para la Cooperación y el Desarrollo Económicos
OEPM	Oficina Española de Patentes y Marcas
OMPI	Organización Mundial de la Propiedad Intelectual
PI	Propiedad Intelectual
PII	Propiedad Industrial
PLN	Procesamiento del Lenguaje Natural
RNN	Redes (neuronales) recurrentes
SVM	Máquinas de vectores de soporte
TRLPI	Texto Refundido de la Ley de Propiedad Intelectual
VAE	Autocodificadores variacionales

INTRODUCCIÓN

La irrupción de la Inteligencia Artificial (IA), en particular en su vertiente generativa, ha suscitado un debate profundo e ineludible en el ámbito jurídico contemporáneo. No se trata de una cuestión menor ni pasajera, sino de la consecuencia lógica de la confrontación entre un sistema normativo construido para proteger la creatividad y la aportación humana —basado en siglos de tradición, jurisprudencia y doctrina— y una tecnología capaz de generar contenidos sin intervención humana directa, reproduciendo estilos, identificando patrones y produciendo resultados que, en ocasiones, resultan prácticamente indistinguibles de los elaborados por un autor humano.

El extraordinario avance experimentado por la IA en los últimos años ha consolidado su presencia en múltiples sectores, entre ellos la economía, la cultura, el derecho y, de manera especialmente significativa, la salud pública y la innovación biomédica. Su irrupción ha transformado los procesos de creación, difusión y explotación de obras y conocimientos, generando tensiones y desafíos sustanciales respecto de los marcos de Propiedad Intelectual e Industrial, diseñados originalmente para regular la actividad creativa humana en un entorno analógico y, posteriormente, digital, pero no para gestionar la autonomía funcional de sistemas artificiales.

Un ejemplo paradigmático de este potencial transformador se evidenció durante el desarrollo acelerado de las vacunas contra la COVID-19. La capacidad de acortar plazos convencionales de años a solo meses demostró el valor estratégico de la IA en la vacunología moderna. No obstante, dicha aceleración fue resultado de una convergencia de factores —como la financiación extraordinaria, la existencia previa de plataformas de ARNm y una cooperación científica global sin precedentes— y no exclusivamente de la IA. A pesar de ello, su contribución en ámbitos como la predicción de estructuras proteicas, el cribado de candidatos vacunales o la optimización de ensayos clínicos evidenció la necesidad de analizar su impacto desde una perspectiva jurídica, ética y regulatoria integral.

La investigación y el desarrollo de vacunas, tradicionalmente estructurados en ciclos de décadas, se enfrentan hoy a la complejidad de procesos como el descubrimiento de antígenos, la identificación de epítomos, la evaluación de seguridad y eficacia y el diseño metodológico de ensayos clínicos. Todos estos procesos plantean interrogantes relevantes en materia de responsabilidad, protección de datos, seguridad farmacológica y, de manera destacada, propiedad intelectual e industrial. La intervención de sistemas de IA en tales etapas requiere revisar la configuración de la autoría, la titularidad de los derechos y el

alcance de la protección jurídica aplicable a los resultados generados o facilitados mediante estas tecnologías.

Estos desafíos se observan igualmente en el ámbito más amplio de la IA generativa, cuya capacidad para producir textos, imágenes, música, modelos biomédicos o incluso invenciones cuestiona la vigencia de conceptos jurídicos clásicos como la originalidad, la autoría o la titularidad de derechos. La proliferación de herramientas autónomas plantea interrogantes acerca de la eventual infracción de derechos preexistentes —por ejemplo, durante los procesos de entrenamiento basados en grandes volúmenes de datos—, así como sobre la adecuación del marco normativo vigente para regular los resultados generados por IA.

De manera paralela, la IA se ha consolidado en el sector salud como una herramienta esencial para afrontar desafíos estructurales, como el agotamiento de la fuerza laboral, el envejecimiento de la población y la creciente prevalencia de enfermedades crónicas. Su impacto abarca desde la mejora de la atención clínica diaria hasta la aceleración de los procesos de innovación biomédica y el desarrollo de nuevas tecnologías sanitarias. Este escenario exige una reflexión jurídica profunda orientada a identificar los límites, obligaciones y garantías necesarias para compatibilizar la innovación tecnológica con la protección de derechos fundamentales, la seguridad jurídica y la integridad de los sistemas de propiedad intelectual e industrial.

La expansión transversal de la IA ha motivado la intervención de organismos internacionales, autoridades regulatorias y operadores jurídicos, quienes coinciden en la urgencia de revisar y actualizar las normas aplicables. El objetivo consiste en alcanzar un equilibrio adecuado entre la promoción responsable de la innovación tecnológica y la protección efectiva de los derechos de propiedad intelectual, garantizando un marco jurídico claro, predecible y adaptable a los nuevos desafíos que plantea la digitalización avanzada y la creciente autonomía de los sistemas artificiales.

En consecuencia, resulta imprescindible promover un análisis sistemático y multidisciplinar que permita redefinir el alcance de la propiedad intelectual en este nuevo paradigma tecnológico, asegurando que la evolución normativa acompañe la transformación digital sin menoscabar la seguridad jurídica, la protección de los creadores y el impulso legítimo a la innovación en ámbitos como la salud, la ciencia y la industria.

1. APROXIMACIÓN AL CONCEPTO DE PROPIEDAD INTELECTUAL E INDUSTRIAL.

Las ideas, cuando no se materializan adecuadamente o carecen de una vía de protección, corren el riesgo de diluirse o de ser apropiadas y utilizadas por terceros que no han intervenido en su concepción ni desarrollo. Precisamente para evitar estos riesgos surgen los sistemas de protección de la propiedad intelectual e industrial, cuyo fin es reconocer derechos exclusivos sobre las creaciones y garantizar que no puedan ser explotadas sin autorización.

Para poder obtener beneficios y salvaguardar de forma efectiva los activos intangibles —ya sean obras, invenciones, diseños o signos distintivos— es imprescindible, en primer lugar, identificar con claridad aquello que constituye nuestra propiedad intelectual o industrial y, posteriormente, utilizar los mecanismos legales adecuados para su registro o protección. Las ideas, por sí mismas, no gozan de tutela jurídica; solo adquieren protección cuando se concretan o cuando encajan dentro de las categorías reconocidas por la ley. Los derechos de propiedad, en este ámbito, funcionan de manera análoga a cualquier otro derecho de dominio: facultan al creador o al titular de una patente, marca o derecho de autor a disfrutar de los beneficios que se deriven de su creación o de la inversión que haya realizado en ella.

Este reconocimiento encuentra apoyo en el artículo 27 de la Declaración Universal de Derechos Humanos⁵⁶⁵, que establece el derecho de toda persona tanto a participar libremente en la vida cultural y científica como a recibir protección sobre los intereses morales y materiales derivados de sus producciones literarias, artísticas o científicas.

La relevancia de la propiedad intelectual quedó reflejada tempranamente en instrumentos internacionales como el Convenio de París para la Protección de la Propiedad Industrial (1883) y el Convenio de Berna para la Protección de las Obras Literarias y Artísticas (1886), actualmente administrados por la Organización Mundial de la Propiedad Intelectual (OMPI). Estos tratados parten de la premisa de que el progreso científico, tecnológico y cultural depende de la capacidad humana para crear e innovar; de que la existencia de una protección jurídica eficaz incentiva la inversión en investigación y desarrollo; y de que la promoción de la propiedad intelectual impulsa el crecimiento económico, fomenta la creación de empleos y contribuye a mejorar la calidad de vida.

Un sistema de propiedad intelectual eficiente y equilibrado permite que los Estados aprovechen el potencial de estos derechos como motores de

⁵⁶⁵ Artículo 27. DUDH: 1. Toda persona tiene derecho a tomar parte libremente en la vida cultural de la comunidad, a gozar de las artes y a participar en el progreso científico y en los beneficios que de él resulten.

2. Toda persona tiene derecho a la protección de los intereses morales y materiales que le correspondan por razón de las producciones científicas, literarias o artísticas de que sea autora.

desarrollo económico y social, a la vez que asegura un punto de equilibrio entre los intereses privados de los innovadores y el interés público. De este modo se genera un entorno propicio para que la creatividad y la invención prosperen en beneficio general.

Los derechos de propiedad intelectual, por tanto, actúan como catalizadores de creatividad y esfuerzo innovador. Sin la protección del derecho de autor, industrias tan relevantes como el cine, la música, la edición o la informática difícilmente habrían alcanzado su dimensión actual. Del mismo modo, sin la garantía de las patentes, los investigadores carecerían del estímulo necesario para continuar desarrollando nuevos y mejores productos. Y sin sistemas eficaces y de alcance internacional para la protección de las marcas y la lucha contra la falsificación y la piratería, los consumidores no podrían confiar en la autenticidad ni en la calidad de los bienes y servicios que adquieren.

La Propiedad Intelectual se entiende como toda creación original de carácter literario, artístico o científico, manifestada en cualquier clase de medio o soporte, ya sea físico o digital, existente o aún por desarrollarse. El formato en que una obra se exprese o difunda no altera la naturaleza de la creación ni su condición jurídica.

En los sistemas jurídicos de tradición anglosajona, la expresión *Intellectual Property* engloba tanto la propiedad intelectual como la propiedad industrial. En cambio, en los países de derecho continental europeo —como España— ambas ramas se consideran ámbitos diferenciados: la propiedad industrial comprende patentes, marcas, indicaciones geográficas, diseños industriales y obtenciones vegetales, mientras que la propiedad intelectual se refiere tanto a los derechos de autor como a los derechos conexos.

Así, la Propiedad Intelectual abarca aquellas obras que pueden ser amparadas por derechos de autor debido a su carácter original —esto es, que no derivan de copias o reproducciones ajenas— y a su concreción en un soporte que permita su percepción, reproducción o comunicación.

La autoría de una obra literaria, artística o científica surge desde el instante mismo de su creación. A partir de ese momento, el autor —y los posteriores titulares de derechos— queda plenamente protegido por la Ley, sin necesidad de realizar registros, depósitos o cualquier tipo de formalidad o divulgación. Así lo establece el artículo 1 del Texto Refundido de la Ley de Propiedad Intelectual (TRLPI).

Por su parte la Propiedad Industrial, al menos en España, comprende una serie de facultades legales cuyo propósito principal es proteger al creador de determinadas obras inmateriales, exigiendo generalmente un procedimiento

de registro para su efectividad. Estos derechos otorgan a su titular una potestad de exclusividad, comparable a un derecho de posesión, sobre las creaciones protegidas.

De manera general, esta clasificación legal abarca conceptos como las patentes, los modelos de utilidad, los diseños industriales, y los signos distintivos como las marcas y los nombres comerciales.

En el ordenamiento jurídico español, la Propiedad Industrial se subdivide en varias categorías específicas:

- Signos Distintivos (Marcas y Nombres Comerciales): Su función es diferenciar los productos o servicios ofrecidos por un agente económico de aquellos similares que se encuentran en el mercado, protegiendo combinaciones denominativas o gráficas.
- Diseños Industriales: Se encargan de salvaguardar la apariencia o estética externa de los artículos.
- Patentes y Modelos de Utilidad: Están destinados a proteger innovaciones, tanto en la forma de productos como de procedimientos, siempre que sean susceptibles de repetición y reproducción para fines industriales.
- Topografías de semiconductores: Protegen el esquema de disposición tridimensional y las interconexiones de las capas y elementos que componen un circuito integrado.

El poseedor de estos derechos puede decidir quién y de qué manera se hace uso de ellos. En España, la Oficina Española de Patentes y Marcas (OEPM) es el organismo competente que otorga esta protección a través de un procedimiento, la cual tiene vigencia en todo el territorio nacional.

La Propiedad Industrial funciona así, como una capa de protección legal que convierte ideas y diseños aplicables a la industria —que son inmateriales— en activos concretos, dándoles un estatus similar al de una posesión física, lo que permite a su dueño controlar su uso y difusión.

2. LA INTELIGENCIA ARTIFICIAL, CONCEPTOS Y ESPECIFICACIONES DE LA INTELIGENCIA ARTIFICIAL GENERATIVA

La inteligencia artificial (IA) es una tecnología que faculta a los sistemas informáticos y a las máquinas para reproducir capacidades humanas tales como el aprendizaje, la comprensión, la resolución de problemas, la toma de decisiones, la creatividad y la actuación autónoma.

Previo al desarrollo de la IA, los ordenadores se limitaban a ejecutar instrucciones explícitas definidas por los programadores y carecían de la capacidad de gestionar excepciones o afrontar situaciones imprevistas. En contraste, la IA aprende a partir de datos y optimiza su rendimiento mediante la identificación de patrones y la incorporación de correcciones, proceso que se denomina aprendizaje automático.

Si bien la base tecnológica de la IA es compleja, sus aplicaciones resultan fácilmente comprensibles. Por ejemplo, en los sistemas de navegación GPS, la IA es capaz de determinar en tiempo real la ruta más eficiente considerando el tráfico y la configuración de las vías. De forma análoga, asistentes virtuales como Siri pueden analizar patrones del habla y reconocer las particularidades lingüísticas de cada usuario.

Los dispositivos y aplicaciones dotados de IA tienen la capacidad de percibir y reconocer objetos, interpretar y responder al lenguaje humano, aprender de datos nuevos y de la experiencia, formular recomendaciones precisas tanto para usuarios como para especialistas y ejecutar acciones de manera autónoma sin intervención humana, como ocurre en el caso de los vehículos sin conductor.

No obstante, en 2025, la mayor parte de la investigación y del interés profesional y mediático en torno a la IA se concentra en los avances de la IA generativa, una tecnología capaz de producir textos, imágenes, vídeos y otros contenidos originales. Para comprender en profundidad la IA generativa, resulta esencial conocer primero las tecnologías que la sustentan: el *machine learning* (ML) y el *deep learning*, del que se tratara algo más en profundidad un poco mas adelante.

La IA permite que las máquinas adquieran conocimientos y mejoren su desempeño con el tiempo. Su funcionamiento se basa en la incorporación y el análisis de datos, la identificación de patrones y la adaptación progresiva de sus modelos internos. En teoría, estas capacidades deberían liberar a los trabajadores para que puedan concentrarse en tareas que requieren mayor creatividad o un nivel elevado de interpretación y matices.

El aprendizaje automático constituye la capacidad de la IA para detectar patrones y modificar sus parámetros internos en función de los datos nuevos que recibe. Se trata de un subcampo de la IA; por consiguiente, todo aprendizaje automático forma parte de la inteligencia artificial, aunque no toda inteligencia artificial se fundamenta en aprendizaje automático. La IA diseñada para ejecutar tareas de manera similar a la humana —conocida como inteligencia artificial débil (IAD)— ya está disponible. En cambio, la inteligencia artificial

general (IAG), que en teoría alcanzaría un nivel de comprensión equiparable al del ser humano, aún no existe... en principio.

Dentro del ámbito de la IA, el *machine learning* se centra en el desarrollo de modelos mediante el entrenamiento de algoritmos capaces de generar predicciones o tomar decisiones basadas en datos. Este campo abarca una variedad de técnicas que permiten a los sistemas informáticos aprender de la experiencia y realizar inferencias sin estar programados explícitamente para cada tarea.

Existen diversos tipos de técnicas y algoritmos de *machine learning*, entre ellos la regresión lineal, la regresión logística, los árboles de decisión, los bosques aleatorios, las máquinas de vectores de soporte (SVM), los algoritmos de k vecinos más cercanos (KNN), el *clustering*, entre otros. Cada uno de estos métodos es apropiado para distintos tipos de problemas y estructuras de datos.

Uno de los enfoques más destacados dentro del *machine learning* es el de las redes neuronales (o redes neuronales artificiales). Estas se inspiran en la estructura y el funcionamiento del cerebro humano. Una red neuronal está compuesta por capas interconectadas de nodos —análogos a las neuronas— que colaboran en el procesamiento y análisis de información compleja. Son especialmente eficaces para tareas que requieren detectar patrones y relaciones profundas en grandes volúmenes de datos.

La forma más elemental de *machine learning* es el aprendizaje supervisado, que utiliza conjuntos de datos etiquetados para entrenar algoritmos capaces de clasificar información o predecir resultados con precisión. En este enfoque, cada ejemplo de entrenamiento se encuentra asociado a una etiqueta o salida definida por humanos. El propósito consiste en que el modelo aprenda la relación entre las entradas y las salidas de los datos de entrenamiento, de modo que pueda predecir correctamente las etiquetas de datos nuevos y no vistos previamente.

La I emplea algoritmos para analizar conjuntos de datos extensos y complejos, con el fin de identificar patrones relevantes. En lugar de limitarse a ejecutar instrucciones rígidas, los modelos de IA aprenden a partir de los datos para efectuar predicciones, extraer conclusiones, llevar a cabo diversas tareas y mejorar progresivamente su rendimiento. No obstante, aunque la IA puede reconocer relaciones y generar información, no “piensa” como un ser humano; simplemente detecta correlaciones y estructuras estadísticas para producir resultados.

Para que un sistema de IA pueda generar predicciones fiables y tomar decisiones óptimas, es indispensable entrenarlo y proporcionarle conjuntos de datos pertinentes y de alta calidad. El aprendizaje automático reproduce el

proceso humano de reconocimiento de patrones, pero lo supera en eficiencia y precisión, especialmente en tareas como la detección de fraude. Gracias a su capacidad para analizar comportamientos con un alto nivel de detalle, puede identificar anomalías de manera inmediata considerando múltiples variables, incluso factores imprevisibles como las condiciones meteorológicas durante un pico de ventas. El aprendizaje profundo —un subcampo más avanzado del aprendizaje automático— perfecciona este proceso mediante el empleo de redes neuronales capaces de detectar patrones altamente complejos.

Los sistemas expertos constituyen programas de IA diseñados para replicar la toma de decisiones humana en dominios especializados. A través del uso de reglas predefinidas y bases de conocimiento estructuradas, ofrecen soluciones en ámbitos como el diagnóstico médico, el análisis financiero y la ciberseguridad. A diferencia del aprendizaje automático, los sistemas expertos se sustentan en lógica establecida y no en mecanismos de autoaprendizaje.

El procesamiento del lenguaje natural (PLN) permite que la IA comprenda textos y el modo en que las personas se expresan verbalmente. La interpretación del habla resulta especialmente compleja para las máquinas, dado que una misma palabra puede adquirir significados diferentes según la entonación, la función gramatical, el acento, el contexto, el estado emocional del hablante o la relación entre los interlocutores. Tras recibir suficiente entrenamiento o exposición al lenguaje, un modelo de IA puede gestionar estos matices con un alto grado de precisión.

a. Tipos de IA

Existen múltiples formas de clasificar la inteligencia artificial (IA), así como numerosos tipos de tecnologías asociadas. Entre las categorías más reconocidas se encuentran la IA generativa (como DALL·E y ChatGPT), la IA orientada al procesamiento del lenguaje natural, la IA aplicada a la robótica y la IA de visión artificial. A continuación, se presentan algunos tipos destacados de inteligencia artificial:

IA generativa: Este tipo de IA es capaz de producir contenido textual, audiovisual o sonoro a partir de los datos con los que ha sido entrenada. Entre los ejemplos más conocidos se encuentran ChatGPT y DALL·E. la IA generativa permite que las máquinas produzcan contenido original, como textos, imágenes, música o código. Mediante el aprendizaje de patrones presentes en grandes volúmenes de datos, estos modelos son capaces de generar resultados altamente realistas y creativos. Su aplicación es frecuente en asistentes de redacción, herramientas de diseño y tecnologías *deepfake*.

IA de memoria limitada: Se caracteriza por almacenar un conjunto reducido de datos directamente relacionados con la tarea que debe realizar. Esto abarca desde la interacción conversacional con el usuario hasta la capacidad de un vehículo autónomo para mantener una distancia segura con otros vehículos en la carretera.

Procesamiento del lenguaje natural (PLN): Su objetivo es comprender el lenguaje humano, tanto escrito como oral. Esta tecnología permite a los sistemas de IA identificar la intención del hablante y mantener diálogos fluidos y coherentes.

Visión artificial: No se limita al reconocimiento facial. Por ejemplo, diversas aplicaciones deportivas la utilizan para analizar el rendimiento físico y corregir posturas o movimientos incorrectos. La visión artificial faculta a la IA para reconocer objetos e interpretar información visual a partir de imágenes o secuencias de vídeo. Mediante el análisis de fotografías o transmisiones en tiempo real, la IA puede “observar” lo que ocurre e incluso reaccionar ante ello. Asimismo, tiene la capacidad de identificar expresiones faciales e inferir emociones, lo que conlleva implicaciones significativas en ámbitos como la seguridad, la atención sanitaria y las tecnologías de conducción autónoma.

IA aplicada a la robótica: Se emplea en sectores como el transporte, la manufactura y la cirugía, donde contribuye a mejorar la precisión, eficiencia y seguridad de los procesos. La robótica basada en IA incrementa la precisión y eficiencia en sectores como la cirugía, la manufactura, la agricultura e incluso en tareas domésticas. Los robots dotados de IA pueden adaptarse a obstáculos mediante algoritmos avanzados y optimizar su desempeño en tiempo real.

b. El entrenamiento y alimentación de la IA

La IA obtiene información a partir de los datos que recibe, los cuales varían según el modelo de aprendizaje empleado. Por ejemplo, los grandes modelos de lenguaje (LLM) utilizan extensas colecciones de libros, sitios web, publicaciones en redes sociales, foros y documentación técnica. Los modelos de visión artificial se entrenan con imágenes y vídeos, a menudo acompañados de etiquetas que facilitan el reconocimiento del movimiento, la distancia y otros elementos visuales. Por su parte, los modelos de aprendizaje por refuerzo aprenden mediante la interacción con su entorno a través de un proceso de ensayo y error, recibiendo recompensas o sanciones según el resultado de sus acciones.

Así la IA adquiere conocimientos mediante una amplia variedad de técnicas, cada una de ellas orientada a un propósito específico. Puede entrenarse para identificar patrones concretos, analizar datos con el fin de detectar relaciones significativas y aprender de los errores cometidos durante su proceso de funcionamiento. A continuación, se describen algunos de los métodos empleados para entrenar distintos modelos de IA:

Aprendizaje supervisado: Consiste en entrenar al sistema con ejemplos previamente etiquetados. Por ejemplo, si se proporcionan numerosas imágenes en las que se indica cuáles contienen tumores, la IA puede aprender a identificarlos posteriormente en imágenes nuevas.

Aprendizaje no supervisado: En este enfoque, la IA busca patrones de manera autónoma sin necesidad de etiquetas. En el ámbito de la detección de fraude, por ejemplo, el sistema analiza los hábitos de consumo de una persona —qué compra y cuándo— y genera alertas cuando detecta actividades inusuales.

Aprendizaje semisupervisado: Combina los dos métodos anteriores. La IA inicia su entrenamiento con una cantidad reducida de datos etiquetados que sirven como referencia, y posteriormente clasifica un volumen mayor de datos sin etiquetar en función de los conocimientos adquiridos.

Aprendizaje por refuerzo: En este caso, la IA aprende mediante un proceso de ensayo y error. Su desempeño se ajusta a partir de predicciones, decisiones y correcciones derivadas de los resultados obtenidos, en un proceso análogo al aprendizaje humano basado en la experiencia.

Aprendizaje profundo: Emplea redes neuronales artificiales para imitar el modo en que el cerebro humano procesa la información. A través del análisis de grandes conjuntos de datos, el aprendizaje profundo permite que la IA lleve a cabo tareas complejas, como el reconocimiento facial, la comprensión del habla y la conducción autónoma de vehículos.

Este último conocido popularmente como *Deep Learning* El *deep learning* es un subconjunto del *machine learning* que emplea redes neuronales multicapa —denominadas redes neuronales profundas—, las cuales reproducen de manera más aproximada la complejidad del proceso de toma de decisiones del cerebro humano.

Las redes neuronales profundas están compuestas por una capa de entrada, una capa de salida y, al menos, tres capas ocultas, aunque en la práctica suelen incluir cientos de ellas. Esto contrasta con las redes neuronales

tradicionales utilizadas en modelos clásicos de *machine learning*, que generalmente cuentan únicamente con una o dos capas ocultas.

La presencia de múltiples capas permite la realización de aprendizaje no supervisado, ya que estas redes pueden automatizar la extracción de características a partir de grandes volúmenes de datos no etiquetados y no estructurados, formulando predicciones propias sobre su significado.

Al no requerir intervención humana directa, el *deep learning* posibilita la aplicación del *machine learning* a gran escala. Esta técnica resulta especialmente adecuada para tareas como el procesamiento del lenguaje natural (PLN), la visión artificial y otras actividades que exigen la identificación rápida y precisa de patrones y relaciones complejas en grandes conjuntos de datos. De hecho, la mayoría de las aplicaciones de inteligencia artificial empleadas en la actualidad se basan, en alguna medida, en *deep learning*.

Asimismo, el *deep learning* posibilita diversas modalidades avanzadas de aprendizaje:

- Aprendizaje semisupervisado: Combina enfoques supervisados y no supervisados mediante la utilización simultánea de datos etiquetados y no etiquetados para entrenar modelos orientados a tareas de clasificación y regresión.
- Aprendizaje autosupervisado: Genera etiquetas implícitas a partir de datos no estructurados, evitando así la necesidad de disponer de grandes conjuntos de datos etiquetados de manera manual.
- Aprendizaje por refuerzo: Se basa en un proceso de ensayo y error en el que el modelo optimiza su comportamiento mediante funciones de recompensa, en lugar de depender exclusivamente de patrones previamente detectados en los datos.
- Aprendizaje por transferencia: Permite aplicar los conocimientos adquiridos en una tarea o conjunto de datos para mejorar el rendimiento del modelo en otra tarea distinta o con datos de diferente naturaleza.

Por otro lado, se debe considerar la recolección automatizada de datos — conocida como *data scraping*— constituye uno de los puntos más conflictivos en la relación entre inteligencia artificial generativa y Propiedad Intelectual. Los modelos avanzados de IA requieren cantidades ingentes de información para su entrenamiento, y esta dependencia estructural ha convertido el *scraping* en un mecanismo esencial para su funcionamiento. No obstante, la obtención masiva

y automatizada de datos procedentes de sitios web, redes sociales o bases de datos plantea una tensión jurídica profunda, ya que gran parte de esos contenidos están amparados por diversos regímenes de protección. La tecnología necesita acceder a datos para operar, mientras que el Derecho debe garantizar la preservación de derechos exclusivos y la integridad de bienes jurídicos protegidos. Esta colisión entre demanda tecnológica y límites legales constituye uno de los desafíos más complejos y menos resueltos del sistema actual de Propiedad Intelectual.

El *scraping* consiste en extraer información disponible en entornos digitales mediante herramientas automatizadas que permiten capturar y organizar datos para usos posteriores. Aunque suele confundirse con la minería de datos, su naturaleza jurídica es distinta: el *scraping* implica la obtención directa de los datos, mientras que la minería se orienta al análisis de patrones. Es precisamente la fase extractiva la que genera mayores riesgos legales, pues puede suponer el acceso y copia de contenidos sujetos a derechos de autor, bases de datos con protección sui generis, signos distintivos, información comercial sensible o elementos que afectan a derechos de la personalidad. La práctica es técnicamente eficiente, veloz y escalable, pero también puede resultar intrusiva, desproporcionada y difícil de supervisar.

El impacto jurídico del *scraping* puede alcanzar numerosos ámbitos: la reproducción no autorizada de textos, imágenes, código o grabaciones puede vulnerar derechos de autor; la extracción masiva de contenidos puede perjudicar la inversión que sustenta la protección sui generis de bases de datos en la Unión Europea; la captación de marcas o signos distintivos puede generar usos indebidos o confusión en entornos digitales; la obtención de información empresarial no divulgada puede comprometer secretos empresariales, con consecuencias económicas graves; y la captura de voces, rostros o estilos personales puede afectar derechos de imagen o de explotación de la personalidad. La amplitud de posibles afectaciones evidencia que el *scraping* no es un acto neutro, sino una operación jurídicamente sensible cuyo impacto se multiplica cuando sus resultados se utilizan para entrenar sistemas capaces de replicar, transformar o reinterpretar contenidos protegidos.

Frente a este escenario, se han propuesto diversas soluciones para mitigar la inseguridad jurídica. El mecanismo de exclusión mediante archivos de control como robots.txt constituye una herramienta tecnológica de autorregulación, aunque su eficacia es limitada por su carácter no vinculante, su falta de granularidad y el hecho de que muchos agentes automatizados lo ignoran. En el ámbito normativo, el Reglamento de Inteligencia Artificial de la Unión Europea introduce obligaciones reforzadas de transparencia, trazabilidad y documentación relativas al origen de los datos utilizados para entrenar modelos de

IA de uso general, así como la necesidad de habilitar mecanismos efectivos para que los titulares de derechos puedan excluir sus obras del entrenamiento. Paralelamente, las tecnologías de trazabilidad —como sistemas de *provenance*, marcas de agua digitales o huellas digitales— buscan incrementar la capacidad de identificar el origen de los contenidos y detectar usos no autorizados, reforzando así la protección de los titulares.

En un entorno digital sin fronteras, la regulación exclusivamente nacional resulta insuficiente. Por ello, organismos internacionales están promoviendo la adopción de códigos de conducta, estándares técnicos armonizados, cláusulas contractuales tipo y principios comunes de transparencia orientados a garantizar que el desarrollo de la inteligencia artificial no erosione los fundamentos de la Propiedad Intelectual. El objetivo no es obstaculizar la innovación, sino asegurar que el avance tecnológico se produzca dentro de un marco que preserve los derechos de quienes generan, financian y mantienen los contenidos sobre los que la IA se entrena.

Por otro lado, debemos destacar la inteligencia artificial generativa que se refiere a modelos de *deep learning* capaces de producir contenido original y complejo —como texto de formato extenso, imágenes de alta resolución, vídeos o audio de apariencia realista, entre otros— en respuesta a instrucciones o solicitudes formuladas por un usuario.

En términos generales, los modelos generativos crean una representación simplificada de los datos con los que fueron entrenados y utilizan dicha representación para generar resultados que guardan similitud con los datos originales, aunque sin replicarlos de manera exacta.

Tradicionalmente, los modelos generativos se empleaban en estadística para el análisis de datos numéricos. No obstante, durante la última década han experimentado una evolución significativa que les ha permitido analizar y producir tipos de datos mucho más complejos. Este avance ha coincidido con la aparición de tres modelos sofisticados de *deep learning*:

Autocodificadores variacionales (VAE): Presentados en 2013, estos modelos posibilitan la generación de múltiples variaciones de contenido a partir de una misma instrucción o mensaje.

Modelos de difusión: Introducidos alrededor de 2014, funcionan añadiendo progresivamente “ruido” a una imagen hasta volverla irreconocible y, posteriormente, eliminándolo de manera controlada para producir imágenes originales en función de las instrucciones del usuario.

Modelos transformadores (*transformers*): Entrenados con datos secuenciales, los transformadores son capaces de generar secuencias extendidas de contenido, como palabras en un texto, formas en una imagen, fotogramas en un vídeo o comandos en un bloque de código. Constituyen la base de la mayoría de las herramientas actuales de IA generativa, entre ellas ChatGPT, GPT-4, Copilot, BERT, Bard y Midjourney.

La inteligencia artificial generativa ha irrumpido con fuerza en el ámbito creativo, transformando la manera en que se produce arte, música, texto y otras formas de expresión. No obstante, este avance tecnológico ha suscitado un debate profundo acerca de los derechos de autor y la protección de la propiedad intelectual.

En el núcleo de esta discusión se encuentra la cuestión de si la IA generativa produce obras verdaderamente originales o si, por el contrario, se limita a recombinar contenidos preexistentes como se analizará más adelante.

Los modelos de IA generativa se basan en técnicas de aprendizaje profundo que permiten a los sistemas identificar patrones y relaciones dentro de grandes conjuntos de datos. El entrenamiento consiste en proporcionar al modelo una cantidad masiva de información —como imágenes, textos o composiciones musicales— a partir de la cual aprende a generar contenido nuevo que, aunque comparte características con los datos de entrenamiento, no constituye una reproducción exacta de estos.

Este proceso plantea interrogantes fundamentales respecto a la originalidad y la autoría. ¿Debe considerarse al modelo de IA como creador de la obra generada? ¿O es dicha obra simplemente el resultado de la recombinación de los datos utilizados para su entrenamiento? Además, surge la cuestión del rol que desempeñan los artistas y creadores cuyos trabajos sirven como base para estos modelos. Estas problemáticas aún carecen de respuestas definitivas y continúan siendo objeto de un activo debate jurídico, ético y social.

c. Redes Generativas Antagónicas

El desarrollo reciente de la inteligencia artificial —y, en particular, de las denominadas redes generativas antagónicas (Generative Adversarial Networks, GAN)— ha permitido que los sistemas automatizados alcancen niveles de precisión extremadamente elevados en la identificación de objetos, patrones visuales y otras entidades presentes en grandes volúmenes de datos. Mediante la exposición a millones de imágenes, estas arquitecturas algorítmicas pueden determinar con notable exactitud en qué fotografías aparece un determinado

elemento, como podría ser una persona montando en bicicleta en vía pública. No obstante, la generación de contenidos completamente novedosos plantea desafíos de mayor complejidad, ya que exige capacidades asociadas tradicionalmente a la imaginación y a la creatividad, atributos no presentes en los sistemas de IA bajo los modelos convencionales.

El enfoque técnico de las GAN se fundamenta en la interacción competitiva de dos redes neuronales artificiales —modelos matemáticos inspirados en el funcionamiento biológico del cerebro humano— que operan de manera confrontativa. Ambas redes se entrenan sobre un mismo conjunto de datos. La primera, denominada red generativa, tiene por función producir variaciones de los ejemplos que ha procesado previamente, es decir, generar nuevas imágenes basadas en patrones aprendidos. La segunda, la red discriminadora, debe determinar si cada imagen que evalúa pertenece realmente al conjunto original de entrenamiento o si, por el contrario, se trata de una creación artificial generada por su contraparte. En términos operativos, la red discriminadora responde continuamente a la cuestión de si la imagen presentada puede considerarse verosímil o real.

A medida que avanza el entrenamiento, la red generativa perfecciona su capacidad de producir imágenes de tal calidad que la red discriminadora deja de ser capaz de detectar la falsedad. Este proceso culmina en la obtención de modelos capaces de generar imágenes, sonidos u otros contenidos con un realismo tal que incluso expertos humanos encuentran dificultades para distinguirlos del material auténtico. Las aplicaciones ya demostradas incluyen la generación de retratos hiperrealistas de personas inexistentes, así como la creación de obras artísticas que imitan estilos pictóricos consagrados. Asimismo, estas tecnologías permiten reinterpretar imágenes, modificando condiciones ambientales o traduciendo elementos visuales de un dominio a otro.

Si bien los resultados no están exentos de imperfecciones —como la producción de imágenes anatómicamente incoherentes—, su elevado grado de realismo ha llevado a parte de la comunidad científica a concluir que estas redes logran modelar estructuras subyacentes presentes en los datos. Sin embargo, otra corriente de análisis, a la cual se adhiere este estudio, sostiene que tanto los algoritmos como las redes neuronales carecen de autoconciencia, intencionalidad o comprensión semántica; son exclusivamente sistemas de procesamiento estadístico sin inteligencia en sentido jurídico-filosófico, y su funcionamiento carece de conocimiento de causa, voluntad o propósito, siendo estas cualidades exclusivas de los operadores humanos que los diseñan y supervisan.

El potencial de las GAN es considerable, pues su estructura matemática les permite aproximarse a cualquier distribución de datos y, en consecuencia,

generar contenidos en múltiples ámbitos —imágenes, música, lenguaje hablado o texto escrito— con altos niveles de verosimilitud. De esta evolución han surgido variantes más sofisticadas, como las denominadas *recycle*-GAN, capaces de transferir el contenido de un dominio a otro manteniendo el estilo propio del dominio original. Todo ello se realiza mediante procesos de aprendizaje profundo no supervisado, lo que intensifica la necesidad de establecer marcos jurídicos robustos que regulen su utilización, dada la creciente capacidad de estas tecnologías para producir contenidos cuyo origen artificial puede resultar difícil de identificar.

2. INTELIGENCIA ARTIFICIAL GENERATIVA Y EL RÉGIMEN JURÍDICO DE LA PROPIEDAD INTELECTUAL (DERECHO DE AUTOR)

El derecho de autor se fundamenta históricamente en una construcción dogmática de carácter estrictamente antropocéntrico. La normativa internacional y nacional en la materia —entre ellas, el Convenio de Berna, el Tratado de la OMPI sobre Derecho de Autor y la Ley de Propiedad Intelectual española (LPI)— asume sin excepción que únicamente las personas naturales pueden ostentar la condición de autores, en tanto sujetos dotados de capacidad jurídica, capacidad de obrar y aptitud creativa protegible. Las máquinas, al carecer de voluntad, discernimiento, conciencia y subjetividad jurídica, quedan excluidas de dicho estatuto.

La irrupción de la inteligencia artificial generativa ha tensionado este paradigma. Estas tecnologías son capaces de producir contenidos de forma autónoma, en volúmenes potencialmente masivos y con grados de sofisticación que pueden resultar indistinguibles de las creaciones humanas. Ello obliga a reexaminar conceptos nucleares del sistema autoral, como autoría, originalidad y titularidad.

d. El principio antropocéntrico de la autoría: núcleo dogmático del sistema

La LPI establece, en consonancia con el Convenio de Berna, que “autor” es la persona natural que crea una obra literaria, artística o científica. Esta definición, aparentemente sencilla, constituye una barrera infranqueable para la atribución de autoría a sistemas de IA.

Si la autoría exige creación humana, entonces:

- la IA no puede ser autora,
- no puede ser titular de derechos morales ni patrimoniales,
- no puede transmitir derechos,
- ni puede reclamar protección bajo el régimen de propiedad intelectual.

En consecuencia, los contenidos generados íntegramente por IA —esto es, sin aportación creativa humana sustancial— carecen de autor y, por tanto, nacen directamente en el dominio público, sin posibilidad de protección por derecho de autor. El sistema jurídico actual mantiene así su coherencia interna: allí donde no existe autor, tampoco puede existir derecho.

- e. La naturaleza jurídica de los *prompts* expresión, idea o instrucción técnica

Desde una perspectiva jurídico-autoral, el *prompt*⁵⁶⁶, por complejo que pueda resultar, no constituye una obra expresiva protegible, sino una idea, directriz o instrucción destinada a guiar el funcionamiento del sistema de IA.

El artículo 2 del Tratado de la OMPI sobre Derecho de Autor excluye de manera categórica las ideas, procedimientos y métodos de operación de la protección. Reconocer protección autoral a un *prompt* implicaría, de facto, blindar ideas durante periodos excesivos, contraviniendo el principio de libre circulación de ideas que preside el sistema autoral.

No obstante, cuando los *prompts* se integran en una estructura técnica compleja —por ejemplo, en procesos de *prompt engineering* con valor competitivo y carácter confidencial— pueden ser objeto de protección como secreto empresarial, conforme a la Ley 1/2019, de 20 de febrero, de Secretos Empresariales. Esta protección, de raíz mercantil, no los convierte en obras creativas, pero sí reconoce su valor económico.

La cuestión relativa a la eventual protección jurídica de los prompts utilizados en sistemas de inteligencia artificial generativa ha adquirido una notable relevancia en el debate contemporáneo sobre Propiedad Intelectual. Numerosos usuarios, tras invertir tiempo y esfuerzo en la elaboración de instrucciones

⁵⁶⁶ Los prompts en IA son las instrucciones, preguntas o textos iniciales que un usuario le da a un sistema de inteligencia artificial (como ChatGPT o generadores de imágenes) para guiarlo y obtener una respuesta o resultado específico, definiendo el rol, tono, formato y objetivo deseado. Son fundamentales para la comunicación con la IA, y el arte de crearlos se llama "prompt engineering", ya que un buen prompt genera respuestas más precisas y útiles.

complejas destinadas a optimizar los resultados de dichos sistemas, se preguntan si tales formulaciones podrían ser objeto de protección exclusiva. Desde una perspectiva estrictamente jurídica, la respuesta es negativa: los prompts no son susceptibles de protección mediante derechos de autor. Esta conclusión no es contingente ni derivada de un vacío normativo, sino que descansa en principios estructurales del Derecho de Propiedad Intelectual, cuya función es precisamente delimitar el ámbito de la protección y evitar una extensión desproporcionada del monopolio jurídico.

El derecho de autor protege únicamente la expresión concreta de una idea, no la idea en abstracto ni las instrucciones, métodos u operaciones que permiten ejecutarla. Este principio, consagrado en instrumentos internacionales como el Tratado de la OMPI sobre Derecho de Autor, impide calificar un *prompt* como obra, incluso en aquellos supuestos en que su formulación pueda parecer compleja, innovadora o técnicamente sofisticada. El *prompt* constituye, en esencia, una instrucción operativa destinada a orientar el funcionamiento de un sistema algorítmico, lo cual lo aproxima más a un método de operación que a una creación expresiva. Reconocerle protección supondría, en la práctica, conceder derechos exclusivos sobre una forma de interactuar con una herramienta tecnológica, generando un efecto restrictivo incompatible con la lógica de apertura y libre aprovechamiento intelectual que subyace al sistema de derecho de autor.

Una eventual protección de los *prompts* resultaría además desproporcionada y contraria al equilibrio interno del sistema. La atribución de derechos de exclusividad —por un plazo que podría extenderse durante la vida del autor y setenta años después de su fallecimiento— sobre formulaciones que, en su naturaleza, son meras instrucciones funcionales, transformaría el entorno tecnológico de la IA generativa en un espacio jurídicamente inoperable. El ecosistema de estos sistemas se basa en el uso masivo, reiterado y evolutivo de indicaciones aportadas por los usuarios. Convertir cada prompt en una obra protegida multiplicaría los riesgos de infracción, bloquearía la innovación, impediría la interacción libre con los sistemas y erosionaría la función social del derecho de autor.

No obstante, la inexistencia de protección por vía de derechos de autor no implica la inexistencia de mecanismos jurídicos para resguardar ciertos *prompts*, especialmente aquellos que puedan poseer un valor económico relevante. En tales casos, el cauce más adecuado no es la Propiedad Intelectual tradicional, sino el régimen de secretos empresariales. Esta vía permite proteger *prompts* particularmente valiosos siempre que se cumplan tres condiciones básicas: que la información tenga carácter secreto, que posea un valor comercial derivado precisamente de dicha confidencialidad y que su titular haya adoptado medidas razonables para preservar dicho secreto. A diferencia del derecho de autor, esta

protección no se otorga de forma automática ni pretende amparar cualquier formulación, sino únicamente aquellas que forman parte de una estrategia empresarial legítima y que se gestionan activamente como información reservada.

En consecuencia, el análisis doctrinal muestra que los prompts no encajan en la estructura conceptual del derecho de autor y que su protección, cuando resulte necesaria, debe articularse a través de instrumentos jurídicos que respeten la naturaleza operativa de estas instrucciones. El equilibrio entre innovación tecnológica y coherencia sistemática —núcleo del Derecho de Propiedad Intelectual— exige mantener esta distinción para evitar expansiones indebidas del monopolio jurídico y preservar un marco normativo funcional para el desarrollo de la inteligencia artificial generativa.

f. Obras generadas por IA: originalidad, reproducción y riesgos jurídicos

La inteligencia artificial generativa ha permitido la creación de textos, imágenes, música, ilustraciones, diseños y contenidos complejos, como cómics o narraciones, a partir de modelos entrenados con grandes volúmenes de datos. Este fenómeno plantea un reto jurídico mayúsculo: determinar si estos resultados pueden considerarse “obras” en el sentido estricto del derecho de autor.

El criterio doctrinal y jurisprudencial dominante sostiene que una obra requiere: primeramente, intervención creativa humana a la par que originalidad suficiente, entendida como expresión de la acción inventiva o creativa del autor.

En virtud de lo anterior cuando la IA actúa de manera autónoma, sin aportación humana relevante, la obra carece de originalidad jurídica en sentido estricto ya que solo cuando una persona introduce decisiones creativas —modificando, seleccionando, organizando o interviniendo con juicio propio sobre el contenido generado— puede reconocerse un nivel de originalidad imputable al ser humano.

El empleo de obras protegidas por derechos de autor en el entrenamiento de modelos de inteligencia artificial generativa constituye uno de los focos de preocupación jurídica más relevantes en el actual panorama regulatorio. La utilización masiva de contenidos sujetos a protección para alimentar sistemas de aprendizaje automático comporta riesgos significativos, entre los que destacan la reproducción no autorizada de materiales originales, la generación de obras derivadas sin consentimiento de los titulares, la potencial vulneración de derechos morales vinculados a la integridad o paternidad de la obra, y los conflictos con el régimen jurídico aplicable al derecho de extracción y reutilización en bases de datos. Estas cuestiones, aún sin una solución plenamente armonizada,

se encuentran en el centro de las discusiones legislativas tanto a nivel europeo como internacional, dada la necesidad de equilibrar el impulso tecnológico con la preservación de los derechos exclusivos reconocidos a los creadores y titulares.

A lo anterior se suma el impacto económico derivado de la capacidad de la IA generativa para producir contenidos de alta calidad a un coste marginal prácticamente nulo. Esta circunstancia plantea riesgos estructurales para el mercado creativo, en la medida en que podría generar fenómenos de saturación, desplazar económicamente a autores humanos, erosionar el valor diferencial de las obras originales y dificultar la adecuada remuneración de quienes aportan creatividad humana. Diversos organismos internacionales, así como autoridades de competencia y entidades reguladoras, han advertido sobre la posibilidad de que la proliferación masiva de contenidos generados por IA distorsione los mercados culturales, modifique las cadenas de valor tradicionales y comprometa la sostenibilidad económica del sector creativo.

En conjunto, estos elementos ponen de manifiesto que la interacción entre IA generativa y propiedad intelectual necesita respuestas normativas sólidas, técnicamente precisas y adaptadas a la complejidad del ecosistema digital contemporáneo. El reto consiste en permitir el desarrollo de tecnologías disruptivas sin menoscabar las garantías esenciales que sustentan la creación humana, el incentivo a la innovación y la seguridad jurídica en los mercados culturales y tecnológicos.

g. Autoría y titularidad de derechos en obras generadas o asistidas por IA

Uno de los debates más complejos en el ámbito de la propiedad intelectual relacionada con la inteligencia artificial se centra en la atribución de autoría y la determinación de la titularidad en escenarios híbridos donde concurren aportaciones humanas y resultados generados por sistemas automatizados. La normativa comparada revela una tendencia uniforme: únicamente las personas naturales pueden ostentar la condición de autor, dado que la creatividad, la intencionalidad y la capacidad jurídica son elementos indispensables que no pueden atribuirse a una máquina.

En consecuencia, cuando una obra es generada íntegramente por un sistema de inteligencia artificial sin intervención humana sustancial, carece de autor jurídico y no puede ser objeto de protección por derechos de propiedad intelectual. Dichos contenidos ingresan directamente en el dominio público, siendo libremente utilizables, reproducibles y transformables por cualquier tercero. Por el contrario, cuando la persona usuaria realiza aportaciones creativas significativas, tales como la selección de contenidos, la modificación sustancial

del material generado, la organización de elementos o la aplicación de criterios estéticos o juicios creativos, puede reconocerse la autoría humana y, por tanto, la obra resultante puede acogerse al régimen ordinario de protección. Los tribunales y las oficinas de registro han comenzado a admitir este enfoque, siempre que la intervención supere el umbral mínimo de originalidad, lo que plantea desafíos importantes en términos de apreciación, valoración y carga de la prueba.

La titularidad patrimonial en las obras híbridas corresponde a la persona que realiza la contribución creativa. Sin embargo, los contratos y términos de uso de las plataformas de inteligencia artificial introducen complejidades adicionales, dado que un acuerdo puede estipular que el usuario “adquiere derechos” sobre los resultados generados, pero no puede conferir protección a aquello que la ley no reconoce como obra protegida. La titularidad contractual de explotación no sustituye ni altera el requisito legal de autoría humana, lo que exige una clara distinción entre derechos contractuales y derechos autorales propiamente dichos.

Hasta la fecha, la exclusión de las obras generadas automáticamente por inteligencia artificial del ámbito de protección no ha generado disfunciones relevantes; el sistema jurídico ha demostrado coherencia y funcionalidad. No obstante, en el futuro podría considerarse oportuno otorgar algún grado de protección económica a este tipo de resultados mediante la creación de un régimen sui generis, adaptado a la naturaleza no humana de estas producciones. Dicho régimen podría caracterizarse por plazos breves, derechos limitados, finalidad incentivadora en lugar de moral y protección basada en la inversión realizada más que en la creatividad. Esta aproximación permitiría equilibrar la promoción de la innovación con la seguridad jurídica, sin alterar los principios esenciales de la dogmática del derecho de autor.

h. La Respuesta de las legislaciones de Propiedad Intelectual al Desarrollo de la IA

La irrupción de la inteligencia artificial generativa ha tensionado de manera significativa la capacidad de adaptación del Derecho de la Propiedad Intelectual, un sistema construido sobre bases históricas estables y diseñado para un mundo en el que la creación era una manifestación exclusivamente humana. El marco internacional de la PI no anticipó un escenario en el que sistemas automatizados pudieran producir, transformar o reinterpretar obras literarias, científicas o artísticas; sin embargo, sus principios estructurales continúan funcionando como referencia normativa indispensable para orientar la respuesta jurídica frente a estos nuevos fenómenos.

En el ámbito internacional, instrumentos como el Convenio de Berna, los Tratados de Internet de la OMPI y el Tratado de la OMPI sobre Derecho de Autor proporcionan el armazón conceptual sobre el que descansa la protección de las creaciones intelectuales. Aunque ninguno de estos acuerdos previó explícitamente la existencia de sistemas de IA generativa, sí consagran reglas cuyo alcance resulta especialmente relevante hoy, como la prohibición de someter la protección autoral a formalidad alguna. Conforme a este principio, toda obra goza de protección desde el momento mismo de su creación, con independencia de que se registre o se notifique su existencia.

Igualmente, el denominado test de los tres pasos, incorporado en diversos tratados, opera como criterio rector para valorar la legitimidad de cualquier excepción a los derechos exclusivos. Este parámetro exige que la limitación se aplique únicamente en supuestos específicos, que no interfiera con la explotación normal de la obra y que no cause un perjuicio injustificado al titular. Su carácter flexible, concebido para equilibrar intereses contrapuestos, lo ha convertido en uno de los ámbitos interpretativos más controvertidos en el contexto de la reutilización masiva de contenidos para entrenar modelos de IA.

A ello se suman las obligaciones internacionales de protección de la información sobre gestión de derechos y de las medidas tecnológicas destinadas a impedir usos no autorizados. En un entorno donde los modelos de IA manipulan, transforman y recombinan datos, la eliminación de metadatos o de mecanismos de protección puede suponer la pérdida de trazabilidad de las obras, lo que dificulta la identificación de su régimen de protección y la determinación de eventuales infracciones.

En el plano sustantivo, la exigencia de autoría humana sigue siendo un requisito ineludible para que exista obra protegida. La normativa española lo expresa de forma categórica al definir al autor como la persona natural que crea la obra literaria, artística o científica, y esta misma concepción se mantiene de forma prácticamente uniforme en el derecho comparado. La inteligencia artificial, al carecer de subjetividad jurídica, voluntad, autonomía decisoria o capacidad para asumir derechos y obligaciones, no puede, bajo la arquitectura actual del sistema, ser titular de derechos de autor. La ausencia de autor humano implica necesariamente la ausencia de obra protegida en los términos que exige el Derecho de la PI.

Algunos ordenamientos contemplan supuestos específicos para obras generadas mediante procesos informáticos, atribuyendo la autoría a quien realiza los actos necesarios para su creación, como ocurre en la legislación británica. No obstante, esta previsión fue diseñada para tecnologías muy anteriores a la

IA generativa contemporánea, y su aplicación sin ajustes normativos genera importantes dudas interpretativas en la doctrina actual.

De lo visto se desprende que, en los últimos años, la expansión acelerada de la inteligencia artificial ha obligado a los ordenamientos jurídicos, y particularmente al de la Unión Europea, a replantear de manera profunda sus marcos normativos. Buena parte de las disposiciones vigentes en materia de propiedad intelectual e industrial fueron concebidas para entornos tecnológicos en los que no existían sistemas capaces de generar obras, innovaciones o resultados técnicos de forma autónoma. La irrupción de la IA generativa y de modelos avanzados de aprendizaje automático ha transformado esta realidad, convirtiendo a la inteligencia artificial en un agente estructural del ámbito creativo, técnico e industrial. En este contexto, el Derecho comienza a adaptarse a una situación en la que la IA deja de ser una anomalía regulatoria para convertirse en un actor permanente cuya actividad debe someterse a parámetros de responsabilidad, transparencia y seguridad.

En el plano europeo, el hito normativo más relevante ha sido la aprobación del Reglamento (UE) 2024/1689, conocido como *AI Act*, considerado la primera legislación integral sobre inteligencia artificial a nivel mundial. Este instrumento jurídico tiene por finalidad garantizar que el desarrollo, despliegue y utilización de sistemas de IA se realicen de forma segura, respetuosa con los derechos fundamentales y compatible con las obligaciones derivadas de la propiedad intelectual y la propiedad industrial. Aunque el reglamento no introduce nuevos derechos de exclusiva, sí impone a los proveedores de sistemas de IA la obligación de cumplir estrictamente con las normas vigentes en materia de propiedad intelectual al entrenar modelos o generar contenidos, incluyendo el respeto a las obras protegidas, las bases de datos reguladas y los signos distintivos registrados.

El *AI Act* incorpora, además, un sistema de gobernanza institucional que incluye la creación de la Oficina Europea de Inteligencia Artificial (*European AI Office*) y de un órgano de coordinación (*AI Board*), así como paneles científicos y foros técnicos destinados a supervisar la aplicación del reglamento, evaluar riesgos y asegurar que la IA se utilice de acuerdo con los marcos legales de la Unión. Este entramado normativo tiene consecuencias directas para la protección de la propiedad intelectual y la propiedad industrial: exige trazabilidad y documentación sobre los datos empleados para entrenar modelos, facilita la atribución de responsabilidad, impone restricciones cuando la IA se utiliza de manera potencialmente lesiva para los derechos fundamentales —incluyendo usos indebidos de contenidos protegidos— y mantiene la exigencia de que las invenciones vinculadas con IA cumplan los requisitos tradicionales de patentabilidad, originalidad y distintividad.

En cuanto a España, la situación jurídica se encuentra en un proceso de transición hacia un marco normativo plenamente adaptado a las exigencias europeas. Hasta fechas recientes, el ordenamiento español carecía de una ley propia dedicada a la inteligencia artificial, lo que implicaba que la regulación aplicable derivaba principalmente de la legislación nacional sobre propiedad intelectual, derechos de autor, protección de datos y responsabilidad civil, complementada por las disposiciones europeas. Con la entrada en vigor del *AI Act*, sus obligaciones se aplican directamente en España, obligando a desarrolladores, implantadores y operadores de sistemas de IA a respetar los derechos de propiedad intelectual existentes y a adoptar medidas de transparencia, gestión de riesgos y control del uso de datos.

En 2025 se presentó un Anteproyecto de Ley para el Buen Uso y la Gobernanza de la Inteligencia Artificial, cuyo objetivo es trasladar al marco estatal determinadas obligaciones del *AI Act* y crear una autoridad nacional especializada —la Agencia Española de Supervisión de la IA— encargada de supervisar el cumplimiento de los requisitos de transparencia, fiabilidad y ética en el desarrollo y utilización de estos sistemas. No obstante, persisten zonas de incertidumbre, especialmente en relación con la naturaleza jurídica de las obras generadas por IA. El marco normativo español, en consonancia con el derecho comparado, exige creatividad humana para reconocer protección mediante derechos de autor. En consecuencia, los contenidos generados de forma autónoma por sistemas de IA que no incorporen una intervención humana creativa significativa difícilmente pueden considerarse obras protegibles. Esta circunstancia genera un espacio jurídico ambiguo tanto para las entidades que desarrollan modelos de IA como para los usuarios que emplean estos sistemas, pues no existe plena claridad sobre la atribución de derechos respecto de los resultados generados ni sobre el tratamiento jurídico de las obras utilizadas para entrenar dichos modelos.

En conjunto, tanto en el plano europeo como en el español, el Derecho avanza hacia un modelo que busca compatibilizar la innovación tecnológica con la preservación de los principios esenciales de la propiedad intelectual y la propiedad industrial. El reto actual consiste en adaptar los marcos jurídicos sin desnaturalizarlos, garantizando que la inteligencia artificial se integre en el ecosistema creativo e industrial sin erosionar los derechos que históricamente han protegido a autores, inventores y titulares de activos intangibles.

3. ALGUNOS ASPECTOS DE PROPIEDAD INDUSTRIAL E INTELIGENCIA ARTIFICIAL

La interacción entre la inteligencia artificial y la Propiedad Industrial está redefiniendo el alcance mismo de la actividad inventiva, obligando a reinterpretar categorías jurídicas tradicionalmente consolidadas. La IA se proyecta simultáneamente como objeto de protección —en tanto tecnología que puede integrar una invención patentable— y como herramienta que interviene de manera decisiva en la generación de soluciones técnicas, llegando incluso a sustituir parcialmente la creatividad humana en determinados procesos de diseño y optimización.

El impacto más visible se aprecia en el ámbito de las patentes asociadas a sistemas de IA generativa, cuya evolución ha sido excepcionalmente acelerada. Desde 2017, las solicitudes relacionadas con esta tecnología han experimentado incrementos superiores al 800 %, fenómeno prácticamente inédito en la historia reciente de la Propiedad Industrial. Organismos internacionales han puesto de manifiesto tendencias claras: el liderazgo absoluto del sector empresarial en la presentación de solicitudes, la posición destacada de jurisdicciones como China —que duplica las publicaciones de cualquier otro país— y la proliferación de decenas de miles de familias de patentes vinculadas a la IA generativa en menos de una década. Todo ello confirma que se trata de una tecnología estructural con capacidad para transformar sectores como la biomedicina, la ingeniería avanzada, el diseño industrial o la robótica.

Este crecimiento abrupto tensiona los fundamentos conceptuales del derecho de patentes y reabre debates jurídicos de gran calado. El más relevante es la cuestión de si una inteligencia artificial puede ser jurídicamente considerada inventora. A pesar de que sistemas avanzados de IA participan activamente en la identificación de soluciones técnicas, el marco normativo vigente en prácticamente todas las jurisdicciones exige que el inventor sea una persona natural. La negativa se fundamenta en criterios jurídicos esenciales: la IA carece de personalidad jurídica, no puede asumir derechos u obligaciones, no posee voluntad ni intención creadora y no es susceptible de recibir el incentivo económico que justifica el monopolio temporal de la patente. En consecuencia, incluso cuando la contribución técnica de la IA resulte determinante, la titularidad inventiva recae necesariamente en los operadores humanos que conciben, dirigen o controlan su funcionamiento.

Junto a esta cuestión, las oficinas de patentes afrontan dificultades adicionales respecto de la patentabilidad de invenciones que incorporan IA como elemento esencial. Surgen desafíos relativos a la delimitación entre materia técnica patentable y algoritmos abstractos excluidos de protección, así como problemas

de suficiencia descriptiva. La divulgación completa y reproducible de la invención —exigencia básica del sistema de patentes— se ve comprometida cuando los modelos se componen de millones de parámetros no explicables y dependen de conjuntos de datos cuyo origen, tratamiento o accesibilidad no siempre resulta verificable. Todo ello obliga a replantear criterios de reproducibilidad, transparencia y alcance de la descripción en contextos donde la lógica interna del sistema es inherentemente opaca.

La evolución reciente del panorama de patentes vinculadas con inteligencia artificial generativa revela un crecimiento de magnitud excepcional. Los datos publicados por la Organización Mundial de la Propiedad Intelectual muestran que, desde 2017, las solicitudes relativas a tecnologías de IA generativa han experimentado un incremento superior al 800 %, cifras que no solo reflejan el ritmo de innovación del sector, sino también la creciente competencia internacional por adquirir posiciones de liderazgo en este ámbito tecnológico. Entre 2014 y 2023 se registraron más de cincuenta y cuatro mil familias de patentes relacionadas con la materia, destacando que la mayor parte de esta actividad procede de grandes corporaciones tecnológicas, que concentran la mayoría de las solicitudes mundiales. En este contexto, determinados Estados han acumulado una clara ventaja competitiva, situándose China a la cabeza con un volumen de publicaciones que duplica al del siguiente país del ranking. Este escenario evidencia la consolidación de una revolución tecnológica en pleno desarrollo y muestra cómo el sistema de Propiedad Industrial se enfrenta a presiones sin precedentes para absorber y regular adecuadamente un flujo creciente de innovaciones.

El marco jurídico internacional en materia de patentes mantiene un principio esencial: la condición de inventora corresponde exclusivamente a una persona natural. Esta exigencia responde a la naturaleza misma del sistema de patentes, construido sobre la idea de incentivar la creatividad humana mediante la concesión de un derecho de exclusiva que recompense el esfuerzo intelectual. Un inventor debe ser susceptible de ostentar derechos, transmitirlos, explotarlos económicamente y asumir las responsabilidades inherentes a dicha titularidad, elementos incompatibles con la naturaleza carente de personalidad jurídica de los sistemas de IA. Además, la actividad inventiva presupone un proceso de creación consciente y deliberado, presupuesto que el derecho comparado reserva a los seres humanos. Por ello, aun cuando la IA pueda generar soluciones técnicas novedosas o innovadoras, no puede ser reconocida jurídicamente como inventora en el sentido estricto del derecho de patentes.

En el marco europeo, la normativa establece que los algoritmos considerados “como tales” carecen de patentabilidad, puesto que se clasifican como métodos abstractos. Sin embargo, cuando forman parte de una solución técnica

concreta —por ejemplo, un sistema que optimiza procesos de ciberseguridad, comunicación de datos o gestión computacional— pueden acceder al sistema de protección siempre que cumplan los requisitos sustantivos exigidos. Entre ellos, la apreciación de la actividad inventiva plantea dificultades particulares: los examinadores deben valorar la obviedad de soluciones generadas mediante procesos cuyo funcionamiento no es transparente, lo que desafía los parámetros tradicionales de evaluación. A ello se añade la exigencia de suficiencia descriptiva, que requiere que la invención pueda ser reproducida por un experto en la materia. En el ámbito de la IA, este requisito se ve comprometido debido a la opacidad inherente de los modelos de gran escala —que pueden integrar millones de parámetros— y a la falta de transparencia sobre los conjuntos de datos utilizados durante el entrenamiento, que en ocasiones no se encuentran debidamente documentados o son objeto de confidencialidad.

La inteligencia artificial no solo desafía los fundamentos del sistema de Propiedad Industrial, sino que también se ha convertido en una herramienta esencial para su modernización y mayor eficiencia. Su implementación en las oficinas de patentes y marcas a nivel internacional abarca diversas funciones. Los sistemas de traducción automática, como los desarrollados específicamente para documentación técnica, permiten superar barreras lingüísticas con un grado de precisión que agiliza significativamente la labor administrativa y reduce costes operativos. Asimismo, los algoritmos destinados a la clasificación automática de solicitudes en sistemas como IPC, CPC, Niza o Viena contribuyen a liberar recursos humanos para tareas de mayor valor añadido. En materia de búsqueda de anterioridades, la IA facilita la identificación de documentos relevantes incluso cuando presentan semejanzas no evidentes a simple vista, mejorando la exhaustividad y fiabilidad del examen. También se han incorporado herramientas para detectar similitudes en marcas e imágenes, lo que permite prevenir registros conflictivos y reforzar la coherencia del sistema. Finalmente, los asistentes digitales de presentación de solicitudes proporcionan apoyo a solicitantes con menor experiencia, favoreciendo el acceso al sistema de Protección Industrial. Este conjunto de innovaciones es especialmente relevante para países en desarrollo, pues posibilita saltos cualitativos en la gestión institucional sin necesidad de inversiones estructurales de gran magnitud.

4. INTELIGENCIA ARTIFICIAL EN EL AMBITO DE LA SALUD, LA BALANZA DE RIESGOS Y BENEFICIOS.

El desarrollo acelerado de las vacunas contra el COVID-19 demostró el potencial transformador de la Inteligencia Artificial (IA) en la vacunología moderna, acortando los plazos convencionales de años a solo meses. No obstante, esta aceleración fue el resultado de una convergencia de factores—incluyendo financiación masiva, plataformas de ARNm preexistentes, y colaboración

global—y no únicamente de la IA. A pesar de los avances, el papel específico de la IA en la investigación, el desarrollo, la distribución y la aceptación de vacunas ha permanecido disperso en múltiples revisiones.

La investigación tradicional en vacunas y su ciclo de I+D (Investigación y Desarrollo), que abarca décadas, se ve desafiada por la complejidad de procesos como el descubrimiento de antígenos, la predicción de epítomos y el diseño riguroso de ensayos clínicos, todos ellos aspectos en que los componentes éticos, jurídicos y de seguridad la propiedad intelectual e industrial deben tenerse muy presentes.

Así la inteligencia artificial se ha consolidado como una herramienta fundamental con el potencial de abordar algunos de los mayores desafíos del sector salud, como el agotamiento de la fuerza laboral, el envejecimiento de la población y la creciente complejidad de las enfermedades crónicas. Su impacto abarca desde la mejora de la atención clínica diaria hasta la aceleración de la innovación biomédica.

4.1. Oportunidades Clave en el Sistema Sanitario

La inteligencia artificial ofrece oportunidades decisivas para fortalecer los sistemas de salud. Una de las más importantes es la reducción de errores médicos mediante una comunicación más eficiente, asegurando que la información correcta llegue a la persona adecuada en el momento oportuno. En 2023 se estimó que unas 163.000 personas murieron en Europa debido a errores médicos, de los cuales cerca del 30% se relacionan con fallos de comunicación. La IA puede mitigar estos errores, mejorar los resultados clínicos y salvar vidas. También puede liberar a los profesionales sanitarios de tareas administrativas al automatizar hasta un 36% de las actividades del sector, permitiendo dedicar más tiempo a la atención directa y contribuyendo a aliviar el déficit proyectado de 3,5 millones de profesionales para 2030 en los países de la OCDE. Otro ámbito prioritario es la ciberseguridad: con pérdidas estimadas en 10,5 billones de dólares por ciberataques en 2025, los sistemas de salud son un objetivo crítico; la IA permite detectar amenazas y reforzar la infraestructura digital. Además, la mayor parte de los datos sanitarios —hasta el 97%— permanece sin utilizar; las capacidades analíticas de la IA permiten extraer valor de ellos para acelerar tratamientos, identificar patrones de enfermedad, impulsar la medicina personalizada y avanzar en la investigación, como se evidenció durante la pandemia con el rápido desarrollo de vacunas.

4.2. Aceleración de la Investigación y el Desarrollo de Vacunas

Una revisión de 27 estudios publicados entre 2020 y 2024 confirma que la IA es fundamental en todas las etapas del ciclo de vida de las vacunas. Tanto el aprendizaje automático tradicional, mediante algoritmos como *random forests*, SVM o regresión logística, como el *deep learning*, a través de arquitecturas CNN, RNN o GAN, han permitido mejorar la predicción de epítomos, el descubrimiento de antígenos y el diseño de vacunas multi-epítomo⁵⁶⁷, así como optimizar cadenas de suministro y simular ensayos clínicos. Gracias a estas técnicas, los tiempos de desarrollo se han reducido de años a meses —como se demostró con la COVID-19— mediante el mapeo acelerado de epítomos y el uso de diseños adaptativos. La integración de datos multi-ómicos ha facilitado identificar antígenos y epítomos conservados con mayor precisión, y herramientas como NetMHCpan o DeepVacPred han mostrado una elevada capacidad predictiva que se traduce en vacunas más inmunogénicas, especialmente en plataformas como ARNm o vacunas peptídicas para enfermedades complejas como el VIH o la malaria.

4.3. Riesgos y Desafíos Críticos en la Implementación de la IA

A pesar de su potencial, la adopción de la IA en salud enfrenta riesgos importantes. Los algoritmos sesgados o entrenados con datos no representativos pueden generar resultados deficientes, discriminatorios o poco confiables, lo que subraya la necesidad de guías nacionales e internacionales sobre transparencia, calidad y criterios de “IA responsable”. La gestión de datos sensibles también plantea desafíos: los sistemas sanitarios son blancos frecuentes de ciberataques y la legislación de protección de datos puede tensionar el desarrollo de modelos avanzados, requiriéndose un equilibrio entre privacidad y beneficios clínicos. A ello se suma la fragmentación del ecosistema digital, que dificulta la escalabilidad y amenaza con aumentar desigualdades si solo las organizaciones con mayores recursos logran aprovechar la IA. En el ámbito laboral, cerca del 27% de los empleos podrían automatizarse, lo que exige integrar a los profesionales en el diseño de herramientas y fortalecer sus competencias digitales para evitar mayor carga y agotamiento. Finalmente, la falta de responsabilidad clara y de marcos regulatorios sólidos —en un mercado que alcanzaría los 188.000 millones de dólares en 2030— genera incertidumbre sobre quién responde cuando la IA causa daño y dificulta la validación y adopción segura de nuevas herramientas.

⁵⁶⁷ <https://www.rekombiotech.com/es/multi-epitopo>

4.4. IA para la Aceptación Pública y la Comunicación

La IA también desempeña un papel creciente en la comprensión del sentimiento público y en la lucha contra la desinformación sobre vacunas. Mediante técnicas de procesamiento del lenguaje natural es posible monitorear redes sociales en tiempo real, identificar focos regionales de desinformación y detectar las preocupaciones que más influyen en la reticencia, como las dudas sobre nuevas tecnologías como las vacunas de ARNm. Los chatbots y agentes conversacionales ofrecen información personalizada y pueden contribuir a contrarrestar la desinformación si se auditan regularmente para asegurar su calidad. Además, modelos predictivos pueden orientar estrategias de salud pública mediante la anticipación de brotes, la asignación de recursos y el diseño de campañas dirigidas, integrando el desarrollo científico con la participación ciudadana para mejorar las tasas de aceptación y cobertura vacunal.

4.5. IA en el ámbito radiológico

La incorporación de sistemas de inteligencia artificial (IA) en los programas de cribado radiológico está transformando la práctica clínica contemporánea, especialmente en el diagnóstico precoz del cáncer de mama. En los servicios de radiología, la carga asistencial se caracteriza por un volumen elevado de estudios que deben revisarse de forma sistemática, pese a que solo una proporción mínima contiene hallazgos malignos. Esta situación se ve agravada por la escasez de especialistas disponibles para asumir la demanda creciente. En este contexto, la IA se perfila como una herramienta eficaz para optimizar la priorización de casos y reducir la carga de trabajo mediante el cribado automático de imágenes, situando los estudios sospechosos en posiciones prioritarias y disminuyendo la necesidad de doble lectura en casos de bajo riesgo.

La radiología constituye un ámbito especialmente propicio para la integración de soluciones basadas en IA debido a la naturaleza visual de los estudios diagnósticos y a la estructura intrínseca de las imágenes médicas, que facilita la identificación automatizada de patrones. La evidencia científica disponible indica que los modelos comerciales pueden reducir de manera significativa el volumen de lecturas necesarias, al tiempo que contribuyen a mejorar la precisión diagnóstica cuando se combinan con la evaluación humana. En el caso del cribado mamográfico, la estandarización del procedimiento —basado en proyecciones constantes— ha favorecido el desarrollo de modelos comerciales capaces de clasificar las imágenes como estándar o sospechosas y de señalar áreas de interés que posteriormente son verificadas por el profesional sanitario. Esta asistencia tecnológica permite priorizar los estudios que requieren atención inmediata y contribuye a minimizar el riesgo de pasar por alto hallazgos sutiles, especialmente en situaciones de elevada carga laboral.

No todas las modalidades de imagen presentan el mismo grado de estandarización. La resonancia magnética, por ejemplo, se caracteriza por una notable heterogeneidad en los protocolos empleados, lo que produce una elevada variabilidad en cuanto a contraste, resolución, artefactos y parámetros de adquisición. Esta diversidad dificulta el entrenamiento de modelos robustos y exige un preprocesamiento complejo, convirtiéndose en un obstáculo significativo para la adopción masiva de soluciones basadas en IA. Aunque los sistemas sanitarios poseen grandes volúmenes de imágenes, estas se encuentran fragmentadas, no estandarizadas y sujetas a fuertes restricciones legales y organizativas que limitan su aprovechamiento para el entrenamiento de algoritmos.

El uso de datos clínicos plantea también desafíos relacionados con la comunicación pública y la percepción social. La falta de claridad sobre la finalidad y el tratamiento de los datos puede generar desconfianza, a pesar de que los modelos destinados a investigación y desarrollo pueden entrenarse eficazmente mediante datos anonimados que no contienen identificadores personales. En respuesta a esta necesidad, diversos sistemas sanitarios avanzan hacia la creación de repositorios públicos anonimizados que permiten utilizar grandes conjuntos de datos con garantías de protección y equidad. Asimismo, algunos centros han comenzado a desarrollar herramientas propias con el fin de garantizar autonomía tecnológica y homogeneidad en las pruebas de cribado, lo que requiere interoperabilidad entre proveedores y plataformas tecnológicas que faciliten la integración de modelos en infraestructuras hospitalarias heterogéneas.

La integración práctica de modelos de IA en entornos asistenciales constituye un reto considerable. La incorporación de estas herramientas implica conectar bases de datos clínicas, adaptar interfaces, ajustar flujos de trabajo y compatibilizar sistemas con décadas de uso. A ello se suma la necesidad de superar la brecha existente entre quienes desarrollan los algoritmos y quienes los utilizan en la práctica clínica. Esta situación demanda la presencia de perfiles profesionales híbridos capaces de traducir los requisitos técnicos a necesidades asistenciales y viceversa, así como el desarrollo de plataformas estandarizadas que faciliten la integración y actualización de modelos.

Desde el punto de vista regulatorio, el marco normativo europeo clasifica como de alto riesgo cualquier sistema algorítmico que intervenga en decisiones relacionadas con la salud, lo que implica evaluar sesgos, asegurar trazabilidad y limitar la autonomía operativa del modelo. En este esquema, la responsabilidad última recae siempre en el profesional sanitario, lo que subraya la importancia de mantener la supervisión humana en todo proceso de decisión clínica asistida por IA.

La evidencia disponible muestra que la IA médica posee la capacidad de identificar patrones que pueden pasar inadvertidos al ojo humano, lo que confirma su potencial como herramienta complementaria en el diagnóstico radiológico. No obstante, su expansión dependerá fundamentalmente de decisiones organizativas y estructurales: la armonización de protocolos, la interoperabilidad entre sistemas hospitalarios, la disponibilidad de datos anonimizados, la integración tecnológica eficiente y la garantía de que los avances lleguen a todos los centros sanitarios de manera equitativa. La IA constituye, así, un recurso con enorme potencial transformador, cuya eficacia futura estará condicionada por la capacidad institucional de gestionar sus implicaciones éticas, regulatorias y operativas.

5. CONCLUSIONES

El análisis realizado permite concluir que el desarrollo y uso de la Inteligencia Artificial (IA), particularmente en sus aplicaciones generativas, ha forzado una revisión estructural de los regímenes de Propiedad Intelectual (PI) y Propiedad Industrial (PII) a nivel internacional, al tensionar un sistema normativo históricamente construido sobre la premisa de la creación exclusivamente humana.

En el ámbito del derecho de autor, persiste un paradigma estrictamente antropocéntrico: tanto la legislación española (LPI) como los tratados internacionales, entre ellos el Convenio de Berna, establecen que únicamente las personas naturales pueden ostentar la condición de autor. En consecuencia, los contenidos generados de manera íntegra por sistemas de IA sin una aportación creativa humana sustancial carecen de autor jurídicamente reconocible y se incorporan directamente al dominio público, sin posibilidad de protección. Únicamente las obras híbridas, en las que el usuario introduce contribuciones creativas relevantes —ya sea mediante la selección, organización, modificación o aplicación de criterios estéticos sobre el material producido por la IA— pueden acogerse al régimen de protección propio del derecho de autor. Asimismo, los prompts o instrucciones utilizadas para generar contenidos no son protegibles por esta vía al considerarse meras ideas, métodos u operaciones, si bien pueden recibir protección como secretos empresariales, de conformidad con la Ley 1/2019, cuando representen un activo competitivo y confidencial dentro de una estrategia empresarial.

En materia de propiedad industrial, y específicamente en el campo de las patentes, se reafirma que la IA no puede ostentar la condición de inventora, puesto que el marco normativo internacional exige que tal titularidad corresponda a una persona natural dotada de capacidad jurídica, voluntad y

responsabilidad. Aunque los sistemas de IA contribuyen de forma decisiva a la identificación de soluciones técnicas, la ausencia de personalidad jurídica impide atribuirles derechos o responsabilidades. Además, la creciente opacidad técnica de modelos avanzados, especialmente los basados en deep learning, plantea dificultades relevantes para satisfacer requisitos esenciales del sistema de patentes, como la suficiencia descriptiva, la accesibilidad técnica y la reproducibilidad de la invención.

Desde la perspectiva regulatoria, destaca el Reglamento (UE) 2024/1689 (AI Act), que establece obligaciones reforzadas de transparencia, trazabilidad y conformidad con las normas de PI y PII para los proveedores de sistemas de IA, con el objetivo de garantizar una integración segura, ética y respetuosa con los derechos fundamentales. Persisten, sin embargo, importantes tensiones jurídicas, como las derivadas del uso masivo de datos (data scraping) para entrenar modelos, práctica que puede originar reproducciones no autorizadas y vulneraciones de derechos de autor o de la protección sui generis de bases de datos. En este contexto, el marco europeo ha impulsado mecanismos que permiten a los titulares excluir sus obras de estos procesos. A medio plazo, parte de la doctrina apunta a la conveniencia de evaluar la creación de un régimen sui generis específico para los resultados generados automáticamente por IA, con una protección de carácter estrictamente económico, basada en la inversión y de duración limitada, que no altere los principios rectores del derecho de autor.

En el sector sanitario, la IA despliega un papel estratégico, al permitir la aceleración de procesos de investigación y desarrollo —como se evidenció durante la pandemia de COVID-19—, la reducción de errores médicos, particularmente los derivados de fallos de comunicación, y la optimización del tiempo de los profesionales. No obstante, su implementación comporta riesgos de considerable relevancia jurídica, entre los que destacan los resultados deficientes derivados de sesgos en los algoritmos, las posibles vulneraciones del derecho fundamental a la protección de datos personales, la fragmentación tecnológica que puede agravar desigualdades existentes y la ausencia de marcos claros de responsabilidad civil o administrativa cuando un sistema de IA provoca daños. En ámbitos clínicos críticos, como la radiología, la normativa europea clasifica estos sistemas como de alto riesgo, imponiendo la obligación de mantener la supervisión humana y reservando la responsabilidad última al profesional sanitario.

En suma, el desafío central consiste en articular un equilibrio adecuado entre la promoción de la innovación tecnológica y la preservación de la seguridad jurídica, los derechos de los creadores y los principios fundamentales del ordenamiento. La evolución normativa deberá ser sistemática, coherente y multidisciplinar para garantizar que el avance digital se produzca sin erosionar los

cimientos del sistema jurídico de PI y PII ni comprometer la protección de los usuarios y de la sociedad en su conjunto.

APÉNDICE BIBLIOGRÁFICO

I. LIBROS Y CAPÍTULOS DE LIBRO

Copyright and Patent Laws for the Age of Artificial Intelligence. 2025. Oxford: Hart Publishing.

Kazeeva, V., S. Meyer, et al. 2025. "Intellectual Property Law in the Age of Artificial Intelligence: Legal Challenges and Regulatory Perspectives." *International Journal Law and Society*.

Lucchi, Nicola. 2025. Generative AI and Copyright. European Parliament Study.
https://www.europarl.europa.eu/RegData/etudes/STUD/2025/774095/IUST_STU%282025%29774095_EN.pdf

Molina, Óscar A. 2024. "Regulación de la inteligencia artificial en la sanidad española." En *Salud y sostenibilidad: Una mirada desde el derecho positivo*. Madrid: Aranzadi.

Plaza Penadés, Javier. 2025. *Propiedad intelectual y protección de sistemas de inteligencia artificial y metaversos*. Pamplona: Aranzadi.

Tschider, C. 2024. "Artificial Intelligence and Intellectual Property in Healthcare." En *Artificial Intelligence and Intellectual Property Law in the Age of Intelligence*. SSRN.

Vanherpe, Jozefien. 2025. "Artificial Intelligence and Intellectual Property Law." En *The Cambridge Handbook of the Law, Ethics and Policy of Artificial Intelligence*, editado por Nathalie A. Smuha, 211–27. Cambridge: Cambridge University Press.

II. ARTÍCULOS ACADÉMICOS (REVISTAS CIENTÍFICAS)

Al-Busaidi, A. S. 2024. "Investigating the Impact of Generative Artificial Intelligence on Intellectual Property Rights." *Journal of Intellectual Property Research*.

<https://www.sciencedirect.com/science/article/pii/S2444569X24001690>

Atkinson, David. 2025. "Unfair Learning: GenAI Exceptionalism and Copyright Law." *arXiv* (2025).

<https://arxiv.org/abs/2504.00955>

Chen, Yang. 2025. "Reviving 'Computer-Generated Works': Should Hong Kong Copyright Law Adapt the Rule to Harness AI Opportunities?" *Journal of Intellectual Property Law & Practice* 20 (9): 584–93.

<https://academic.oup.com/jiplp/issue/20/9>

Hanna, Kai. 2025. "Artificial Intelligence: Redefining Intellectual Property Frameworks." *Washington University Law Review*.

<https://journals.library.wustl.edu/wuulr/article/9079/galley/25837/view/>

Kaur, G. 2025. "The Intersection of Law, Ethics, and Healthcare AI: Policy Pathways for Responsible Innovation." *Legal Health Policy Review*.

Lim, Daryl. 2025. "Banana Republic: Copyright Law and the Extractive Logic of Generative AI." *Journal of Intellectual Property Law & Practice* 20 (9): 573–83.

<https://academic.oup.com/jiplp/issue/20/9>

Lou, Chanhon. 2025. "Develop-Fair Use for Artificial Intelligence: A Sino-U.S. Copyright Law Comparison." *arXiv* (2025).

<https://arxiv.org/abs/2509.07365>

Matulionyte, Rita, y Jyh-An Lee. 2025. "Copyright in AI-Generated Works: Lessons from Patent Law." *arXiv* (2025).

<https://arxiv.org/abs/2503.04738>

Pham, T., et al. 2025. "Ethical and Legal Considerations in Healthcare AI." *Royal Society Open Science* 12 (5): 241873.

<https://royalsocietypublishing.org/rsos/article/12/5/241873>

Pham, T., y S. Sayın. 2024. "Understanding Artificial Intelligence within Legal and Ethical Issues." *Journal of Oncological Sciences* 10 (2): 105–14.

<https://journalofoncology.org/pdf/50c58720-e50d-4950-b386-1d4e1867ef9a/articles/jos.2023-99598/105.pdf>

Podar, A., et al. 2024. “Evolving Intellectual Property Landscape for AI-Driven Biomedical Innovations.” *Frontiers in Artificial Intelligence* 7: 1372161.

<https://www.frontiersin.org/articles/10.3389/frai.2024.1372161/full>

Popov, V., et al. 2025. “Artificial Intelligence and Intellectual Property Law in the Age of Intelligence.” *International Journal Law and Society*.

<https://www.researchgate.net/publication/393492920>

Repanovici, Angela, y Gabriela Ivanus. 2025. “Industrial Property of Patents Generated by Artificial Intelligence.” *International Maritime Transport and Logistic Journal*.

<https://apc.aast.edu/ojs/index.php/MARLOG/article/view/MARLOG.2025.14.1.57>

Romero, M. 2025. “Aproximación o amenaza de la inteligencia artificial en la práctica de propiedad industrial.” *Revista de Actualidad Mercantil* 9: 67–77.

<https://revistas.pucp.edu.pe/index.php/actualidadmercantil/article/view/30906>

Šarčević, Tanja, et al. 2024. “U Can’t Gen This? A Survey of Intellectual Property Protection Methods for Data in Generative AI.” *arXiv* (2024).

<https://arxiv.org/abs/2406.15386>

van Kessel, R. 2025. “Artificial Intelligence, Intellectual Property, and Human Rights.” *npj Health Systems*.

<https://www.nature.com/articles/s44401-025-00050-3>

Wubineh, B. Z., et al. 2024. “Exploring Opportunities and Challenges of Artificial Intelligence in Health.” *Health Informatics Review*.

III. LEGISLACIÓN, NORMATIVA Y DOCUMENTOS OFICIALES

Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS). 1995.

https://www.wto.org/english/docs_e/legal_e/27-trips.pdf

- España. 2019. Ley 1/2019, de 20 de febrero, de Secretos Empresariales. Boletín Oficial del Estado, núm. 44. <https://www.boe.es/buscar/act.php?id=BOE-A-2019-2182>
- España. 1996. Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el Texto Refundido de la Ley de Propiedad Intelectual (TRLPI). Boletín Oficial del Estado. <https://www.boe.es/buscar/act.php?id=BOE-A-1996-8930>
- European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 12 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (AI Act). Official Journal of the European Union. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689>
- European Union. 2016. Directive (EU) 2016/943 on Trade Secrets. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016L0943>
- European Union. 2024. Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act). <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Organización Mundial de la Propiedad Intelectual (OMPI/WIPO). 1886. Berne Convention for the Protection of Literary and Artistic Works (Berne Convention). Ginebra: OMPI. <https://www.wipo.int/treaties/en/ip/berne/>
- Organización Mundial de la Propiedad Intelectual (OMPI/WIPO). 1883. Paris Convention for the Protection of Industrial Property. Ginebra: OMPI. <https://www.wipo.int/treaties/en/ip/paris/>
- Organización Mundial de la Propiedad Intelectual (OMPI/WIPO). 1996. WIPO Copyright Treaty (WCT). Ginebra: OMPI. <https://www.wipo.int/treaties/en/ip/wct/>
- Organización Mundial de la Propiedad Intelectual (OMPI/WIPO). 1996. WIPO Internet Treaties (WCT & WPPT). Ginebra: OMPI. <https://www.wipo.int/treaties/en/>

Organización de Naciones Unidas. 1948. Universal Declaration of Human Rights. Artículo 27. Nueva York: ONU. <https://www.un.org/en/about-us/universal-declaration-of-human-rights>

United States Congress. 2024. Generative AI Copyright Disclosure Act (H.R. 7913).
https://en.wikipedia.org/wiki/Generative_AI_Copyright_Disclosure_Act

U.S. Copyright Office. 2025. Copyright and Artificial Intelligence, Part 3: Generative AI Training.
<https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>

World Health Organization. 2021. Ethics and Governance of Artificial Intelligence for Health.
<https://www.who.int/publications/i/item/9789240029216>

World Intellectual Property Organization. 2023. Artificial Intelligence and Intellectual Property: Policy and Technical Reports.
<https://www.wipo.int>

IV. OTRAS FUENTES INSTITUCIONALES / RECURSOS ACADÉMICOS

España. Ministerio para la Transformación Digital. 2025. Anteproyecto de Ley para el Buen Uso y la Gobernanza de la Inteligencia Artificial. Madrid. (Iniciativa legislativa anunciada).
<https://www.mintdigit.gob.es>

European Commission. 2024. Artificial Intelligence in Healthcare.
https://health.ec.europa.eu/ehealth-digital-health-and-care/artificial-intelligence-healthcare_es

Organización Mundial de la Propiedad Intelectual (OMPI/WIPO). 2024. WIPO IP Statistics Data Center. Ginebra: OMPI.
<https://www3.wipo.int/ipstats/>

Organización para la Cooperación y el Desarrollo Económicos (OCDE/OECD). 2024. Health Workforce Outlook 2024. París: OCDE. <https://www.oecd.org/health/>

Okoro, D. C. 2025. An Intellectual Property Approach to the Use of Generative AI.

<https://www.iiardjournals.org/get/WJIMT/VOL.%209%20NO.%204%202025/An%20Intellectual%20Property%2097-116.pdf>

Verma, Manish, Vivek Sharma, y Vishal Singh. 2025. "Artificial Intelligence in Patent and Market Intelligence: A New Paradigm for Technology Scouting." arXiv (2025).

<https://arxiv.org/abs/2507.20322>



