TRABAJO FIN DE MÁSTER

MÁSTER EN DERECHOS HUMANOS Y MECANISMOS DE PROTECCIÓN NACIONAL E INTERNACIONAL

CURSO ACADÉMICO 2024-2025

El Impacto de la Inteligencia Artificial en los Derechos Humanos y su Protección Internacional

The Impact of Artificial Intelligence on Human Rights and their International Protection

AUTORA: YAIZA SANTOS DÍEZ

TUTORA: YAELLE CACHO SÁNCHEZ

Junio de 2025

RESUMEN

El presente trabajo aborda el impacto de la inteligencia artificial (IA) sobre los derechos humanos, analizando tanto los beneficios que presenta para mejorar su disfrute como los riesgos que entraña, tales como la existencia de sesgos algorítmicos discriminatorios. Este diagnóstico confirma la necesidad de un marco regulatorio que vaya más allá del ámbito nacional. El objetivo de nuestro trabajo es precisamente analizar las propuestas que ya existen en la escena internacional con el fin de valorar en qué medida afrontan los riesgos que la IA tiene para los derechos humanos.

Comenzamos examinando las propuestas de *soft law* de algunos organismos internacionales, que incluyen una serie de principios y valores que han de preservarse durante todo el ciclo de vida de los sistemas de IA. Pero sobre todo nos centraremos en el estudio de la Ley de inteligencia artificial de la Unión Europea y del Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, democracia y Estado de Derecho, que suponen un gran paso al tratarse en ambos casos de normas de carácter jurídicamente vinculante.

Palabras clave: Derechos Humanos, discriminación, ética, Inteligencia Artificial, sesgos algorítmicos, supervisión.

ABSTRACT

This paper addresses the impact of artificial intelligence (AI) on human rights, analysing both the benefits it presents to enhance their enjoyment and the risks it entails, such as the existence of discriminatory algorithmic biases. This diagnosis confirms the need for a regulatory framework that goes beyond the national level. The aim of our work is precisely to analyse the proposals that already exist on the international scene in order to assess the extent to which they address the risks that AI poses to human rights.

We begin by examining the soft law proposals of some international bodies, which include a series of principles and values that must be preserved throughout the life cycle of AI systems. But above all, we will focus on the study of the European Union´s Artificial Intelligence Act and the Council of Europe´s Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, both of which are legally binding norms.

Keywords: Human Rights, discrimination, ethics, Artificial Intelligence, algorithmic biases, supervisión.

ÍNDICE

ABREVIATURAS

CAI	Comité sobre Inteligencia Artificial
CAHAI	Comité Ad Hoc sobre Inteligencia Artificial
COMPAS	Correctional Offender Management Profiling for Alternative Sanctions
IA	Inteligencia Artificial
OCDE	Organización para la Cooperación y el Desarrollo Económico
RIA	Reglamento de Inteligencia Artificial de la Unión Europea
UE	Unión Europea
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

I. INTRODUCCIÓN

Los sistemas de inteligencia artificial contribuyen a generar muy diversos beneficios económicos, medioambientales y sociales y ofrecen oportunidades sin precedentes para proteger y promover los derechos humanos, la democracia y el Estado de Derecho. Pero también existen graves peligros derivados de determinadas actividades realizadas como parte de su ciclo de vida. Su impacto negativo se acrecienta en la medida en que cubre un amplio campo de tecnologías en rápida evolución, que pueden desplegarse fácilmente a escala transfronteriza. Por otro lado, aunque algunos Estados ya han estudiado la adopción de normas nacionales destinadas a garantizar que la IA sea fiable y segura y se desarrolle y utilice de conformidad con las obligaciones relativas a los derechos humanos, no todos han seguido este camino. A ello se le suma el riesgo de que surjan marcos regulatorios divergentes, dificultando una regulación coherente de estas tecnologías a nivel internacional.

A través del presente trabajo se pretende abordar el examen de los nuevos desafíos a los que nos enfrenta la IA en materia de derechos humanos y el estudio en profundidad de los mecanismos internacionales que se están adoptando para hacerles frente. En primer lugar, se examinarán algunas de las propuestas de *soft law* más relevantes debido a su contenido basado en principios y valores éticos en el diseño, desarrollo y uso de sistemas de inteligencia artificial. A continuación, se analizarán dos instrumentos jurídicos que resultan del máximo interés, el Reglamento (UE) 2024/1689 de la Unión Europea, de 13 de junio de 2024, y el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y derechos humanos, democracia y Estado de derecho, de 5 de septiembre de 2024, en tanto que no solo es el primer tratado internacional jurídicamente vinculante en la materia, sino que además está abierto a países no europeos. Por último, se analizaría en qué medida estos nuevos instrumentos internacionales adoptados para mitigar los riesgos de la IA o paliar sus daños garantizan efectivamente el respeto de la igualdad, incluida la igualdad de género, la prohibición de la discriminación y el derecho a la intimidad, entre otros derechos humanos cuyo disfrute se haya más directamente afectado por la IA.

2. LA IA Y SU IMPACTO EN LOS DERECHOS HUMANOS

No cabe duda de la importancia que está adquiriendo la Inteligencia Artificial (IA) en la configuración de nuevas formas de desarrollo de tareas que hasta ahora llevábamos a cabo las personas. Aún no conocemos todo el potencial que la IA presenta, tanto en su vertiente positiva como negativa, pero ya es posible identificar algunos de sus mayores beneficios, aunque también muchos de sus peligros, especialmente en cuanto al disfrute de los derechos humanos.

Antes de abordar el impacto que tiene la IA sobre los derechos humanos, parece necesario identificar qué debe entenderse por IA. Sin embargo, aún no se ha proporcionado una definición concreta que pueda abarcar todas las funcionalidades que la IA tiene, sino que se está trabajando sobre la base de definiciones que dejan un amplio espacio a las posibles modificaciones que puedan ir produciéndose. De este modo, la Comisión Europea se ha referido a la IA como "sistemas de software (y posiblemente también de hardware) diseñados por humanos que, ante un objetivo complejo, actúan en la dimensión física o digital: percibiendo su entorno, a través de la adquisición e interpretación de datos estructurados o no estructurados, 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, y también pueden adaptar su comportamiento al analizar cómo el medio ambiente se ve afectado por sus acciones previas"[1]. En el plano español, esta es la definición que se ha adoptado también en la Estrategia Nacional de Inteligencia Artificial (ENIA), la cual pretende promover que estos sistemas respeten los derechos fundamentales, la equidad en el acceso y prevención contra la discriminación[2].

Se puede deducir entonces que la IA se trata de un conjunto de algoritmos capaz de procesar información, aprender datos y tomar decisiones a partir de ellos, es decir, genera comportamientos inteligentes a través de un aprendizaje automático. Su capacidad de interpretación de datos que se documentan, generan y almacenan en dispositivos

[1] High-Level Expert Group on Artificial Intelligence, European Comission. (2019). *A Definition of AI: Main Capabilities and Disciplines*, p. 6.

[2] Ministerio de Asuntos Económicos y Transformación Digital, Gobierno de España (2020). *Estrategia Nacional de Inteligencia Artificial*, p. 66.

electrónicos, así como la reacción a dichos datos y la comunicación entre los mismos, han generado lo que ha venido en llamarse *«big data»*, o macrodatos[3].

Comprender mejor el fenómeno de la IA requiere de una serie de precisiones terminológicas adicionales. De un lado nos encontramos la categoría del *Machine Learning* que desarrolla algoritmos permitiendo a las máquinas aprender a partir de una serie de datos y crear sus propias reglas para encontrar soluciones óptimas a problemas específicos[4]. Esta categoría y sus amplias capacidades de búsqueda de patrones en los datos se han llegado a aplicar en una variedad de funciones empresariales entre las que se incluye el marketing, ya que ofrece a los profesionales del ámbito la oportunidad de obtener nuevos conocimientos sobre el comportamiento de los consumidores. De modo que, este tipo de aprendizaje mejora su conocimiento y rendimiento en las tareas en base a la experiencia[5].

Dentro del *Machine Learning* existe una subcategoría conocida como *Deep Learning*, la cual imita el proceso de aprendizaje del ser humano a través del uso de redes neuronales artificiales profundas y logra modelar y entender patrones complejos en grandes volúmenes de datos. Su efectividad se expone mayormente a la hora de identificar y clasificar objetos en imágenes y vídeos para aplicaciones como la vigilancia de seguridad o en el caso del procesamiento del lenguaje natural empleado en asistentes virtuales[6].

Por lo tanto, las diversas funcionalidades que tiene la inteligencia artificial y su complejidad llevan a concluir que se trata de una poderosa herramienta que puede traer numerosos beneficios para la sociedad, sin embargo, comprende al mismo tiempo grandes riesgos en su desarrollo y utilización.

2.1. LOS BENEFICIOS DE LA IA

La inseguridad que las nuevas tecnologías provocan en las personas no debe frenar el avance científico-tecnológico y deben aprovecharse todas las utilidades que pudiera

[3] Kriebitz, A. y Lütge, C. (2024). La inteligencia artificial y sus consecuencias para los derechos humanos: una evaluación desde la ética de la empresa. En E. González-Esteban y J. C. Siurana Aparisi (Eds.), *Inteligencia Artificial: concepto, alcance, retos* (p. 149). Tirant Humanidades.

[4] Rodríguez Torres, Á. F., Rodríguez Alvear, F. S., Collaguazo Lapo, D. R., y Rodríguez Alvear, J. C. (2024). Diferencias y Aplicaciones de Big Data, Inteligencia Artificial, Machine Learning y Deep Learning. *Dominio de las Ciencias,* 10 (3), 963-966.

[5] Herhausen, D., Bernritter, S. F., Ngai, E. W. T., Kumar, A., y Delen, D. (2024). Machine learning in marketing: Recent progress and future research directions. *Journal of Business Research,* 170, 1- 2.

[6] Rodríguez Torres, Á. F., Rodríguez Alvear, F. S., Collaguazo Lapo, D. R., y Rodríguez Alvear, J. C. (2024), *op. cit.*, p. 966.

proporcionarnos. Podemos afirmar entonces que la IA puede tener grandes beneficios para la sociedad en diferentes ámbitos. Entre todos los aspectos sobre los que incide, cabe mencionar algunos de ellos por su capacidad para tener un impacto sobre la protección y mejor disfrute de los derechos humanos.

Sin duda uno de los avances más importantes que permite esta herramienta se da en el campo de la salud, mejorando la atención al paciente al garantizar una mejora en los suministros tecnológicos y la maquinaria empleada para llevar a cabo cirugías de diversa índole. La creación de estas máquinas va a ayudar a agilizar el análisis de datos médicos y de las imágenes extraídas de la realización de las pruebas diagnósticas, lo que ciertamente constituye una ventaja a la hora de resolver ciertas patologías y que va a fortalecer la protección del derecho a la salud de las personas y a la asistencia sanitaria[7].

Un claro ejemplo es el caso de los asistentes robóticos para cirugías que permiten al profesional médico realizar operaciones sin la necesidad de que doctor y paciente se encuentren en el mismo espacio geográfico[8]. Igualmente, la IA ha sido aplicada en el área de la investigación biomédica, tanto en enfermedades prevalentes como el cáncer, como en enfermedades raras[9].

Entre las máquinas de IA en este sector nos encontramos con Corti, un sistema capaz de salvar vidas al diagnosticar un ataque al corazón con el uso de un teléfono móvil. Su funcionamiento se da a través del análisis de datos al detectar la voz en una llamada de emergencia, lo que facilita la tarea de los centros médicos y servicios de emergencia[10].

La IA tiene también gran influencia en el cuidado de enfermería porque a través de la automatización de procesos y la asistencia en la toma de decisiones clínicas se permite a los profesionales sanitarios una mejor gestión organizativa y ofrecer así una atención más eficiente y de mayor calidad[11]. Los aportes de la IA han facilitado la personalización y precisión de los tratamientos y han hecho más efectiva la atención de los pacientes al adaptar los cuidados a sus características individuales, ya que esta herramienta ayuda a la

---

[7] Márquez Muñoz, L. (2023). La Inteligencia Artificial aplicada al ámbito sanitario: retos en el futuro. En F. Pérez Tortosa (Coord.). *La Justicia en la Sociedad 4.0: Nuevos Retos para el Siglo XXI* (p. 85). Colex.

[8] Martínez García, D. N., Dalgo Flores, V. M., Herrera López, J. L., Analuisa Jiménez, E. I., y Velasco Acurio, E. F. (2019). Avances de la inteligencia artificial en salud. *Dominio de las Ciencias*, *5* (3), p. 609.

[9] Sánchez Rosado, J. C. y Díez Parra, M. (2022). Impacto de la Inteligencia Artificial en la transformación de la sanidad: beneficios y retos. *Economía Industrial* (423), p.131.

[10] Angulo, S. (2018, 7 febrero). *4 ejemplos de inteligencia artificial para la medicina*. Apps y Software, Enter.co. Disponible en: https://www.enter.co/chips-bits/apps-software/ejemplos-inteligencia-artificial-salud/ .

[11] Carrión Bósquez, N. G., Castelo Rivas, W. P., Alcívar Muñóz, M. M., Quiñonez Cedeño, L. P. y Llambo Jami, H. S. (2022). Influencia de la COVID-19 en el clima laboral de trabajadores de la salud en Ecuador. *Revista Información Científica*, *101* (1).

identificación de patrones y tendencias en los datos. Se mejoran así los tiempos y la calidad de los cuidados aumentando la satisfacción del paciente[12]. Por lo tanto, la IA facilita y mejora el trabajo humano en la salud, sin suplantar la labor de los profesionales médicos[13].

Se puede observar también cómo la IA puede ser muy útil en el ámbito educativo, especialmente por su capacidad para ajustar el método de enseñanza a las necesidades y destrezas particulares de cada alumno[14]. Se produce así una personalización del contenido educativo gracias al análisis de datos de los estudiantes y la filtración de la información irrelevante para poder hacer recomendaciones de estudio más específicas para cada uno de ellos. Igualmente, no sólo beneficia a los estudiantes de los distintos niveles de educación, sino que facilita la labor del docente al asistirle en tareas de planificación didáctica o de evaluación de los alumnos. La creación de tutores virtuales para la resolución de dudas de los alumnos logra proporcionar una retroalimentación inmediata sobre su trabajo que les ayudará a mejorar su rendimiento[15].
Desde una perspectiva más amplia, la IA permite activar el proceso para la consecución del número 4 de los Objetivos de Desarrollo Sostenible en el que se apuesta por "Garantizar una educación inclusiva, equitativa y de calidad y promover oportunidades de aprendizaje durante toda la vida para todos"[16].

Asimismo, a nivel administrativo la IA tiene un gran potencial. El uso de softwares en las tareas de gestión y administración de la justicia o de la Seguridad Social permite agilizar procesos mecánicos en los que no es tan necesaria la intervención humana. Como organismos públicos se hace necesario responder a las dudas que plantean los ciudadanos de forma más ágil y eficiente, de modo que los asistentes virtuales que se emplean en las Administraciones permiten dar respuestas automatizadas a las preguntas más comunes o redirigir a la persona al departamento correspondiente en función de sus necesidades[17].

[12] Jaramillo Verduga, M. J. y Alarcón Dalgo, C. M. Á. (2024). Influencia de la Inteligencia Artificial en el Cuidado de Enfermería y su Reto. *Ciencia Latina Revista Científica Multidisciplinar, 8* (5), p. 5.
[13] Demir Kaymak, Z. T. (2024). Efectos de la preparación de los estudiantes de obstetricia y enfermería sobre la inteligencia artificial médica en la ansiedad por la inteligencia artificial. *Science Direct*, 7.
[14] Fitria, T. N. (2021). The use technology based on artificial intelligence in english teaching and learning. *ELT Echo: The Journal of English Language Teaching in Foreign Language Context, 6* (2), p. 218.
[15] Romero García-Aranda, C. (2024). La Inteligencia Artificial en la Educación: una visión de la UNESCO. *Dignitas: Revista On-line sobre Derechos Humanos y Relaciones Internacionales,* (7), 24-27.
[16] *Ibid*., p. 15.
[17] Fernández Ramírez, M. (2023). Inteligencia artificial, algoritmos predictivos y gestión tecnológica de la Seguridad Social. En Asociación Española de Salud y Seguridad Social (Coord.), *Las transformaciones de*

Es creciente el número de gobiernos que utilizan estos sistemas para tomar de manera automática decisiones que inciden en derechos individuales, tales como la selección de los beneficiarios de servicios sociales o la determinación de los servicios sanitarios a los que tiene derecho cada persona. Incluso se está produciendo una automatización de procedimientos administrativos sancionadores, tal como ha ocurrido en España con el RD 688/2021, de 3 de agosto, que modifica el *Reglamento general sobre procedimientos para la imposición de sanciones por infracciones de orden social y para los expedientes liquidatorios de cuotas de la Seguridad Social*. Esto permite que la actividad de la Inspección se realice de forma automatizada, es decir, sin requerir intervención humana, basándose en un análisis masivo de datos para detectar incumplimientos[18].

El empleo de sistemas de inteligencia artificial resulta muy útil para los juristas en sus tareas de investigación al facilitar la sistematización y búsqueda de información jurídica relevante. De este modo, no incide solamente en el ámbito administrativo, sino que también está provocando cambios en el funcionamiento del sistema judicial e incluso en el rol judicial[19]. Claro ejemplo de ello son los sistemas de análisis predictivo que permiten la predicción a partir de decisiones judiciales ya adoptadas que podrían ser replicadas[20]. Estas técnicas de predicción judicial son las grandes innovaciones que introduce la IA en el curso de un procedimiento legal, cuya función es predecir el comportamiento de un individuo determinando su nivel de riesgo y se utilizan para tomar decisiones, por ejemplo, en cuanto a la aplicación de medidas o tras la condena en un proceso judicial[21]. Uno de los algoritmos que ha tenido mayor análisis en su aplicación es el COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), una herramienta que valora el riesgo de reincidencia del procesado y las necesidades criminológicas del sujeto. Su objeto es reducir la población carcelaria en base al uso de una serie de algoritmos que, según los antecedentes penales del acusado, predicen el posible nivel de reincidencia[22]. El uso de este algoritmo ha sido implementado en la

la Seguridad Social ante los retos de la era digital. Por una salud y Seguridad Social digna e inclusiva (p.14). Laborum.

[18] Solar Cayón, J. I. (2022). Inteligencia artificial y justicia digital. En Llano Alonso, F. H. (Dir.), *Inteligencia artificial y Filosofía del derecho* (p. 384). Laborum.

[19] Sourdin, T. (2018). Judge v. Robot? Artificial Intelligence and judicial decision-making. *UNSW Law Journal, 41* (4), p. 1115.

[20] Megías Quirós, J. J. (2022). Derechos humanos e Inteligencia Artificial. *Dikaiosyne,* 37, p.141.

[21] Miró Linares, F. (2018). Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots. *Revista de Derecho Penal y Criminología*, 20, p.107.

[22] Muñoz Rodríguez, A. B. (2020). El impacto de la inteligencia artificial en el proceso penal. *Anuario de la Facultad de Derecho*, 36, 702-703.

justicia norteamericana, suponiendo un hito el caso de Wisconsin State vs Loomis, donde un tribunal se pronunció por primera vez sobre la admisibilidad del uso de herramientas de IA dentro del proceso penal. En el mencionado caso fue admitido el informe aportado por la Fiscalía en el que se exponía el resultado del algoritmo COMPAS estableciendo el alto grado de riesgo de reincidencia que presentaba el acusado[23].

Otra experiencia reseñable viene constituida por el empleo de PROMETEA en el Tribunal Superior de Justicia de la Ciudad Autónoma de Buenos Aires, un sistema de *Machine Learning* supervisado que es utilizado por la Fiscalía General Adjunta de este tribunal para generar automáticamente propuestas de decisión en el área contencioso-administrativo. Su funcionamiento es completamente trazable ya que emplea los documentos que ha agrupado manualmente la fiscalía, de manera que el sistema aplica las reglas de decisión jurídica que esta ha formulado, basándose en árboles de decisión. Los resultados que arroja PROMETEA permiten reseñar un incremento notable en la eficiencia de la fiscalía y sus propuestas de decisiones automatizadas pueden alcanzar un alto grado de calidad y fiabilidad. Se trata de un sistema algorítmico transparente, cuyos resultados pueden ser explicados y justificados conforme a razones jurídicas, garantizando la independencia judicial y la libre toma de decisiones[24].

Por lo tanto, los sistemas de IA en la Administración de Justicia están cumpliendo diversas funciones que pueden ser esencialmente instrumentales y auxiliares al proceso, algunas tareas procesales durante la tramitación que pueden incidir en la determinación de elementos necesarios para la toma de decisión o tareas directamente de decisión sobre el litigio que se haya producido[25].

Incluso desde una perspectiva de género la IA ha contribuido a la configuración de herramientas para proteger a las víctimas de violencia de género. Es el caso del *chatbot* SARA que fue implementado por el Proyecto Regional Infosegura, una iniciativa del Programa de las Naciones Unidas para el Desarrollo (PNUD) en colaboración con la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID). El objeto de este chat es ser aplicado en países del Caribe como República Dominicana o Costa Rica

[23] Roa Avella, M. P., Sanabria- Moyano, J. E. y Dinas-Hurtado, K. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. *Revista Brasileira de Direito Processual Penal, 8* (1), p. 287.

[24] Solar Cayón, J. I. (2025). *Inteligencia artificial jurídica e imperio de la ley* (pp. 320-322). Editorial Tirant Lo Blanch.

[25] Solar Cayón, J. I. (2022)., *op. cit*., p. 386.

para ofrecer información y orientación a mujeres sobre el riesgo de sufrir o haber sufrido violencia, proporcionando también servicios de asesoramiento legal para conocer los derechos que le asisten para saber dónde acudir. Se trata de una herramienta digital capaz de desarrollar aprendizaje automático que permite la comunicación de forma sencilla a través de un chat[26]. Estas tecnologías sirven además para mejorar las ya existentes, como ocurre con Atenpro en España, el Servicio Telefónico de Atención y Protección para víctimas de violencia contra las mujeres. Este es un sistema que utiliza la comunicación telefónica móvil para que las víctimas puedan ponerse en contacto en cualquier momento con un centro de atención personal para atender sus necesidades[27]. La modernización de este servicio se va a producir con la implementación de una aplicación dotada con IA que mejore la prevención y atención a las víctimas permitiendo clasificarlas en función del nivel de riesgo para detectar los casos de mayor urgencia con rapidez[28].

Uno de los principales beneficios que estamos percibiendo que la IA puede tener es facilitar las tareas humanas que con normalidad necesitan de una mayor inversión de tiempo y que gracias a la programación de algoritmos pueden agilizar muchos procesos. En el ámbito de la Administración de Justicia, la sobrecarga que existe en los tribunales podría darse con mayor celeridad y ser más eficiente y accesible a los ciudadanos si se aplican de forma adecuada estas tecnologías. Será crucial diseñar herramientas respetuosas con las garantías procesales y los derechos del justiciable[29].

Igualmente, se podrá conformar un mejor acceso a nuestros derechos sanitarios gracias al avance en tecnología biomédica con herramientas de IA dando viabilidad a ciertos tipos de cirugías que sin la aplicación de estos instrumentos no podrían haberse dado.

Ante los potenciales beneficios que iremos descubriendo que puede aportar a los diferentes ámbitos de la sociedad, no se ha de descuidar el hecho de todos los peligros que puede tener sobre nuestros derechos si no existe un desarrollo y uso ético y regulado.

---

[26] Efeminista (2023, 16 mayo). *`Sara´, la chatbot española que asesora a víctimas de violencia machista en el Caribe.* Disponible en: https://efeminista.com/sara-chatbot-espanola-violencia-machista-caribe/ .

[27] Montesinos García, A. (2024). Inteligencia Artificial en la Justicia con Perspectiva de Género: Amenazas y Oportunidades, *Actualidad Jurídica Iberoamericana*, 21, 583-585.

[28] "Los fondos europeos han consignado 32 millones a modernizar Atenpro y también se ha incrementado la partida presupuestaria estatal". Martín, P. (2023, 23 septiembre). España usará la IA y el "big data" para proteger mejor a las víctimas del machismo. *Diario el Periódico*. Disponible en: https://www.elperiodico.com/es/sociedad/20230923/violencia-genero-machista-inteligencia-artificial-proteccion-big-data-victimas-92338576 .

[29] Solar Cayón, J. I. (2022)., *op. cit.,* p. 385.

2.2. LOS RIESGOS DE LA IA

Al realizar una aproximación a los riesgos, resulta imprescindible considerar una de las grandes problemáticas que generan los sistemas de IA y que sin duda nos concierne, los sesgos algorítmicos y su incidencia sobre los derechos humanos. Se entiende por «*sesgo algorítmico*» aquellos sistemas cuyas predicciones benefician sistemáticamente a un grupo de individuos frente a otro, resultando así injustas o desiguales[30]. Estos sesgos pueden tener lugar en cualquier etapa del proceso de desarrollo del algoritmo y suelen originarse porque se utilizan datos de entrenamiento sesgados por parte de quienes los diseñan. Las personas que diseñan los sistemas de inteligencia artificial tienen sus propias visiones del mundo, prejuicios, valoraciones de los hechos y han adquirido sesgos durante su vida, pudiendo estos filtrarse al definir los criterios de evaluación para estos modelos[31]. De este modo, los sesgos reflejan los prejuicios humanos que se encuentran en los datos históricos que emplean estos sistemas, perpetuando y amplificando las desigualdades existentes si no son diseñados de forma adecuada. El sesgo algorítmico plantea importantes desafíos éticos y sociales pudiendo dar lugar incluso a nuevas formas de discriminación, de modo que es crucial que los desarrolladores de IA tengan en consideración estas implicaciones y trabajen activamente por mitigar los sesgos que pueden surgir. Para abordar de forma adecuada estos desafíos se requiere de un esfuerzo colaborativo en el que juegan un papel clave las normas y principios jurídicos que han de propiciar la garantía del respeto de todas las personas, de su dignidad, libertad y el derecho a ser tratados por los demás conforme al autorreconocimiento de su identidad[32].

El incontrolable avance de la IA está generando numerosos casos de vulneración de derechos humanos especialmente debido a un uso poco ético de los algoritmos por parte de quienes producen el software en cuestión. Los sesgos que se introducen en dichos algoritmos producen y perpetúan grandes discriminaciones de distinto tipo hacia los usuarios afectados. Algunos de los factores que provocan la incorporación de sesgos en estos sistemas se encuentran en el diseño de los mismos debido a los prejuicios y sesgos

[30] Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos? *Nueva sociedad* (294), p. 29.
[31] *Ibid*., p. 35.
[32] López Martínez, F. y García Peña J. H. (2024). IA y sesgos: una visión alternativa expresada desde la ética y el derecho. *Informática y Derecho. Revista Iberoamericana de Derecho Informático, 1* (15), p. 120.

sociales y culturales de los creadores de aplicaciones o a las decisiones que se toman sobre el origen y alcance de los conjuntos de datos con que se entrenan[33].

La privacidad de las personas es sin duda uno de los derechos más desprotegidos en este ámbito. La información privada que depositamos en Internet puede ser utilizada de forma maliciosa por las grandes empresas tecnológicas que almacenan nuestros datos. A través de la IA, la información puede ser sometida a tratamientos masivos en los que, en ocasiones, no se cuenta con el conocimiento y el consentimiento informado del titular del dato. Toda la información que se recaba puede llegar a exponer aspectos privados de la vida de las personas, como pueden ser sus relaciones familiares o su ideología política o religiosa[34].

Del mismo modo, el ejercicio de la libertad de expresión y de pensamiento en los medios o plataformas en línea puede verse afectado por la moderación automatizada de contenidos mediante el uso de IA. Su finalidad es filtrar diversos contenidos que pueden ir de la desnudez, al acoso y al discurso del odio. Sin embargo, se desconoce hasta qué punto las empresas usan la automatización sin intervención humana en casos determinados[35]. Estos sistemas actualmente están limitados por su incapacidad de evaluar el contexto, los usos idiomáticos y los aspectos culturales de los seres humanos. Estas limitaciones tienen como consecuencia que la respuesta automatizada a cualquier contenido o expresión que pueda considerarse inadecuado sea limitarlo o removerlo rápidamente sin tener en cuenta que ello puede afectar de forma considerable al derecho humano a recibir, investigar y difundir información[36]. La libertad de tener una opinión sin injerencia es un derecho consagrado en el artículo 19 de la Declaración Universal de Derechos Humanos, que puede verse fácilmente menoscabado por la afectación de estos modos de personalización y administración del contenido sobre la capacidad de la persona de formar y desarrollar opiniones[37]. Se puede concluir entonces que la remoción automática de contenidos en línea realizada a través de algoritmos predictivos vulnera los estándares internacionales de libertad de expresión tal como la prohibición de censura previa[38].

---

[33] Naciones Unidas. (2018, 29 agosto). *Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión.* (Doc. A/73/348), p. 15.

[34] Mendoza Enríquez, O. A. (2021). El derecho de protección de datos personales en los sistemas de inteligencia artificial. *Revista del Instituto de Ciencias Jurídicas de Puebla, México 15,* (48), 191-192.

[35] Naciones Unidas. (2018, 29 agosto)., *op. cit.*, p. 8.

[36] Grandi, N. M. (2021). Inteligencia Artificial, algoritmos y libertad de expresión. *Universitas* (34), p. 183.

[37] Naciones Unidas. (2018, 29 agosto)., *op. cit.*, p. 11.

[38] Grandi, N. M. (2021)., *op. cit*., p. 191.

El uso de algoritmos en las redes sociales tiene cierta incidencia sobre nuestras decisiones en el ámbito político. Existen en la actualidad algoritmos predictivos que, a través del análisis masivo de datos, adquieren la capacidad de conocer ciertas características personales de los usuarios y estudiar su comportamiento. Un ejemplo demostrativo de ello es el caso Facebook, en el cual la empresa Cambridge Analytica obtuvo datos de los usuarios de esta red social acerca de sus gustos, orientación sexual, política y religiosa y otras cuestiones relativas a datos sensibles de la personalidad. El estudio de esta información a través de la IA sirvió para realizar un perfilamiento y plantear estrategias digitales de coerción electoral al emplearse los resultados durante el proceso electoral presidencial de Estados Unidos en 2016[39]. Esto supuso no solo una violación a la protección de datos personales de los usuarios, sino también una erosión plena del sistema democrático al haber pretendido incidir en la voluntad popular.

Los algoritmos tienen la capacidad de personalizar la información que recibimos en función de nuestro perfil y esto nos lleva a reafirmarnos en nuestros pensamientos, incluso llegando a incidir en nuestras decisiones políticas cuando solo nos muestran aquellas propuestas con las que coincidamos[40]. Esta capacidad de centralizar la información mediante algoritmos opacos que aplican sesgos ideológicos condiciona las opiniones al reducir la variedad y autonomía de las personas[41]. Se refuerzan entonces las ideas similares al generarse grupos cerrados que se convencen mutuamente, dificultando la mejora del sistema democrático en el que se requiere abordar problemas que traspasan los intereses privados y personales. La pérdida de control sobre los datos incide, consecuentemente, en derechos tan importantes como la identidad, la dignidad o el desarrollo de la personalidad[42]. Todo ello supone una vulneración del derecho humano a la privacidad, que ha repercutido sobre las libertades políticas de las personas y que, en consecuencia, erosiona la democracia.

Otro de los aspectos más preocupantes es el factor de los sesgos discriminatorios existentes en la información que se procesa, lo que ocurre, por ejemplo, con algunas

[39] Céspedes Gutiérrez, O. Y. y Carrillo Cruz, Y. A. (2024). Democracia en riesgo: la amenaza a las libertades políticas en la era de la inteligencia artificial. *Justicia, 29* (45), p. 8.
[40] Merino Rus, R. (2021). Democracia y derechos humanos en peligro: una advertencia sobre el impacto de la inteligencia artificial y los algoritmos. *Economistas sin Fronteras, 42*, p. 33.
[41] Sánchez Martínez, M. O. (2022). El impacto de la sociedad digital en los derechos humanos. En J. I. Solar Cayón y M. O. Sánchez Martínez (Dirs.), *El impacto de la inteligencia artificial en la teoría y la práctica jurídica* (p. 129). La Ley.
[42] *Ibid.*, p. 130.

aplicaciones de reconocimiento facial que muestran mayor precisión en la identificación de personas de piel blanca que en la de personas de piel oscura. Se ha demostrado que, en el uso de distintos sistemas de reconocimiento facial, casi el 100% de los hombres blancos eran reconocidos de manera exitosa, mientras que esta tasa disminuía hasta un 35% en el caso de las mujeres racializadas. El origen de esta problemática se halla en un entrenamiento de los sistemas en base a datos sesgados y poco representativos, lo que provoca que la IA identifique patrones en el grupo sobrerrepresentado que les son beneficiosos, repitiendo estos patrones al conocer cada vez mejor las cualidades de dicho grupo[43].

Una forma de reforzar los estereotipos de género a través de la IA se produce con el diseño de los asistentes personales virtuales como Alexa y Siri a los cuales se les asignan nombres y voces femeninas, reforzando así una realidad social en la que tradicionalmente las personas que se encargan del cuidado o que prestan servicios secretariales o de asistencia personal en el sector público o privado son mujeres. Se entiende que la causa principal de estos sesgos de género no es solo el uso de datos ya sesgados, sino también la subrepresentación de las mujeres en el diseño y desarrollo de productos y servicios de IA[44].

En relación con la IA se viene produciendo un fenómeno alarmante especialmente para las mujeres, pudiéndose considerar como una nueva forma de violencia de género. Es el caso de los *deepfakes*, un término que combina *deep learning* (aprendizaje profundo) y *fake* (falso), y que trata de la creación de representaciones realistas de personas en situaciones en las que nunca estuvieron. Esta forma de manipulación de imágenes, audios o vídeos está afectando plenamente a las mujeres en las plataformas digitales al permitir crear vídeos hiperrealistas en los que se coloca a las mujeres en situaciones explícitas, sin su consentimiento ni participación. Estas técnicas comenzaron a ser muy notorias en el año 2017 especialmente en plataformas como Reddit, donde varios usuarios comenzaron a compartir vídeos usando esta tecnología para reemplazar los rostros de mujeres famosas en situaciones sexuales no consentidas. La inmensa mayoría de estos *deepfakes* están relacionados con pornografía no consensuada, lo que demuestra que el uso de estas herramientas de explotación y manipulación del cuerpo femenino intensifica las

[43] Ortiz de Zárate Alcarazo, L. (2023). Sesgos de género en la inteligencia artificial. *Revista de Occidente*, (502), 9-10.

[44] Pérez- Ugena Coromina, M. (2024). Sesgo de género (en IA). *Eunomía. Revista en Cultura de la Legalidad,* (26), 314-315.

dinámicas históricas de poder y control en el espacio digital[45]. Las mujeres son principalmente las víctimas perjudicadas por este tipo de acciones, independientemente de su condición social o económica. Como muestra de ello está la aplicación DeepNude, mediante la que se puede generar desnudos de forma artificial incorporando una foto de un rostro a un cuerpo desnudo obtenido de una base de datos que únicamente contiene imágenes de mujeres[46].

La generación de este tipo de imágenes supone una peligrosa herramienta que atenta directamente contra la imagen, dignidad e integridad de las mujeres y que constituye también una violación de su privacidad.

En el ámbito laboral, la implantación de sistemas de IA puede afectar a los derechos de los trabajadores, incluso desde los procesos de selección para acceder al puesto de trabajo, donde se pueden producir grandes discriminaciones. El caso más ejemplificativo es el que ocurrió con la empresa Amazon que empleó la IA para realizar sus procesos de selección de personal de forma que resultaran más precisos a la hora de puntuar y analizar los currículums. Se descubrió entonces que el sistema puntuaba sistemáticamente de forma más positiva los currículums de hombres que los de mujeres al haber sido entrenado con datos históricos de las personas que habían sido contradas con anterioridad en la empresa, siendo predominantemente hombres[47]. De modo que el sistema encontraba una relación entre los patrones de los hombres, evidentemente sobrerrepresentados, y su idoneidad para desempeñar el puesto vacante[48]. Al no ser capaces de lograr la neutralidad del algoritmo, la empresa abandonó el proyecto ante las características discriminatorias que presentaba. Las implicaciones que tienen estos sesgos son significativas y llevan a amplificar el riesgo de que aumenten las desigualdades existentes en la sociedad por el aprendizaje automático de las máquinas cuando no están sometidas a medidas de supervisión y control.

Igualmente, la automatización en la gestión empresarial está teniendo gran afectación sobre algunos derechos de los trabajadores, como son el derecho a la intimidad, a la protección de datos, a la igualdad y a la salud laboral. Estas intrusiones vienen dadas por

[45] Rosa Castillo, A. y Mont Verdaguer, M. (2024). Inteligencia artificial, deepfakes y nueva explotación del cuerpo femenino: implicaciones éticas y sociales. En T. Áranguez y O. Olariu (Coords.), *Los derechos de las mujeres en la sociedad digital* (p. 155). Dykinson.
[46] Montesinos García, A. (2024)., *op. cit*., p. 579.
[47] Pérez- Ugena Coromina, M. (2024)., *op. cit*., p. 320.
[48] Ortiz de Zárate Alcarazo, L. (2023)., *op. cit*., p. 11.

una vigilancia invasiva y un uso intensivo de datos de los trabajadores, ya que para implementar adecuadamente algoritmos en los procesos se exige utilizar *big data*. Estos aspectos intrusivos han sido expuestos en casos como el del algoritmo *FranK* utilizado por la plataforma Deliveroo para la organización de las reservas de trabajo de los *riders*. Ha sido un caso pionero al haberse declarado por primera vez en Europa, gracias a la Sentencia del Tribunal de Bolonia, que ciertos algoritmos pueden ser discriminatorios con las personas consideradas vulnerables[49]. La plataforma utilizaba un modelo organizativo basado en un ranking de reputación digital, privilegiando a aquellos trabajadores que contasen con mayor disponibilidad para efectuar la actividad de reparto y penalizando a aquellos que cancelaban su sesión, sin tener en cuenta la justificación de la situación. Esta preferencia situaba de forma prevalente en el ranking a los trabajadores que contaban supuestamente con mayor disponibilidad, de forma que se excluía de ciclo productivo a los repartidores que no aseguraban estar disponibles, dejando que fuesen los últimos en elegir franja horaria. Se producía entonces una discriminación de aquellos trabajadores que cancelaban el servicio, al ignorar si se trataba de cuestiones de salud, conciliación familiar o acciones sindicales como la huelga[50]. Esta automatización de las decisiones empresariales no es neutral y su actuación discriminatoria incide directamente sobre el ejercicio de los derechos laborales de estos trabajadores.

En el ámbito de la justicia también se han llegado a producir vulneraciones de derechos fundamentales de las personas implicadas en litigios, especialmente por la existencia de sesgos discriminatorios en las decisiones adoptadas mediante sistemas de IA. La principal problemática que plantean los algoritmos de *machine learning* es la opacidad en la toma de decisiones, que puede generar una lesión en el derecho de defensa de las partes de un proceso. Este gran obstáculo supone la quiebra de un derecho fundamental que sitúa al perjudicado en una situación de indefensión y que, en definitiva, produce un menoscabo a la tutela judicial efectiva[51]. Por otra parte, automatizar las decisiones predictivas tiene como consecuencia la estandarización de pautas y argumentos jurídicos, condicionando así el futuro de los sujetos que se vean sometidos a estas técnicas por haberse aplicado

[49] Fernández Sánchez, S. (2021). Frank, el algoritmo consciente de Deliveroo. Comentario a la Sentencia del Tribunal de Bolonia 2949/2020, de 31 de diciembre. *Revista de Trabajo y Seguridad Social, CEF,* (457), p. 182.
[50] Fernández Sánchez, S. (2021)., *op. cit.*, p. 183.
[51] San Miguel Caso, C. (2021). La aplicación de la Inteligencia Artificial en el proceso: ¿un nuevo reto para las garantías procesales? *Ius et Scientia, 7* (1), p. 295.

una interpretación basada en datos pasados, que generalmente conllevará resultados erráticos[52]. Esta circunstancia afecta a la imparcialidad judicial al producir resultados subjetivos y puede suponer una quiebra de la presunción de inocencia de los sujetos[53].

Se ha determinado que algunos softwares empleados para predecir futuros delincuentes demuestran una predisposición contra las comunidades afrodescendientes, asignándoles mayor puntuación en cuanto a posible riesgo frente a los acusados blancos. En relación a esta cuestión cabe traer a colación el caso anteriormente tratado acerca del algoritmo COMPAS en el caso State v. Loomis. El acusado fue identificado por este sistema como individuo de alto riesgo para la comunidad, con alto grado de reincidencia, de modo que se le negó la libertad condicional. La defensa entendió que esta decisión vulneraba el derecho al debido proceso a causa de la opacidad del algoritmo que impedía conocer razonablemente la argumentación de la decisión judicial. El recurrente afirmó además que se desconoció el derecho a la individualización de su caso porque los datos que manejaba el algoritmo se basaban en el riesgo de incidencia en grupos; y sostuvo también que se había creado un sesgo por razón del género y la raza[54]. En este caso se podía entender que el error no se produjo por sesgos racistas de los desarrolladores del software, sino porque COMPAS intenta encontrar patrones basados en ejemplos que se dan en la sociedad[55]. Por lo tanto, estas situaciones revelan que existe una discriminación directa e indirecta en la aplicación de la IA en el proceso.

Pese a que algunas tareas puedan ser automatizadas, incluida la toma de decisiones judiciales, ello no implica necesariamente la posibilidad de prescindir del juez en su realización[56]. El empleo de estos sistemas en determinadas actividades podrá restringirse dada su imposibilidad de interpretar o explicar los resultados que adoptan, lo que afecta directamente a la motivación de las decisiones que toman[57]. La labor de los sistemas de IA en el ámbito judicial puede ser de apoyo y asistencia al juez en sus decisiones, ya que colabora a una mayor eficiencia y a agilizar el sistema judicial. No obstante, en ningún

---

[52] *Ibid.*, p. 292.
[53] *Ibid.*, p. 95.
[54] Duitama Pulido, E. A. (2022). Sesgos algorítmicos e inteligencia artificial en el poder judicial. En A. Padilla-Muñoz (Ed.), *El derecho como laboratorio de saberes: meditaciones sobre epistemología* (p. 59). Editorial Universidad del Rosario.
[55] Casacuberta, D. y Guersenzvaig, A. (2019). Using Dreyfus´ legacy to understand justice in algorithm-based processes. *AI y Society, 34*(2), p. 314.
[56] Solar Cayón J. I. (2022). ¿Jueces-robot? Bases para una reflexión realista sobre la aplicación de la inteligencia artificial en la administración de justicia. En J. I Solar Cayón y M. O Sánchez Martínez (Dirs.), *El impacto de la inteligencia artificial en la teoría y la práctica jurídica* (p. 257). La Ley.
[57] *Ibid.*, p. 256.

caso la potestad jurisdiccional podrá desempeñarse por los denominados jueces-robots[58]. Todos los obstáculos que presentan revelan que se han de primar las garantías y derechos de las partes ante el riesgo de que se sigan produciendo vulneraciones de los derechos procesales tales como el debido proceso, el derecho de defensa, la imparcialidad judicial, e incluso el derecho a la igualdad y no discriminación.

Los riesgos que presenta la IA sobre los derechos humanos, especialmente debido a la problemática de los sesgos algorítmicos, plantean importantes desafíos éticos y sociales que han de abordarse para evitar perpetuar las desigualdades existentes en la sociedad y crear nuevas formas de discriminación. Ello supone que entre los desarrolladores de estos sistemas se encuentren personas diversas con distintas experiencias culturales, de modo que se pueda garantizar que estas tecnologías sean justas, lo que implica también un seguimiento y revisión sobre cómo afectan a nuestros derechos. La clave para que estas prácticas sean efectivas es la creación de marcos regulatorios que regulen su uso y desarrollo.

3. LOS PRIMEROS INTENTOS DE REGULACIÓN INTERNACIONAL DE LA IA

En un inicio, el intento de regulación de la IA ha venido marcado por una serie de propuestas de *soft law* por las que algunas organizaciones han establecido principios y recomendaciones dirigidos a los Estados con objeto de marcar unas pautas comunes. Ante el desconocimiento de todas las implicaciones que esta tecnología pudiera tener en nuestras vidas, todas estas normas prevén la protección de derechos a través de indicaciones concretas como son la transparencia y la ética tanto en el desarrollo como en el uso de sistemas de inteligencia artificial.

De entre los marcos normativos que se han ido adoptando, nos ocuparemos a continuación de los que claramente han tenido una mayor influencia en la posterior elaboración de normativa internacional jurídicamente vinculante, como son los Principios de la OCDE y la Recomendación ética de la UNESCO sobre inteligencia artificial.

[58] San Miguel Caso, C. (2023). Inteligencia artificial y algoritmos: la controvertida evolución de la tutela judicial efectiva en el proceso penal. *Estudios Penales y Criminológicos, 44*, p. 8.

3.1. LOS PRINCIPIOS DE LA OCDE SOBRE INTELIGENCIA ARTIFICIAL

En primer lugar, desde la Organización para la Cooperación y el Desarrollo Económico (OCDE) fue adoptada el 22 de mayo de 2019 la Recomendación del Consejo sobre Inteligencia Artificial, estableciendo en ella la necesidad de regularla con el fin de lograr sistemas seguros, fiables e imparciales[59]. Este instrumento agrupa una serie de principios mínimos que deben cumplir los sistemas de IA para ser gestionados de manera responsable. Estos estándares no son legalmente vinculantes, pero se insta a los adherentes a que los promuevan e implementen a la hora de diseñar políticas públicas y normativa. Su objetivo principal es desarrollar un enfoque de la IA centrado en el ser humano, que promueva la innovación y respete los derechos humanos, la diversidad y los valores democráticos[60]. Se establecen así normas prácticas y lo suficientemente flexibles como para logar adaptarse a los cambios tecnológicos que puedan darse a lo largo del tiempo.

En la primera sección de la Recomendación se exponen una serie de principios de administración responsable en aras de una IA fiable, para cuya aplicación se recomienda a los Miembros y no Miembros de la Organización a adherirse a la Recomendación. Mientras que los dos primeros principios pretenden alcanzar la inclusividad y proponen un enfoque centrado en el ser humano, los tres últimos aportan una perspectiva interseccional sobre la ética de los datos y la seguridad física en el uso de la IA[61]. Los cinco principios que se promueven incluyen los siguientes elementos[62]:

1) Crecimiento inclusivo, desarrollo sostenible y bienestar.

A través de este principio se establece una orientación de la IA hacia la prosperidad y resultados beneficiosos tanto para las personas, como para el planeta. Se reconoce también el hecho de que la IA puede perpetuar los sesgos existentes en la sociedad y tener un impacto desigual en las personas vulnerables, riesgo que debe abordarse utilizando esta herramienta. De este modo, será crucial la gestión responsable durante todo el ciclo de vida del sistema de IA mediante la colaboración de las múltiples partes interesadas

[59] Uriol, L. M. (2024). Análisis de las recomendaciones de ONU y OCDE sobre la regulación de la Inteligencia Artificial. El estado de situación en Argentina. *Aequitas, 16* (16), p. 9.
[60] *Ibid*., p. 9.
[61] Nguyen A., Ngo H. N., Hong, Y., Dang, B. y Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies, 28* (4), p. 4226.
[62] Toda la información acerca de los principios de administración responsable ha sido extraída de la Sección 1 del instrumento de: Organización para la Cooperación y el Desarrollo Económico. (2019, 22 mayo). *Recomendación del Consejo sobre Inteligencia Artificial*. (Doc. OECD/LEGAL/0449), pp. 8-9.

para obtener resultados beneficiosos, dado que un diálogo público significativo puede mejorar la confianza en la IA y su comprensión.

2) Respetar el Estado de Derecho, los derechos humanos y los valores democráticos, incluidas la equidad y la privacidad.

Este principio incide en la importancia del desarrollo de una IA acorde con valores centrados en el ser humano, incluyendo la igualdad, la dignidad, la privacidad y la protección de datos, el Estado de Derecho y la justicia social. Esto supone que los sistemas de IA se diseñen con las salvaguardas adecuadas a través de la intervención y supervisión humanas para garantizar que su comportamiento protege y promueve los derechos humanos ante el riesgo de que estos puedan ser vulnerados deliberada o accidentalmente. En este sentido juegan un importante papel las medidas como la evaluación de impacto sobre los derechos humanos, los códigos de conducta ética o las certificaciones de calidad.

3) Transparencia y explicabilidad

La transparencia supone un compromiso para los actores de IA de divulgar de forma responsable la información sobre estos sistemas, fomentando una comprensión general de los mismos y que permita a aquellos afectados negativamente cuestionar los resultados que les han sido proporcionados. Ha de ser factible para las personas la comprensión del desarrollo y funcionamiento de los sistemas de IA para permitir una toma de decisiones más informada. La información habrá de ser adecuada para la situación específica y tendrá que ajustarse al estado actual de la tecnología.

4) Robustez, seguridad y protección

Estos principios deben caracterizar a los sistemas de IA durante todo su ciclo de vida para garantizar su adecuado funcionamiento y no plantear riesgos irrazonables de seguridad. Esto va a suponer que los actores empleen un enfoque de gestión de riesgos para protegerse contra el uso indebido previsible utilizando mecanismos que garanticen la invalidación o corrección de los sistemas de IA en caso de amenaza de provocar daños indeseados.

5) Rendición de cuentas

Los actores de la IA han de ser responsables del correcto funcionamiento de los sistemas y respetar los principios establecidos por la OCDE en esta materia. De acuerdo con sus

funciones han de garantizar la trazabilidad de las decisiones tomadas durante todo el proceso de desarrollo para permitir el análisis de los resultados obtenidos y asegurar que las respuestas sean adecuadas al contexto específico y coherentes con el estado de la tecnología. La rendición de cuentas supone acciones como la aportación de documentación sobre decisiones clave a lo largo del ciclo de vida del sistema o permitir auditorías debidamente justificadas.

Por otro lado, se establecen también en la segunda sección cinco recomendaciones a los Miembros y no Miembros que se adhieran a la Recomendación para que las apliquen en sus políticas nacionales y de cooperación internacional[63]:

1. Invertir en investigación y desarrollo de la IA.

Se insta a los gobiernos a invertir en investigación y desarrollo de ciencia abierta para impulsar y dar forma a la innovación en IA fiable. Esta consideración supone tener en cuenta herramientas que respeten la privacidad y la protección de datos para que el entorno investigador se halle libre de sesgos perjudiciales. Dado su amplia implicación en múltiples facetas de la vida, la recomendación exige además invertir en investigación interdisciplinaria sobre las implicaciones sociales, legales y éticas de la IA.

2. Fomentar un ecosistema inclusivo que propicie la IA

Igualmente, los gobiernos deberían fomentar el desarrollo y el acceso a un ecosistema digital inclusivo proporcionando la infraestructura y las tecnologías digitales necesarias para la IA, promoviendo además mecanismos que garanticen una difusión segura, justa, legal y ética de los datos. Esto debe hacerse atendiendo a los riesgos que pueden existir en un intercambio de datos para las personas, organizaciones y países, incluyendo violaciones de la confidencialidad y la privacidad o riesgos para la seguridad digital.

3. Configurar un entorno político y de gobernanza interoperable propicio para la IA

Los gobiernos tienen el papel de promover un entorno controlado de políticas de IA empleando para tal fin la experimentación. Se destaca la importancia de adaptar sus marcos normativos y regulatorios para que sean lo suficientemente flexibles como para

[63] Toda la información acerca de las recomendaciones de aplicación a las políticas nacionales y de cooperación internacional ha sido extraída de la Sección 2 del instrumento de: Organización para la Cooperación y el Desarrollo Económico. (2019, 22 mayo)., *op. cit.*, pp. 9-10.

seguir el ritmo de los avances y promover la innovación. Se ha de tener en cuenta, sin embargo, que ese entorno sea seguro y ofrezca seguridad jurídica, de modo que debe haber mecanismos de supervisión y evaluación para complementar estos marcos, existiendo un margen para alentar a los actores de la IA a desarrollar mecanismos de autorregulación.

4. Reforzar las capacidades humanas y prepararse para la transformación del mercado laboral

Debe darse una colaboración entre los grupos de interés y los gobiernos con objeto de una preparación ante la transformación del mundo laboral y de la sociedad. Se ha de garantizar especialmente, a través del diálogo social y la negociación colectiva, una transición justa para los trabajadores permitiendo que accedan a nuevas oportunidades en el mercado laboral. Los cambios que ha provocado la IA en este contexto pueden requerir la adaptación o adopción de normas laborales y acuerdos para abordar los posibles desafíos a la igualdad, la diversidad y la equidad, de modo que se garanticen lugares de trabajo fiables, seguros y productivos.

5. Cooperación internacional para una IA fiable

La cooperación a nivel internacional entre los gobiernos y los grupos de interés es clave para promover los principios que propone la OCDE y su trabajo conjunto en foros internacionales debe servir para difundir los conocimientos sobre la IA. Se han de desarrollar normas técnicas internacionales consensuadas por los diferentes actores implicados para conseguir una IA interoperable y fiable, junto con la utilización de parámetros como la equidad y el progreso de los objetivos sociales para evaluar el desempeño de los sistemas de IA.

Hasta la fecha, además de los 38 Miembros de la OCDE, 9 países no miembros se han adherido a la Recomendación. Los Informes sobre la implementación de los principios de la OCDE arrojan grandes resultados acerca de cómo los Adherentes los han ido implementando en sus políticas a través de diversas iniciativas, lo que demuestra que

estos principios continúan teniendo gran relevancia a la hora de guiar el desarrollo de una IA responsable acorde con los derechos humanos y los valores democráticos[64].

El 8 de noviembre de 2023, el Consejo de la OCDE se reunió para realizar ciertas revisiones de la definición de sistemas de IA incluida en la Recomendación, modificando el concepto para, entre otras actualizaciones, clarificar que los objetivos de los sistemas de IA pueden ser explícitos o implícitos, reflejar el hecho de que algunos de estos sistemas pueden seguir evolucionando después de su diseño e implementación y recalcar el rol de la información que pueden proporcionar los humanos o las máquinas[65].

Del mismo modo, siguiendo las conclusiones expuestas en el Informe al Consejo del año 2024, la Recomendación fue revisada en la reunión del Consejo a nivel ministerial celebrada en mayo de ese mismo año, con objeto de facilitar su aplicación cinco años después de su adopción. En dicha revisión se incluyeron actualizaciones en las que se pretende especialmente subrayar la necesidad de que las jurisdicciones cooperen para promover entornos de gobernanza y políticas interoperables para la IA, frente al aumento de las iniciativas de políticas de IA en todo el mundo. Se busca también aclarar la información que los actores de la IA han de proporcionar sobre los sistemas que diseñan para garantizar la transparencia y la divulgación responsable y abordar con mayor relevancia las cuestiones de seguridad. Ello implica que, si se observan riesgos de que lo sistemas de IA causen daños indebidos, puedan ser anulados o reparados de forma segura mediante la intervención humana[66].

La Recomendación de la OCDE pretende fomentar la innovación y la confianza en la IA y vela porque estos sistemas se guíen por la transparencia y la ejecución responsable. La gran crítica de este instrumento se centra en que sus principios no son vinculantes y que no aborda algunas preocupaciones más amplias en relación con el impacto que la IA pueda tener en la sociedad. Sin embargo, se trata de una propuesta que sienta las bases para promover el bienestar general en beneficio de las personas y el medio ambiente y un crecimiento equitativo en la creación de sistemas de IA[67].

---

[64] Organización para la Cooperación y el Desarrollo Económico. (2024, 24 abril). *Informe sobre la implementación de la Recomendación de la OCDE sobre Inteligencia Artificial*. (Doc. C/MIN (2024)17), p. 3.

[65] Organización para la Cooperación y el Desarrollo Económico. (2019, 22 mayo)., *op. cit*., p. 4.

[66] Organización para la Cooperación y el Desarrollo Económico. (2024, 24 abril)., *op. cit.*, p. 33.

[67] Morandín-Ahuerma, F. (2023). Principios normativos para una ética de la Inteligencia Artificial. *Concytep,* 1, p. 100.

3.2. RECOMENDACIÓN DE LA UNESCO SOBRE LA ÉTICA DE LA IA

No cabe duda de que, habida cuenta de la implicación que tiene la IA sobre los Derechos Humanos, se ha de mantener un uso ético de esta herramienta, tanto en su configuración y desarrollo como en su propia utilización. De este modo, la continua evolución de la IA y su impacto en los diferentes sectores hacen necesario el establecimiento de mecanismos de supervisión para asegurar el cumplimiento de los principios éticos, promoviendo así un entorno donde la tecnología contribuya al bienestar colectivo y al respeto por los derechos humanos[68]. Indudablemente, las cuestiones éticas y sus correspondientes regulaciones han de establecerse como pilares fundamentales a fin de garantizar que su uso se produzca en beneficio del bien común y que no se convierta, por tanto, en una amenaza para la humanidad. Este marco ético involucra a todos los actores de la IA, incluidos los usuarios y la sociedad civil en su conjunto, garantizando así que sea una herramienta poderosa para el progreso social y se promueva a través de ella no solo la innovación tecnológica, sino también la justicia social[69].

Dado que estos sistemas se basan normalmente en grandes cantidades de datos, será esencial la elaboración de políticas que incorporen mecanismos para prevenir la discriminación dada por sesgos cognitivos que puedan incluso vulnerar la seguridad y privacidad de los usuarios[70].

Por lo tanto, será necesario que los distintos actores de la IA se coordinen para alcanzar un consenso acerca de los principios éticos que deben guiar su desarrollo y uso, que servirá para construir una sociedad más segura, justa y equitativa[71].

En este contexto parece situarse la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO) al adoptar en el año 2021 la Recomendación sobre la Ética de la Inteligencia Artificial. Se trata de la primera norma mundial sobre la ética de la IA y tiene aplicación para los 194 Estados miembros, en su calidad de autoridades responsables de la elaboración de marcos jurídicos y reguladores durante el ciclo de vida de los sistemas de IA. En esta Recomendación se propone una evaluación del impacto ético de estos sistemas, de modo que se proporciona una

[68] Guzmán Huayamave, K. V. (2024). IA y ética: regulaciones y políticas para una sociedad digital. En *II Congreso Internacional Multidisciplinario. Nuevas Perspectivas en Ciencia y Tecnología: El Papel Transformador de la Inteligencia Artificial* (p. 104).
[69] *Ibid.,* pp. 105-106.
[70] *Ibid*., pp. 108.
[71] *Ibid.*, pp. 109.

orientación acerca de cómo actuar a todos los actores de la IA, tanto del sector público como privado[72].

Conjuntamente se trata de un marco integrador de valores, principios y acciones que vienen impulsados por un conjunto de fines y objetivos y que cuenta con once ámbitos de acción política. En este instrumento se reconoce que las cuestiones éticas que rodean a la IA requieren de una cooperación entre las partes interesadas a nivel internacional, regional y nacional, posibilitando que en base a los fines y valores que se establecen en la Recomendación, logren un diálogo mundial y multidisciplinario que lleve a compartir una responsabilidad sobre dichas cuestiones[73].

Entre sus diversos objetivos cabe destacar que se pretende proteger, promover y respetar los derechos humanos y las libertades fundamentales, la dignidad humana y la igualdad, incluida la igualdad de género, en todas las etapas del ciclo de vida de los sistemas de IA. Estos fines toman forma en los cuatro valores y diez principios que establece la Recomendación con objeto de que sean respetados por todos los actores y, de resultar conveniente, se modifiquen las leyes, reglamentos y directrices empresariales existentes, en consonancia con lo previsto en la Carta de las Naciones Unidas, así como en otros instrumentos de derecho internacional y en las obligaciones de los Estados miembros en materia de derechos humanos[74].

Entre los valores que establece la citada Recomendación nos interesa, en primer lugar, el relativo al respeto, protección y promoción de los derechos humanos, las libertades fundamentales y la dignidad humana. Se reconoce que ningún ser humano ni comunidad humana debería sufrir daños o sometimiento, ya sean de carácter físico, económico, social, político, cultural o mental, durante ninguna etapa del ciclo de vida de los sistemas de IA. Por tanto, su función debería ser mejorar la calidad de vida de los seres humanos y deberá entonces asegurarse que en las interacciones de personas en situación de vulnerabilidad con los sistemas de IA no se menoscabe nunca su dignidad y no se produzca ninguna violación o abuso de los derechos humanos y libertades fundamentales[75].

---

[72] UNESCO (2021, 23 noviembre). *Recomendación sobre la ética de la inteligencia artificial*. (Doc. SHS/BIO/REC-AIETHICS-2021).

[73] Reyes Vásquez, P. A (2023). Ética de la Inteligencia Artificial. Recomendación de la UNESCO, noviembre 2021. *Compendium, 26* (50), p. 2.

[74] UNESCO (2021, 23 noviembre)., *op. cit.,* p. 18.

[75] *Ibid*., p. 18.

En una aplicación práctica de los valores y principios que prevé la Recomendación podrían surgir tensiones, de manera que se recalca que, en cualquier caso, de producirse una posible limitación de los derechos humanos, esta ha de tener una base jurídica y seguir los principios de razonabilidad, necesidad y proporcionalidad, además de ser conforme a las obligaciones de los Estados con arreglo al derecho internacional[76].

El otro valor de interés para nuestro trabajo es el de garantizar la diversidad y la inclusión a lo largo del ciclo de vida de los sistemas de IA, promoviendo la participación activa de todas las personas o grupos, con independencia de su raza, color, género, edad, religión, condición económica o social o cualquier otro motivo. Se invita a la cooperación internacional para tratar de paliar la falta de infraestructura, educación y competencias tecnológicas necesarias que afecta a algunas comunidades[77].

Los otros dos valores que se incluyen son la prosperidad del medio ambiente y los ecosistemas y vivir en sociedades pacíficas, justas e interconectadas. Esto exige que se promuevan la paz, la inclusión y la interconexión durante el ciclo de vida de los sistemas de IA, así como reconocer y proteger la prosperidad del medio ambiente y los ecosistemas para que todos los seres vivos puedan beneficiarse de los avances de la misma[78].

Por otro lado, los principios que comprende la Recomendación son los siguientes: proporcionalidad e inocuidad, seguridad y protección, equidad y no discriminación, sostenibilidad, derecho a la intimidad y protección de datos, supervisión y decisión humanas, transparencia y explicabilidad, responsabilidad y rendición de cuentas, sensibilización y educación, y gobernanza y colaboración adaptativas y de múltiples partes interesadas.

Cabe destacar el principio de equidad y no discriminación, por el cual se pretende que los actores de la IA intenten reducir al mínimo y evitar reforzar resultados discriminatorios o sesgados durante estos procesos, facilitando recursos efectivos contra la discriminación y la determinación algorítmica sesgada con el fin de garantizar la equidad de dichos sistemas[79]. Este principio conecta bastante con el de transparencia y explicabilidad, por el que se exige que se informe completamente a las personas en caso de que una decisión se base en algoritmos de IA, especialmente si afecta a su seguridad o sus derechos humanos. Además, dicha decisión habría de estar motivada y se habría de posibilitar la

[76] Reyes Vásquez, P. A (2023)., *op. cit.,* p. 2.
[77] UNESCO (2021, 23 noviembre)., *op. cit*., p. 19.
[78] *Ibid.*, pp. 19-20.
[79] *Ibid.,* p. 21.

presentación de alegaciones a un miembro del personal del sector correspondiente si la persona ha entendido vulnerados sus derechos o libertades[80].

Asimismo, la Recomendación prevé una serie de acciones políticas en once distintos ámbitos de actuación para poner en práctica los valores y principios enunciados: evaluación del impacto ético, gobernanza y administración éticas, política de datos, desarrollo y cooperación internacional, medio ambiente y ecosistemas, género, cultura, educación e investigación, comunicación e información, economía y trabajo, salud y bienestar social.

Entre ellos el más relevante en relación con el tema que se pretende abordar en el presente trabajo es la evaluación del impacto ético. Se recomienda a los Estados miembros el establecimiento de medidas de prevención, atenuación y seguimiento de los riesgos de los sistemas de IA con miras a revelar las posibles repercusiones en los derechos humanos y las libertades fundamentales, así como sus consecuencias éticas y sociales. Estos mecanismos de gobernanza deben ser inclusivos, transparentes, multidisciplinarios y multilaterales[81]. De tal modo que, en el marco de la evaluación del impacto ético, habrían de ser ampliamente probados, antes de sacarlos al mercado, todos aquellos sistemas que puedan ser considerados riesgos potenciales para los derechos humanos. Esto demuestra que se considera de gran importancia el desarrollo de mecanismos de diligencia debida, supervisión y vigilancia en todas las etapas del ciclo de vida de los sistemas de IA, en consonancia con las obligaciones de los Estados miembros en materia de derechos humanos, que han de estar necesariamente ligadas a los aspectos éticos de las evaluaciones de estos sistemas[82].

En definitiva, la Recomendación sobre la ética de la inteligencia artificial supone un instrumento de amplio alcance para los diversos agentes sociales y políticos, tanto gobiernos como empresas y organizaciones civiles. Constituye una guía a los países para responder a los impactos actuales que tienen las tecnologías que emplean inteligencia artificial en los distintos ámbitos. En particular, a través de los valores y principios que toma como base, se pretende atender a las repercusiones éticas de estos sistemas.

[80] *Ibid.*, p. 22.
[81] Vázquez Pita, E. (2021). La UNESCO y la gobernanza de la inteligencia artificial en un mundo globalizado. La necesidad de una nueva arquitectura legal. *Anuario de la Facultad de Derecho. Universidad de Extremadura 37*, p. 298.
[82] UNESCO (2021, 23 noviembre)., *op. cit.*, p. 26.

No obstante, sigue tratándose de un texto que carece de fuerza jurídica vinculante. De este modo, las propuestas de *soft law* no resultan suficientes para paliar los efectos que han tenido la falta de transparencia y ética en el comportamiento de los desarrolladores al configurar estos sistemas.

4. MARCO NORMATIVO INTERNACIONAL SOBRE IA

Ante el aumento de producción de sistemas de IA a nivel mundial y su gran impacto en la vida de las personas, así como el riesgo de vulnerar sus derechos, se ha hecho necesaria en los últimos años pasar de la mera proclamación de principios a la elaboración de un marco normativo jurídicamente vinculante que regule el uso de estas tecnologías. La regulación jurídica no puede quedar atrás ante el veloz desarrollo tecnológico que se está produciendo. En particular, parece imprescindible que los agentes nacionales e internacionales deben llegar a acuerdos para que los derechos humanos queden amparados en este contexto.

A lo largo de nuestra investigación hemos observado la toma de conciencia creciente por parte de los Estados de la necesidad de regular la IA, lo que se ha traducido en los últimos años en toda una serie de propuestas normativas, algunas de las cuales han trascendido el ámbito puramente nacional. Nos interesa ahora detenernos en el examen de las principales iniciativas observadas en el escenario internacional con el fin de verificar si identifican la protección de derechos humanos entre sus objetivos y valorar en qué medida contribuyen realmente a su logro.

4.1. INTENTOS DE REGULACIÓN DE LA IA EN EL ÁMBITO NACIONAL E INTERNACIONAL

El mayor número de propuestas de regulación de la IA existentes en la escena internacional son de naturaleza nacional, aunque no exclusivamente. Como cabía esperar, las grandes potencias tecnológicas en este ámbito han sido las primeras en dar un paso al frente en la medida en que compiten por obtener una mayor ventaja en el campo de la IA. Sin embargo, presentan grandes diferencias en cuanto al aspecto regulatorio de esta cuestión, especialmente debido a que sus intereses son distintos y generalmente se priman

aspectos como la innovación, el mercado y la seguridad frente a la protección de derechos humanos. Con objeto de considerar las diversas propuestas regulatorias y teniendo en cuenta la relevancia en el mercado tecnológico, conviene mencionar los desarrollos normativos en Estados Unidos y China, pero también nos han llamado la atención algunas iniciativas en Iberoamérica, en las que la ética y los derechos humanos han ocupado un importante papel; y en Europa, por las razones añadidas de afectarnos más directamente y tener un carácter colectivo, incluso con vocación de universalidad.

Comenzando por Estados Unidos, el Gobierno de Biden anunció el 30 de octubre de 2023 la Orden Ejecutiva sobre el desarrollo y la utilización segura y fiable de la inteligencia artificial, imponiendo medidas de transparencia a las empresas tecnológicas que operen en el país. A través de ella se reconoció la existencia de violaciones de derechos civiles en relación con el uso de la IA, por lo que se instó al fiscal general a tomar medidas para lograr una reducción de las vulneraciones de derechos como la discriminación algorítmica. Con ello se buscaba que las grandes empresas que desarrollan sistemas de IA tuvieran la obligación de compartir la información disponible con el gobierno. Algunas de las acciones específicas que se pretendían estaban enfocadas en conseguir que los líderes de las agencias federales desarrollasen estrategias sobre el uso responsable de la IA y que presentasen informes al presidente sobre su impacto y sobre las mejores prácticas para su uso[83]. Con esta Orden Ejecutiva se pretendía enfatizar la necesidad de regular la IA de alto riesgo y se reconocía la vinculación que esta tiene con la privacidad, además de prometer apoyo federal para el desarrollo de tecnologías que mejoren estos aspectos[84].

Sin embargo, con la toma de posesión del presidente Trump, se han revertido numerosas políticas de la Administración de Biden, entre las cuales se encuentra la mencionada acerca de la inteligencia artificial. De modo que el intento por reducir los riesgos que la IA puede tener sobre los consumidores ha sido rápidamente rechazado por parte del nuevo Gobierno, el que apoya un desarrollo de la IA basado en la libertad de expresión y el libre desarrollo tecnológico. En la Cumbre de Acción sobre Inteligencia Artificial celebrada

[83] Lluís Iribarren, A. (2024). *Análisis comparativo de la regulación de la inteligencia artificial en la Unión Europea y Estados Unidos* (p. 22). [Trabajo Fin de Grado, Universidad Autónoma de Barcelona]. Dipòsit digital de documents de la UAB. https://ddd.uab.cat/record/303428 .

[84] Williams, K. (2023, 7 noviembre). *Summary: What Does Biden´s Executive Order on Artificial Intelligence Actually Say?* Electronic Privacy Information Center. Disponible en: https://epic.org/summary-what-does-bidens-executive-order-on-artificial-intelligence-actually-say/ .

en París en febrero de 2025, en la que se abogaba por una IA inclusiva y sostenible, el Gobierno estadounidense rechazó sumarse a la declaración final del encuentro argumentando que "una regulación excesiva del sector de la IA podría matar a una industria transformadora", y recalcando que "Estados Unidos es el líder en inteligencia artificial y queremos que siga siendo así"[85]. En definitiva, se puede observar que el enfoque que adopta ahora Estados Unidos tiene por objeto no frenar la innovación, ya que considera peligroso e innecesario imponer un control gubernamental sobre el desarrollo de esta tecnología. Su propósito es mantener una posición competitiva en el mercado tecnológico y ello supone no imponer obligaciones a los agentes implicados en el desarrollo de IA, con el riesgo que ello implica para los derechos de todos los usuarios.

Otra potencia mundial en el ámbito de la inteligencia artificial es China, la que ha optado por un enfoque más centralizado y proactivo en la regulación de esta tecnología priorizando la seguridad y la estabilidad social. En junio de 2023 publicó el Reglamento de Síntesis Profunda consistente en una serie de normas que exigen a los proveedores de servicios de IA el registro de sus algoritmos ante el Gobierno en el caso de que sus servicios puedan influir en la opinión pública, lo que implica revelar los contenidos generados de manera artificial y los datos empleados para el entreno de modelos lingüísticos de gran tamaño, además de por supuesto revisar la seguridad de estos sistemas. Un mes más tarde se publicó el Reglamento de IA Generativa con objeto de regular «modelos y tecnologías con capacidad de generar textos, imágenes, sonidos, vídeos y otros contenidos». Al ser los usuarios de sistemas de IA quienes pueden ver afectados sus derechos, las obligaciones de control y protección deben recaer sobre los desarrolladores. Estos tendrán que informar a los usuarios de las circunstancias en las que se utilizan los algoritmos de recomendación en sus servicios, haciendo además responsables a las plataformas de cualquier contenido ilegal o indeseable que sea recomendado a través de dichos algoritmos[86].

El escaso intento de regulación de la IA en China no busca tampoco restringir los avances en este campo, sino un control estatal de la tecnología, estableciendo a los proveedores

[85] Milmo, D. (2025, 11 febrero). EEUU acusa a Europa de "regulación excesiva" y se niega a firmar la declaración de la cumbre de París sobre IA "inclusiva". *ElDiario.es*. Disponible en: https://www.eldiario.es/internacional/eeuu-rechaza-firmar-declaracion-cumbre-paris-ia-inclusiva-hablar-excesiva-regulacion-europa_1_12044385.html .

[86] Pérez-Ugena, M. (2024). Análisis comparado de los distintos enfoques regulatorios de la inteligencia artificial en la Unión Europea, EE. UU, China e Iberoamérica. *Anuario Iberoamericano de Justicia Constitucional*, *28* (1), 146-148.

obligaciones especialmente centradas en informar sobre aquellos contenidos que hayan sido generados artificialmente. Se impone además la obligación de moderación de contenidos y la revisión del contenido que se distribuye a los usuarios. En definitiva, la supervisión gubernamental es extensa en el marco regulatorio chino y su principal objetivo[87].

Como ya adelantamos, otros ejemplos interesantes de regulación de esta cuestión se han observado en Iberoamérica en la medida en que se han orientado más a establecer una serie de principios acerca del uso ético y responsable de la IA. En 2019, la Red Iberoamericana de Protección de Datos (RIPD) elaboró un documento que contiene unas Recomendaciones Generales para el Tratamiento de Datos en la Inteligencia Artificial. Su aproximación a la ética de la IA se realiza desde la protección de los datos con un enfoque preventivo dirigido a evitar la vulneración y proteger así los derechos humanos comprometidos en el tratamiento de datos personales[88].

Por otro lado, varios países iberoamericanos han intentado llevar a cabo propuestas de regulación que reflejan las diferentes prioridades que tienen cada uno de ellos, aunque destaca siempre el papel de la ética en este ámbito. A modo de ejemplo, Colombia ha sido uno de los países pioneros en implementar las Recomendaciones sobre la Ética de la Inteligencia Artificial de la UNESCO. En el año 2021 se elaboró un Marco Ético para la Inteligencia Artificial con el objetivo de orientar el diseño, desarrollo, implementación y evaluación de sistemas de IA, asegurando el respeto de los derechos[89]. Supone una guía de *soft law* de recomendaciones y sugerencias a las Entidades Públicas para abordar la formulación y gestión de los proyectos que incluyan el uso de IA[90].

Entre las diversas iniciativas regulatorias sobre la IA que se han ido adoptando en cada país de América Latina, Brasil se sitúa a la vanguardia tras haberse aprobado en el Senado el Proyecto de Ley 2338/2023, que propone la creación de un marco regulatorio para la inteligencia artificial en este país[91]. El texto tiene como objetivo establecer los derechos

---

[87] Recla, E. y Zurita, C. (2024, 2 julio). *Diferencias y similitudes de la legislación de IA en UE, EEUU y China.* Computing. Disponible en: https://www.computing.es/opinion/legislacion-ia-en-el-mundo-eeuu-china-ue/ .

[88] Red Iberoamericana de Protección de Datos (2019, 21 junio). *Recomendaciones Generales para el Tratamiento de Datos en la Inteligencia Artificial*, p. 8.

[89] Pérez-Ugena, M. (2024)., *op. cit*., p. 151.

[90] Gobierno de Colombia (2021, mayo). *Marco ético para la inteligencia artificial en Colombia*, p. 20.

[91] Senado Federal de Brasil. (2023, mayo). *Proyecto de Ley nº 2338/2023, que dispone sobre el desarrollo, fomento y uso ético y responsable de la inteligencia artificial con base en la centralidad de la persona humana*.

de protección de las personas físicas dada su vulnerabilidad ante el impacto de los sistemas de inteligencia artificial, así como proveer herramientas de gobernanza, inspección y supervisión para crear condiciones de predictibilidad en cuanto a su interpretación[92]. De este modo, Brasil está cumpliendo con las Recomendaciones elaboradas por la OCDE y pretende adoptar un modelo lo suficientemente flexible para adaptarse a la evolución de las nuevas tecnologías. Podrá servir de ejemplo para otras naciones que aún se encuentran en debate sobre qué medidas y políticas públicas implementar para proteger la ciberciudadanía[93].
En todo caso, los distintos ritmos en la regulación de la IA de cada uno de los países de Iberoamérica reflejan la necesidad de un marco global más armonizado.

Pero es en el ámbito europeo donde se advierte una posición de liderazgo en el proceso de regulación de la IA al elaborar normas jurídicamente vinculantes a nivel regional con el fin de que tengan una mayor expansión y eficacia. La Unión Europea fue pionera con la conocida Ley de Inteligencia Artificial y a ella le ha seguido la adopción de un tratado internacional por parte del Consejo de Europa para regular esta materia. Dada la implicación que van a tener sobre la IA que podría superar el marco exclusivamente europeo y la atención que prestan a los derechos humanos en este ámbito, cabe realizar un análisis más exhaustivo de su objeto y alcance. Por ello, procederemos a un análisis por separado en los dos próximos epígrafes.

4.2. REGLAMENTO DE INTELIGENCIA ARTIFICIAL DE LA UNIÓN EUROPEA

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 (RIA), conocido como Ley de IA, tiene como objetivo mejorar el funcionamiento del mercado interior y promover la adopción de una IA centrada en el ser humano y fiable. Como el propio texto indica, esta norma busca establecer un marco jurídico uniforme para los sistemas de IA en la Unión, garantizando al mismo tiempo un nivel elevado de

[92] Pimentel Cavalcante, A. (2023). Regulación brasileña de la inteligencia artificial. *Revista de la Facultad de Derecho de México, 73* (287), p. 17.
[93] *Ibid.*, pp. 27-28.

protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta de los Derechos Fundamentales de la Unión Europea[94].

a) Elaboración y estructura del Reglamento

La elaboración de este instrumento jurídicamente vinculante se inició con la propuesta de Reglamento de la UE por el que se establecen normas armonizadas en materia de IA de 21 de abril de 2021, que se insertaba en el marco de la Estrategia sobre Inteligencia Artificial de la Unión Europea, publicada en 2018 y dirigida a promover el liderazgo de Europa en la promoción del desarrollo de una IA centrada en el ser humano, sostenible, segura, inclusiva y fiable[95]. La culminación de las negociaciones entre el Consejo y el Parlamento Europeo se produjo con la publicación final del texto en el DOUE el 12 de julio de 2024, que se compone de 180 considerandos, 113 artículos y trece anexos. Su estructura es compleja y extensa. Por un lado, los considerandos son desarrollados de forma más detallada a lo largo de todo el articulado, mientras que en los anexos se contiene información adicional que complementa aquellos artículos que requieren de mayor precisión, tales como los listados de documentación técnica que han de presentar los proveedores o la descripción de procedimientos de evaluación. Por otro lado, los artículos se organizan en trece Capítulos, centrándose los cinco primeros en los diferentes sistemas de IA que existen en atención al riesgo que presentan para la salud, la seguridad y los derechos fundamentales y estableciendo las obligaciones pertinentes a los responsables de su despliegue; y los siete capítulos siguientes dedicados a imponer una serie de directrices acerca de cómo han de introducirse estos sistemas en el mercado, su control y supervisión y las sanciones que conlleva el incumplimiento de las obligaciones.

La base jurídica de este Reglamento es el art. 114 TFUE en lo referente a la adopción de medidas para garantizar el establecimiento y el funcionamiento del mercado interior. No obstante, entre sus objetivos se incluyen cuestiones relacionadas con los derechos humanos, tales como garantizar que los sistemas de IA respeten la legislación vigente en

[94] 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. *Diario Oficial de la Unión Europea, L 1689*, de 12.7.2024, art. 1.
[95] Solar Cayón, J. I. (2025)., *op. cit*., p. 60.

materia de derechos fundamentales y los valores de la Unión[96]. De tal modo que, su estructura trata de conjugar dos marcos regulatorios de distinta naturaleza. Por un lado, en los dos primeros capítulos referentes a las Disposiciones generales (artículos 1 a 4) y a las Prácticas de IA prohibidas (artículo 5) se centra en la protección de derechos fundamentales. Por otro, los Capítulos III (artículos 6 a 49), IV (artículo 50) y V (artículos 51 a 56) se ocupan respectivamente de los sistemas de alto riesgo, obligaciones "de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA" y de los sistemas de uso general, cuyo contenido también pretende responder a ese objetivo último de protección de la salud, la seguridad y los derechos fundamentales, aunque más enfocado a la regulación del mercado. Por último, los Capítulos VI a XII se dedican al sistema de certificaciones para desarrollar una legislación de producto que permita poner en el mercado de la UE los productos de IA. Esta intrincada conexión presente en el Reglamento ha entrañado una cierta dificultad a nuestro análisis, más centrado en identificar el modo en que protege los derechos humanos que en los aspectos económicos. Ambos marcos regulatorios se entremezclan a lo largo de todo el articulado, tal como puede observarse en el Capítulo III que desarrolla los sistemas de alto riesgo en cinco secciones, dos de ellas acerca de la clasificación y los requisitos que han de cumplir para proteger la salud, la seguridad y los derechos fundamentales y las otras tres sobre las obligaciones de las partes implicadas y los procedimientos de notificación y evaluación que han de llevar a cabo.

b) Ámbito de aplicación

El RIA establece su ámbito de aplicación determinando las personas o entidades que quedan afectadas por su normativa, pero no hace una distinción de sus destinatarios en función de su carácter público o privado[97].

El artículo 2 establece que se aplicará principalmente a los proveedores y responsables del despliegue que pongan en servicio o comercialicen en la Unión sistemas de IA y modelos de uso general y que tengan su lugar de establecimiento o estén ubicados en la UE, así como a los responsables del despliegue y proveedores de sistemas de IA que estén

[96] Hernández Ramos, M. (2024). El marco jurídico regulatorio europeo de la inteligencia artificial. La relación de complementariedad entre el Reglamento de la UE y la Convención Marco del Consejo de Europa. *Revista Española de Derecho Europeo, 92*, p. 16.
[97] *Ibid.*, p. 27.

establecidos en un tercer país, cuando los resultados generados por sus sistemas se utilicen dentro del ámbito de la Unión. Igualmente será de aplicación para los importadores y distribuidores de sistemas de IA y para las personas afectadas que estén ubicadas en la Unión[98]. Su ámbito de aplicación trasciende, por tanto, estrictamente el espacio geográfico puramente europeo.

Sin embargo, quedan fuera de su alcance aquellos sistemas de IA exclusivamente con fines militares, de defensa o de seguridad nacional. Esta exclusión se justifica por el hecho de que la Unión debe respetar las funciones esenciales del Estado y porque la seguridad nacional sigue siendo responsabilidad exclusiva de los Estados miembros[99]. La exención afecta también a los sistemas que hayan sido desarrollados y puestos en servicio con la finalidad específica de la investigación y desarrollo científico, y a aquellas actividades de investigación, prueba o desarrollo relativas a sistemas de IA antes de su introducción en el mercado o puesta en servicio. En este caso, el objetivo que buscaría esta exclusión es no obstaculizar la innovación tecnológica y respetar la libertad de ciencia[100].

Tampoco será de aplicación para las autoridades públicas de terceros países u organizaciones internacionales que empleen sistemas de IA en el marco de acuerdos o de la cooperación internacionales con fines de aplicación del Derecho y de cooperación judicial con la UE o con sus Estados miembros, siempre que estos garanticen la protección de los derechos y libertades fundamentales de las personas[101]. La finalidad de esta exclusión es preservar los acuerdos existentes y las necesidades especiales de cooperación futura con socios extranjeros con los que se intercambian información y pruebas en el marco de acuerdos de cooperación policial y judicial[102].

c) Clasificación de los sistemas en función del riesgo

El aspecto más destacado de esta ley y por lo que resulta del máximo interés para el objeto de nuestro trabajo es la clasificación de los sistemas de IA en función del riesgo que presentan para la salud, la seguridad y los derechos fundamentales de las personas. Este enfoque permite establecer distintos niveles de obligaciones en base al riesgo, llevando

[98] Barrio Andrés, M. (2024). Objeto, ámbito de aplicación y sentido del Reglamento Europeo de Inteligencia Artificial. En M. Barrio Andrés (Dir.), *El Reglamento Europeo de Inteligencia Artificial* (p. 34). Tirant lo Blanch.
[99] Reglamento (UE) 2024/1689., *op. cit*., Considerando 24.
[100] Hernández Ramos, M. (2024)., *op. cit.*, p. 24.
[101] Reglamento (UE) 2024/1689., *op. cit*., Artículo 2.4.
[102] Reglamento (UE) 2024/1689., *op. cit.*, Considerando 22.

así al RIA a prohibir determinados usos considerados de riesgo inaceptable, imponer obligaciones estrictas para usos de riesgo alto, establecer obligaciones de transparencia para ciertos sistemas de riesgo limitado, y dejar sin regulación imperativa al resto de sistemas de riesgo mínimo[103]. Se expondrán a continuación los diferentes tipos teniendo en cuenta los requisitos que deben cumplir para ser considerados como tales, así como las obligaciones de los sujetos a quienes se aplica.

Sistemas prohibidos

El Capítulo II del RIA recoge en su artículo 5 todas aquellas situaciones en las que la comercialización, puesta en servicio o uso de un sistema de IA deben considerarse prohibidas bajo ciertos condicionantes. A continuación, se exponen los ocho supuestos que la norma desarrolla en estos términos, aunque en algunos casos se prevén determinadas excepciones a la prohibición.

a) En primer lugar, los sistemas que empleen técnicas subliminales o técnicas deliberadamente manipuladoras o engañosas para alterar el comportamiento de una persona o grupo, mermando su capacidad para tomar una decisión informada y haciendo que tome una decisión que de otro modo no habría tomado, de modo que provoque, o sea probable que provoque, perjuicios considerables a esa persona o a otras.
b) Los sistemas que aprovechen las vulnerabilidades de una persona o grupo específico derivadas de su edad, discapacidad o de una situación social o económica específica, con la intención de alterar su comportamiento, provocando o pudiendo provocar perjuicios considerables a esa persona o a otras.
c) Sistemas utilizados para evaluar o clasificar a personas o a grupos en función de su comportamiento social o características personales o de su personalidad conocidas, inferidas o predichas, de manera que la puntuación o clasificación social resulte en un trato perjudicial o desfavorable hacia determinadas personas en contextos sociales no relacionados con aquellos donde se generaron los datos originalmente; o que se provoque un trato perjudicial o desfavorable injustificado respecto de su comportamiento social o su gravedad.
d) Aquellos sistemas predictivos de delitos en base únicamente a la evaluación de los rasgos y características de la personalidad del sujeto. Esta prohibición no se

[103] Barrio Andrés, M. (2024)., *op. cit.*, p. 30.

aplicará a los sistemas que se empleen para apoyar la valoración humana de la implicación de la persona en el delito basándose en hechos objetivos y verificables directamente relacionados con una actividad delictiva. Se protegen de esta manera los derechos procesales del acusado y se garantiza que la IA sea una herramienta de apoyo al órgano judicial para evitar los sesgos que puedan introducirse.

e) Se prohíben también los sistemas que creen o amplíen bases de datos de reconocimiento facial mediante la extracción no selectiva de imágenes faciales de internet o de circuitos cerrados de televisión. En este caso el derecho fundamental más en peligro es el derecho a la intimidad y la privacidad de las personas debido a la utilización de sus imágenes sin su autorización[104].
f) Los sistemas empleados para inferir las emociones de las personas en los lugares de trabajo y en los centros educativos. Se exceptúa la prohibición si su fin es introducirlos en el mercado por motivos médicos o de seguridad. En algunos contextos como el ámbito laboral, la ley trata con mayor atención el impacto sobre los derechos por entender que estas personas se encuentran en una situación de mayor vulnerabilidad o desventaja[105].
g) Los sistemas de categorización biométrica para la clasificación individual de personas físicas sobre la base de sus datos biométricos para deducir o inferir sus datos personales, tales como la raza, opiniones políticas, convicciones religiosas u orientación sexual. Este tipo de sistemas pueden ejercer un control social sobre las personas al asignarlas a distintas categorías en función de estos datos, resultando en prácticas contrarias al derecho a la no discriminación y necesitando una protección más severa.
h) Por último, se prohíbe el uso de sistemas de identificación biométrica remota en tiempo real en espacios de acceso público con fines de garantía de cumplimiento del Derecho. No obstante, se prevén tres excepciones en caso de que su uso sea estrictamente necesario para los siguientes objetivos; la búsqueda selectiva de víctimas de secuestro, trata de seres humanos, explotación sexual o personas desaparecidas; para prevenir amenazas para la vida o la seguridad física de las personas o amenazadas de atentado terrorista; y localizar o identificar personas sospechosas de la comisión de un delito. De modo que se protege el derecho a la

[104] Laukyte, M. (2024). Reflexión sobre los derechos fundamentales en la nueva Ley de la Inteligencia Artificial. *Derechos y Libertades: Revista de Filosofía del Derecho y Derechos Humanos, 51,* p. 161.
[105] *Ibid.*, p. 163.

intimidad de las personas, salvo en casos extremos en los que se vean en peligro otros derechos como la vida o la integridad física.

La prohibición de este tipo de sistemas se realiza con la finalidad de proteger los derechos humanos de aquellas personas que se ven afectadas por las decisiones que adopta el algoritmo, especialmente en aquellos casos en que hay una mayor vulnerabilidad o cuando entran en juego datos sensibles, pero también se establecen excepciones por las mismas razones cuando se trata de proteger otros derechos. En este sentido, el impacto de la recopilación y uso de datos biométricos de forma inadecuada suponen un mayor peligro para los derechos, ya que estos tienen una conexión directa con los valores de autonomía y dignidad humanas[106].

Sistemas de alto riesgo

Dentro de la clasificación de los sistemas de inteligencia artificial, el RIA pone especial atención en aquellos que pueden considerarse como de alto riesgo y dedica un extenso desarrollo en su Capítulo III a su delimitación y a las obligaciones que implican.

Un sistema de IA será clasificado como de alto riesgo cuando tenga un efecto perjudicial considerable en la salud, la seguridad y los derechos fundamentales de las personas de la Unión[107]. Entre las numerosas reglas para clasificar los sistemas de alto riesgo, cabe mencionar algunas de las previstas en el Anexo III del RIA por su especial impacto en los derechos humanos.

En primer lugar, aquellos casos en los que los sistemas de uso de datos biométricos no estén prohibidos, serán considerados de alto riesgo debido a su probabilidad de dar lugar a resultados sesgados y tener efectos discriminatorios, siendo el riesgo pertinente en lo que respecta a la edad, la etnia, la raza, el sexo o la discapacidad. Se excluyen de esta clasificación aquellos vinculados a la ciberseguridad y a las medidas de protección de datos personales[108].

En referencia al uso de la IA en la Administración de Justicia, se considerarán sistemas de alto riesgo aquellas herramientas de IA empleadas para ayudar a las autoridades judiciales a investigar e interpretar los hechos y el Derecho y a aplicar la ley a unos hechos

[106] Garriga Domínguez, A. (2024). Los derechos ante los sistemas biométricos que incorporan Inteligencia Artificial. *Derechos y Libertades: Revista De Filosofía Del Derecho y Derechos Humanos* (51), p. 123.
[107] Reglamento (UE) 2024/1689., *op. cit*., Artículo 6.2.
[108] Reglamento (UE) 2024/1689., *op. cit.*, Considerando 54.

concretos. Este enfoque supone un avance en la protección jurídica frente a los riesgos de la automatización mediante IA y destaca la relevancia de que siempre haya supervisión humana en su uso en el sistema judicial[109].

También en el ámbito laboral se ha de tener en cuenta que los sistemas empleados para evaluar a candidatos y filtrar las solicitudes de empleo en los procesos de selección y contratación de personas suponen un alto riesgo para sus derechos. Serán igualmente considerados los sistemas de IA que se utilizan para tomar decisiones que afectan a las condiciones laborales o a la promoción o rescisión de relaciones contractuales en este ámbito, por ejemplo, para la asignación de tareas a partir de comportamientos individuales. Se reconoce que estos sistemas afectan a los derechos de los trabajadores y pueden perpetrar patrones históricos de discriminación contra determinados colectivos como las mujeres o las personas con discapacidad. Además, su uso con la finalidad de control del rendimiento y comportamiento de los trabajadores puede socavar sus derechos fundamentales a la protección de los datos personales y a la intimidad[110].

En el ámbito educativo, un mal diseño y utilización de los sistemas, especialmente aquellos que son empleados para determinar el acceso o la admisión a un centro educativo o para evaluar los resultados de aprendizaje de las personas, pueden vulnerar el derecho a la educación y la formación, así como el derecho a no sufrir discriminación[111].

Asimismo, se ha de prestar especial atención al contexto de la migración, el asilo y la gestión del control fronterizo donde las personas se encuentran en una situación de vulnerabilidad. Los sistemas de IA que utilicen las autoridades públicas competentes para, entre otras funciones, evaluar los riesgos que presenten las personas físicas que entren en el territorio de un Estado miembro o las destinadas a ayudar a examinar las solicitudes de asilo, visado y permiso de residencia, serán considerados de alto riesgo por su posible afectación a los derechos a la libre circulación, la no discriminación, la intimidad personal y la protección internacional[112].

Entre las obligaciones que se establecen para los responsables del despliegue de estos sistemas, el artículo 14 destaca la importancia de la supervisión humana en el diseño y desarrollo de los sistemas de IA de alto riesgo, que deberán ser vigilados efectivamente

[109] Conde Fuentes, J. (2024). Sistemas de alto riesgo en el Reglamento Europeo de Inteligencia Artificial: La ponderación de los derechos fundamentales en juego. En T. J. Aliste Santos (Coord.), *Eficiencia Procesal I* (p. 121). Atelier.
[110] Reglamento (UE) 2024/1689., *op. cit*., Considerando 57.
[111] Reglamento (UE) 2024/1689., *op. cit.*, Considerando 56.
[112] Reglamento (UE) 2024/1689., *op. cit.*, Considerando 60.

por personas físicas con la finalidad de prevenir o reducir los riesgos para la salud, la seguridad o los derechos fundamentales a que pueden dar lugar. Se ha de garantizar un sistema de gestión de riesgos robusto para lograr que las medidas de supervisión sean proporcionales a los riesgos. Ello va a permitir que se adopten las medidas pertinentes en caso de que surja la necesidad de detener el sistema o mitigar los riesgos residuales que pudieran convertirse en inaceptables[113].

En este sentido, cabe destacar lo previsto en el artículo 27 acerca de la evaluación de impacto relativa a los derechos fundamentales. Se pretende con ella someter a una evaluación a algunos sistemas de alto riesgo antes de su despliegue. Los responsables serán organismos de Derecho público o entidades privadas que presten servicios públicos y, entre otros datos, deberán informar de las categorías de personas físicas y grupos que puedan verse afectados, los riesgos de perjuicios específicos para aquéllos y las medidas que se adoptarán en caso de que los riesgos se materialicen[114]. Se observa aquí la influencia que ha podido tener la Recomendación ética de la UNESCO que, entre sus propuestas de acción política, prevé las evaluaciones de impacto ético.

Sistemas de riesgo limitado

El artículo 50 del Reglamento está dedicado a establecer una serie de obligaciones limitadas de transparencia para proveedores e implantadores de determinados sistemas de IA destinados a interactuar con personas físicas *(chatbots*) o a generar contenidos que pueden presentar ciertos riesgos, independientemente de si cumplen las condiciones para ser considerados como de alto riesgo o no[115]. Los riesgos que pueden plantear pueden ser de suplantación o engaño, de modo que se obliga a que estas personas sean debidamente informadas de que están interactuando con un sistema de IA, especialmente teniendo en cuenta si pertenecen a colectivos vulnerables debido a su edad o discapacidad.

En el caso de los proveedores de sistemas de IA que generen contenido sintético de audio, imagen, vídeo o texto, tendrán que etiquetar tales contenidos con una marca de agua para reflejar que han sido generados de manera artificial[116].

[113] Miguez Macho, L. y Torres Carlos, M. (2024). Sistemas de IA prohibidos y sistemas de IA de alto riesgo. En M. Barrio Andrés (Dir.). *El Reglamento Europeo de Inteligencia Artificial* (p. 68). Tirant lo Blanch.
[114] Reglamento (UE) 2024/1689., *op. cit.*, Artículo 27.
[115] Reglamento (UE) 2024/1689., *op. cit.*, Artículo 50.1.
[116] Barrio Andrés, M. (2024)., *op. cit*., p. 44.

Además, aquellos sistemas de reconocimiento de emociones y sistemas de categorización biométrica que no sean clasificados como prohibidos, deberán informar de que están siendo usados y de su funcionamiento[117].

Por último, se incluyen las ultrasuplantaciones, tales como las *deepfakes*, consistentes en la generación o manipulación de contenidos audiovisuales y que constituyen una falsificación profunda. La obligación de los responsables del sistema será la de revelar que el contenido ha sido generado artificialmente.

En cualquiera de los casos previstos, están exentos de estas obligaciones aquellos sistemas de IA autorizados por ley para detectar, prevenir, investigar o enjuiciar delitos penales.

Sistemas de riesgo mínimo

El resto de los sistemas de IA que no cumplen los requisitos para ser considerados como de riesgo en los términos anteriormente expresados, quedan excluidos del ámbito de aplicación y, por tanto, de las obligaciones que impone el Reglamento. No obstante, mediante esta regulación se busca fomentar a los proveedores de este tipo de sistemas a crear códigos de conducta con objeto de que apliquen voluntariamente de forma total o parcial los requisitos aplicables a los sistemas de riesgo adaptándolos a sus condiciones[118].

d) Régimen sancionador

El RIA contiene un sistema de sanciones que impone una serie de multas administrativas bastante significativas a aquellos operadores que infrinjan las disposiciones del Reglamento. Dichas sanciones serán proporcionadas y disuasorias y se aplicarán en función de la gravedad del incumplimiento de las obligaciones. Así pues, para las infracciones más graves en relación con los sistemas de IA prohibidos, se prevén multas de hasta 35 millones de euros o de hasta el 7% del volumen de negocio total anual. Para los incumplimientos relacionados con sistemas de alto riesgo o con las obligaciones de transparencia previstas en el artículo 50, este tipo de infracciones graves podrán acarrear sanciones de 15 millones de euros o hasta el 3% de su volumen de negocios mundial total. Finalmente, para infracciones leves como la presentación de información inexacta,

[117] Solar Cayón, J. I. (2025)., *op. cit*., p. 75.
[118] Solar Cayón, J. I. (2025)., *op. cit*., p. 76.

engañosa o incompleta, se fijan sanciones de hasta 7,5 millones de euros o del 1% del volumen del negocio[119].

e) Entrada en vigor

En cuanto a la entrada en vigor y aplicación del Reglamento, tiene ciertas complejidades ya que los distintos preceptos se irán aplicando de manera escalonada en un amplio espacio de tiempo, dejando así un margen a los Estados para que se adecúen a las actuaciones que la norma exige. Este escalonamiento responde a la necesidad de incorporar la IA en el mercado de manera flexible y progresiva para lograr que se amolde correctamente. De tal modo que, el Reglamento no se aplicará de forma completa hasta que hayan transcurrido tres años desde su publicación en el DOUE, siendo además obligatorio en todos sus elementos y directamente aplicable en cada Estado miembro.
Su entrada en vigor se produjo el 2 de agosto de 2024 y este se empezará a aplicar dos años después, salvo determinados preceptos que ya se podrán aplicar durante este transcurso de tiempo. De hecho, los Capítulos I y II referentes a las disposiciones generales del Reglamento y a las prácticas de IA prohibidas se vienen aplicando desde el 2 de febrero de 2025, es decir, 6 meses después de la entrada en vigor. A partir del 2 de agosto de este mismo año se aplicarán las disposiciones relativas a las autoridades notificantes y organismos notificados (Sección 4, Capítulo III), a los modelos de IA de uso general (Capítulo V), a la estructura de gobernanza (Capítulo VII), a las sanciones, salvo las correspondientes a proveedores de modelos de IA de uso general (Capítulo XII) y el artículo 78 sobre la confidencialidad de la información. Por último, se establece que lo dispuesto acerca de los sistemas clasificados como de alto riesgo (artículo 6.1) y las obligaciones correspondientes del RIA serán aplicados a partir del 2 de agosto de 2027[120].

En definitiva, esta reciente norma supone un gran avance y un precedente para la elaboración de futura normativa en materia de IA a nivel internacional. Su enfoque se centra en el tratamiento de la IA como producto y su comercialización en el mercado de la Unión. No obstante, reconoce los riesgos de esta tecnología para los derechos fundamentales y los clasifica para establecer distintas obligaciones. Su carácter

[119] Miranzo Díaz, J. (2025). El Reglamento de Inteligencia Artificial de la Unión Europea: regulación de riesgos y sistemas de estandarización. *A&C- Revista de Direito Administrativo y Constitucional, 96*, p. 61
[120] Reglamento (UE) 2024/1689., *op. cit.*, Artículo 113.

vinculante para los Estados miembros va a posibilitar que se adopten las medidas necesarias para poder frenar cualquier tipo de vulneración que ocurra en estos términos. Se espera que sirva de inspiración y de empuje a otros países fuera de Europa, siguiendo "el efecto Bruselas", especialmente en lo referente a aquellas disposiciones del RIA que establecen estándares regulatorios estrictos, como ocurre con la clasificación de los sistemas de IA en función del riesgo[121]. También contribuirá a ello el hecho de que la aplicación del RIA se extienda a aquellos proveedores y responsables del despliegue no europeos, cuyos sistemas se utilicen dentro del ámbito de la Unión.

4.3. CONVENIO MARCO DEL CONSEJO DE EUROPA SOBRE INTELIGENCIA ARTIFICIAL Y DERECHOS HUMANOS, DEMOCRACIA Y ESTADO DE DERECHO

El Consejo de Europa se ha sumado a la regulación de la inteligencia artificial con la adopción del Convenio marco sobre inteligencia artificial y derechos humanos, democracia y Estado de Derecho, adoptado el 17 de mayo de 2024 por el Comité de Ministros y abierto a la firma el 5 de septiembre de 2024. Se trata del primer tratado internacional jurídicamente vinculante en este ámbito[122]. El objeto de este instrumento, así como las obligaciones que impone a quienes lo ratifiquen, toman como base los valores propios del Consejo de Europa, pretendiendo que las actuaciones que se adopten vayan dirigidas a proteger los derechos humanos, la democracia y el Estado de Derecho. Se obliga así a las Partes especialmente a garantizar una protección de los derechos humanos en todas las actividades de diseño, desarrollo e implementación de los sistemas de inteligencia artificial.

a) Elaboración y estructura del Convenio

La elaboración del Convenio se inició por el Comité *Ad Hoc* sobre Inteligencia Artificial (CAHAI), creado en el año 2019 con el mandato de examinar la viabilidad del

[121] López- Tarruella Martínez, A. (2025). Claroscuros del "efecto Bruselas" del Reglamento de Inteligencia Artificial. *Revista Española de Derecho Internacional, 77* (1), p. 242.
[122] Consejo de Europa. (2024). *Convenio Marco sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho (CETS Nº 225)*, de 5 de septiembre de 2024.

instrumento ante la necesidad de establecer un marco legal globalmente aplicable que previera una serie de principios generales y normas para regular el ciclo de vida de los sistemas de inteligencia artificial. Su prioridad se basaba en que en la actividad de dichos sistemas se preservaran los valores compartidos en la Organización y se favorecieran al mismo tiempo el progreso tecnológico y la innovación. No obstante, el órgano que se encargó de la negociación y redacción del texto fue el Comité sobre Inteligencia Artificial (CAI), que sucedió en su mandato al CAHAI en el 2022. Su intención era crear un instrumento que trascendiera fronteras y fuese atractivo para un gran número de Estados más allá de Europa[123]. Para ello se contó con la participación de los 46 Estados miembros del Consejo de Europa, los 5 Estados observadores, 6 Estados no miembros (Australia, Argentina, Costa Rica, Israel, Perú y Uruguay), la Unión Europea e incluso representantes internacionales de la sociedad civil y algunas organizaciones internacionales como la OCDE[124].

Pese al riesgo de no llegar a un acuerdo por la presión ejercida por Estados Unidos para que el Convenio no se aplicara a sujetos privados[125], finalmente se adoptó el 17 de mayo de 2024 este extenso texto que incluye diecisiete considerandos y treinta y seis artículos repartidos en ocho capítulos. Aunque se adopta un enfoque con base en el riesgo e impacto, este no condiciona su estructura como ocurre en el RIA que desarrolla su contenido distinguiendo los distintos tipos de prácticas de IA en función de sus riesgos[126]. Además del Capítulo I (artículos 1 a 3) que contiene una serie de disposiciones generales, y el Capítulo VIII (artículos 27 a 36) de cláusulas finales, el Convenio engloba los siguientes capítulos: el Capítulo II (artículos 4 y 5) sobre obligaciones generales, entre las que se encuentra la protección de derechos humanos (artículo 4); el Capítulo III (artículos 6 a 13) en relación con los principios relativos a las actividades llevadas a cabo en el marco del ciclo de vida de los sistemas de inteligencia artificial; el Capítulo IV (artículos 14 y 15) sobre recursos; Capítulo V (artículo 16) dedicado a la evaluación y mitigación de los riesgos y de los impactos negativos; el Capítulo VI (artículos 17 a 22) sobre la aplicación del Convenio; y el Capítulo VII (artículos 23 a 27) relativo al mecanismo de seguimiento y a la cooperación.

---

[123] Cotino Hueso, L. (2024). El Convenio sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho del Consejo de Europa. *Revista Administración y Ciudadanía, EGAP, 18* (2), p. 4.
[124] Presno Linera, M. A. y Meuwese, A. (2024). La regulación de la Inteligencia Artificial en Europa. *Teoría y Realidad Constitucional, 54*, p. 137.
[125] Volpicelli, G. (2024, 11 marzo). International AI rights treaty hangs by a thread. *Politico*. Disponible en: https://www.politico.eu/article/council-europe-make-mockery-international-ai-rights-treaty/ .
[126] Hernández Ramos, M. (2024)., *op. cit.*, p. 19.

b) Ámbito de aplicación

Este Convenio establece una importante distinción en cuanto a su aplicación en el sector público o en el sector privado, lo que supone una significativa diferencia respecto del RIA. Por un lado, el art. 3. a) abarca el uso de sistemas de inteligencia artificial por parte de las autoridades públicas o agentes privados que actúen en su nombre. Por otro lado, la aplicación a los actores privados es más flexible dado que en estos casos las Partes en el Convenio son quienes tendrán que abordar los riesgos e impactos surgidos de sus actividades. Para dar cumplimiento a esta disposición podrán optar por aplicar los principios y obligaciones establecidos en los Capítulos II a VI; o adoptar otras medidas adecuadas para cumplir las disposiciones del tratado respetando plenamente sus obligaciones internacionales en materia de derechos humanos, democracia y Estado de Derecho[127]. La forma que adopten para implementar esta obligación deberá especificarse en una declaración dirigida al Secretario General en el momento de firma o depósito del instrumento de ratificación, aceptación, aprobación o adhesión. Hasta el momento, solo Noruega había efectuado esta declaración para acogerse a las disposiciones pertinentes del Convenio en el momento en que firmó el instrumento con fecha de 5 de septiembre de 2024. No obstante, la reciente firma del Convenio por parte de Ucrania el 15 de mayo de 2025 ha incluido también esta declaración, pero indicando que adoptará otro tipo de medidas respecto de las actividades que desarrollen los actores privados. Además, ha realizado una especificación en cuanto al ámbito territorial de aplicación, declarando que el Convenio no se aplicará al territorio ucraniano ocupado por Rusia hasta que esta agresión armada termine y se haya restablecido la integridad territorial de Ucrania[128].

Por otra parte, el texto excluye de manera explícita su aplicación sobre la seguridad y defensa nacional, aclarando que será cada Estado quien las regule, pero de forma compatible con el Derecho internacional aplicable, de modo que en este ámbito se concede una muy fuerte discrecionalidad a los Estados[129]. Igualmente, se producen exclusiones de aplicación a las actividades de investigación y desarrollo relativas a sistemas de inteligencia artificial que aún no estén disponibles para su uso, a no ser que se realicen pruebas o actividades similares que tengan el potencial de interferir con los

[127] Hernández Ramos, M. (2024)., *op. cit.*, p. 28.
[128] Página web del Consejo de Europa. Disponible en: https://www.coe.int/en/web/conventions/full-list?module=declarations-by-treaty&numSte=225&codeNature=0 . Información consultada por última vez el 16.06.2025.
[129] Cotino Hueso, L. (2024)., *op. cit.*, p. 9.

derechos humanos, la democracia y el Estado de Derecho[130]. Esto evidencia que la exclusión no puede suponer la derogación de la vigencia de los derechos fundamentales reconocidos[131]. No obstante, esta excepción a la investigación debe darse sin perjuicio de las medidas de innovación que establezca cada Parte para poder probar sistemas de inteligencia artificial bajo la supervisión de sus autoridades competentes[132] o de los intercambios de información que deben darse entre las Partes acerca de los efectos que hayan surgido en contextos de investigación[133]. Ambas exclusiones guardan, por tanto, un importante paralelismo con el RIA.

c) Principios y obligaciones

El Capítulo III (artículos 6 a 13) está dedicado a establecer una serie de principios clave en relación con las actividades implicadas durante el ciclo de vida de los sistemas de inteligencia artificial. La condición de Convenio Marco supone que, pese a su carácter vinculante y a que imponga conductas de naturaleza obligatoria, no se establecen obligaciones muy concretas en todos los casos, sino que mayormente se incorporan indicaciones normativas genéricas[134]. De este modo, se otorga un amplio margen de discrecionalidad a las Partes para adecuarlos a las características de sus ordenamientos nacionales. A este respecto, los principios previstos son los siguientes: la dignidad humana y autonomía individual, transparencia y supervisión, rendición de cuentas y responsabilidad, igualdad y no discriminación, privacidad y protección de datos personales, fiabilidad e innovación segura. Se puede observar que coinciden sustancialmente con los ya recogidos en la Recomendación sobre la ética de la inteligencia artificial elaborada por la UNESCO en 2021, de modo que se denota cierta influencia de las iniciativas de *soft law* previas a este instrumento jurídicamente vinculante.

En cuanto al primero de estos principios, no cabe duda de que la dignidad humana constituye un elemento esencial del sistema internacional de protección de derechos humanos y está estrechamente vinculada a la autonomía individual que permite facilitar

[130] Consejo de Europa. (2024)., *op. cit.*, Artículo 3.3.
[131] Cotino Hueso, L. (2024)., *op. cit*., p. 8.
[132] Consejo de Europa. (2024)., *op. cit.*, Artículo 13.
[133] Consejo de Europa. (2024)., *op. cit.*, Artículo 25.2.
[134] Díaz Galán, E. C. (2025). 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, 17* (1), p. 375.

el ejercicio de otros derechos y libertades[135]. Por otro lado, se insta además a las Partes a que adopten medidas para asegurar que los sistemas de IA respetan la igualdad, incluida la igualdad de género, así como la prohibición de discriminación y la no vulneración del derecho a la privacidad de los individuos. El problema radica en que, dado que estas obligaciones han de ser aplicadas teniendo en cuenta el Derecho Internacional y nacional, su contenido real podrá variar significativamente en función de los diferentes sistemas jurídicos de cada Estado Parte. Su aplicación en la regulación vendrá determinada por especificades del régimen político nacional tales como la eficacia de las instituciones que determinan las normas de comportamiento de los participantes en el ciclo de vida de los sistemas de IA o el papel que juegan la sociedad civil y las organizaciones de protección de los derechos humanos y libertades[136].

En cambio, hay algunos principios más específicos como son el de transparencia y supervisión que pueden considerarse condiciones previas para garantizar el respeto, la protección y la promoción de los derechos humanos. Con ellos se recalca la importancia de que las decisiones que toman los sistemas de IA puedan ser claramente explicadas y se facilite así la posibilidad de impugnarlas, ya que la falta de transparencia podría llegar a vulnerar el derecho a un juicio imparcial y a un recurso efectivo[137].

Por su parte, el Capítulo IV (artículos 14 y 15) contempla las obligaciones en materia de acceso a recursos efectivos si se produjeran violaciones de derechos humanos a causa del desarrollo de actividades de IA. Ello implica que las personas que interactúan con estos sistemas sean informadas de que están interactuando con ellos y no con humanos. Las Partes tendrán que proporcionar las garantías procesales, salvaguardias y derechos efectivos a las personas que vean afectado el disfrute de los derechos humanos debido al uso de los sistemas de IA, lo cual implica la posibilidad de presentar una denuncia ante las autoridades competentes.

El Capítulo V del Convenio consta de un único artículo en el que se prevé que las Partes lleven a cabo la gestión de riesgos e impactos sobre los derechos humanos, la democracia y el Estado de Derecho adoptando medidas diferenciadas en función de la severidad y

[135] Cano Linares, M. Á. (2025). Inteligencia Artificial y derechos humanos: el Convenio Marco del Consejo de Europa. En M. Á. Cano Linares y A. I. Carreras Presencio (Dirs*.), Nuevas Realidades: un enfoque desde el derecho* (p. 46). Dykinson.

[136] Martynova, E. (2024). Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law: a Commentary. En L. Belli y W. B. Gaspar (Eds.), *AI From the Global Majority: Official Outcome of the UN IGF Data and Artificial Intelligence Governance Coalition* (p. 114). FGV- Direito Rio.

[137] Cano Linares, M. Á. (2025)., *op. cit.*, p. 47.

probabilidad de que se produzcan efectos adversos. De este modo, habrán de llevarse a cabo evaluaciones respecto de las repercusiones reales y potenciales para posteriormente implementar las debidas medidas de mitigación que permitan abordar efectivamente los riesgos identificados. Este tipo de medidas incluyen la realización de pruebas de los sistemas de IA antes de ponerlos a disposición para su primer uso. Se hace especial mención a la posibilidad de que las Partes evalúen la necesidad de una moratoria o prohibición o cualquier medida apropiada en caso de que consideren los riesgos identificados como incompatibles. Se observa aquí una diferencia con lo establecido en el artículo 5 del RIA que contiene un amplio listado de prohibiciones, dado que el Convenio no prevé la prohibición de sistemas de IA concretos, pese a que el CAHAI en sus inicios sugería prohibir total o parcialmente ciertas aplicaciones de IA[138].

d) Mecanismos de seguimiento del Convenio y cooperación

El Convenio Marco establece como mecanismo de seguimiento principal la Conferencia de las Partes, un foro en el que participan representantes oficiales de las Partes y en el que se consultan periódicamente sobre diversas cuestiones del Convenio[139]. De este modo, se posibilita el intercambio de información para facilitar recomendaciones interpretativas o la resolución de las controversias existentes e incluso se prevé la realización de audiencias públicas[140]. Esta Conferencia se podrá convocar por el Secretario General del Consejo de Europa o por solicitud de una mayoría de las Partes o del Comité de Ministros[141]. La previsión de un órgano de seguimiento del Convenio es de gran relevancia, y se alinea con la práctica habitual del Consejo de Europa en el marco de los tratados sobre derechos humanos. Bien es cierto que se trata de un órgano con una composición de naturaleza fuertemente gubernamental y política, muy diferente de los comités de expertos creados en otros casos.

Los redactores del Convenio entendieron que la cooperación internacional entre las Partes es clave para que se produzca un intercambio de información útil concerniente a aspectos sobre los efectos positivos y negativos de la inteligencia artificial en el disfrute de los derechos humanos[142].

[138] Cotino Hueso, L. (2024)., *op. cit*., p. 4.
[139] Consejo de Europa. (2024)., *op. cit.*, Artículos 23.1 y 23.2.
[140] Cano Linares, M. Á. (2025)., *op. cit.*, p. 49.
[141] Consejo de Europa. (2024)., *op. cit.*, Artículo 23.3.
[142] Consejo de Europa. (2024)., op. cit., Artículo 25.

De este modo, con objeto de garantizar un seguimiento efectivo del cumplimiento de sus obligaciones, las Partes deberán proporcionar un informe a la Conferencia de las Partes una vez transcurridos dos años desde su adhesión, y posteriormente con carácter periódico. En este informe deberán indicar detalladamente cómo se han abordado los riesgos e impactos sobre los derechos humanos, la democracia y el Estado de Derecho que hayan podido tener las actividades desarrolladas durante el ciclo de vida de los sistemas de inteligencia artificial por parte de actores públicos y privados[143].

Por último, el Convenio prevé que las Partes tienen el deber de designar uno o varios mecanismos para supervisar el cumplimiento de las obligaciones contraídas en el Convenio. Tendrán que ejercer sus funciones de manera independiente e imparcial, pero de existir varios tendrán que tomar medidas para facilitar la cooperación entre ellos[144].

Se observa entonces que el Convenio Marco no establece un sistema de gobernanza tan exhaustivo como el Reglamento, otorgando la función de seguimiento de su aplicación a un órgano fundamentalmente político de representación de los Estados, la Conferencia de las Partes, y dando un amplio margen de discrecionalidad a las Partes para la supervisión de sus obligaciones.

El Consejo de Europa a través del CAI ha dado un paso más allá, no sólo codificando la importancia de la protección de los derechos humanos en el ámbito de la inteligencia artificial y estableciendo un mecanismo específico para el seguimiento de la aplicación del Convenio por las Partes, sino que también ha creado una herramienta para dar directrices a los Estados en la aplicación de las disposiciones del Convenio. Se trata de HUDERIA, una guía diseñada para llevar a cabo evaluaciones de riesgo e impacto de los sistemas de IA que pretende ayudar a los agentes públicos y privados a identificar y abordar los riesgos en los derechos humanos, la democracia y el Estado de derecho[145]. Esta guía no es legalmente vinculante, sino que los Estados parte en el Convenio tienen flexibilidad para usarla en el desarrollo de sus obligaciones de evaluación de riesgos. Su función es asesorar acerca de cómo prevenir y mitigar los riesgos en la aplicación de tecnologías de inteligencia artificial. De este modo, el Consejo de Europa no sólo ha desarrollado una labor de codificación en este ámbito, sino que la ha completado con su asistencia técnica a las Partes a través de una herramienta creada al efecto.

---

[143] Consejo de Europa. (2024)., op. cit., Artículo 24.

[144] Artículo 26.

[145] Comité sobre Inteligencia Artificial. (2024). *Metodología para la evaluación del riesgo y el impacto de los sistemas de inteligencia artificial desde el punto de vista de los derechos humanos, la democracia y el Estado de Derecho (HUDERIA Methodology), (CAI(2024)16rev2)*, de 28 de noviembre de 2024, p. 3.

En concreto, esta guía adopta un enfoque sociotécnico al considerar que los aspectos del ciclo de vida de los sistemas de IA están interconectados con las elecciones humanas y las estructuras sociales. Su estructura combina orientaciones generales y específicas y ofrece gran flexibilidad para adaptarse a diferentes contextos, necesidades y capacidades. A nivel general, se establece la metodología HUDERIA con el objetivo de definir criterios claros y concretos para identificar contextos y aplicaciones en los que el uso de estos sistemas podría entrañar riesgos significativos[146]. A nivel más específico se ha configurado el modelo HUDERIA que tendrá su desarrollo a lo largo del 2025 y cuya función será proporcionar materiales y fuentes para ayudar a implementar la metodología[147]. Por lo tanto, con esta nueva herramienta se podría garantizar el seguimiento y la revisión iterativa de los sistemas que empleen IA, colaborando al mismo tiempo con las prácticas de evaluación y control existentes en la industria[148].

e) Entrada en vigor

El proceso de entrada en vigor de este Tratado requiere que cinco signatarios, tres de los cuales han de ser Estados miembros del Consejo de Europa, expresen su consentimiento para estar obligados por el mismo. Para ello, habrán de depositar ante el Secretario General del Consejo de Europa los correspondientes instrumentos de ratificación, aceptación o aprobación del Convenio. Desde esa fecha, entrará en vigor el primer día del mes siguiente después de que haya transcurrido un período de tres meses[149]. Se prevé igualmente la posibilidad de invitar a cualquier Estado no miembro que no haya participado en la elaboración a acceder al Convenio, tras obtener el consentimiento unánime de las Partes una vez este haya entrado en vigor[150].

Actualmente este Convenio Marco ha sido firmado por once Estados miembros del Consejo de Europa (Andorra, Georgia, Islandia, Liechtenstein, Montenegro, Noruega, Moldavia, San Marino, Suiza, Ucrania y Reino Unido), la Unión Europea y cuatro países observadores (Canadá, Estados Unidos, Israel y Japón); si bien hasta la fecha ninguno ha

[146] Graf, F., Obrecht, L. y Weiner, S. (2022). Erste Erkenntnisse zu Transparenz, Diskriminierung und Manipulation. Rechtliche Rahmenbedingungen für Künstliche Intelligenz in der Schweiz. *Jusletter, 12*, 3-4.
[147] Comité sobre Inteligencia Artificial. (2024)., *op. cit.*, p. 5.
[148] Graf, F., Obrecht, L. y Weiner, S. (2022)., *op. cit.*, p. 4.
[149] Consejo de Europa. (2024)., *op. cit.*, Artículo 30.
[150] Consejo de Europa. (2024)., *op. cit.*, Artículo 31.

prestado aún su consentimiento en quedar obligado por este Tratado[151]. Pese a ello, se puede observar que tanto los Miembros como aquellos que actúan como observadores vienen demostrando un compromiso por mostrarse obligados al cumplimiento de las disposiciones de este instrumento.

Su proyección universal va a ser crucial para que esta regulación de la IA amplifique su ámbito. La condición de Convenio Marco facilita la cooperación internacional y el establecimiento de estándares comunes. Su carácter flexible permite que las Partes adapten las disposiciones de acuerdo con su situación específica mediante sus legislaciones y políticas nacionales. Se establecen unos principios generales y las Partes los desarrollan posteriormente, de modo que se asegura una mayor coherencia entre todas ellas y se logra una mayor adaptación a los nuevos desafíos que se planteen[152].

Este instrumento, al igual que el RIA, tiene un enfoque de carácter protector y preventivo ya que se basa en la gravedad y probabilidad de que los sistemas de IA afecten negativamente a los derechos humanos, la democracia y el Estado de Derecho, es decir, se refiere a los potenciales riesgos que esta tecnología pudiera tener, aunque no adopta una clasificación de sistemas de IA sobre esta base, como sí hace el RIA[153].

[151] Página web del Consejo de Europa. Disponible en: https://www.coe.int/en/web/conventions/full-list?module=signatures-by-treaty&treatynum=225 . Información consultada por última vez el 16.06.2025.
[152] Cano Linares, M. Á. (2025)., *op. cit.*, p. 42.
[153] *Ibid.*, p. 42.

5. CONCLUSIONES

1. La inteligencia artificial presenta grandes beneficios para la sociedad y debe utilizarse como una herramienta de apoyo al ser humano. Su utilidad es clara en tareas mecánicas tales como la búsqueda de información, recopilación de datos, o la investigación. La automatización permite agilizar muchos procesos y hacerlos más eficientes, especialmente en la gestión administrativa. En muchas ocasiones, su uso puede contribuir a un mejor disfrute de nuestros derechos en ámbitos como la salud, la educación o la justicia. Por lo tanto, esta tecnología debe tener como objetivo último aumentar el bienestar humano y ello implica un desarrollo y uso éticos.

2. Aunque existen beneficios, se ha demostrado que en la mayoría de las tareas se necesita de supervisión humana por los numerosos y graves riesgos que puede entrañar esta tecnología para los derechos humanos. La existencia de sesgos algorítmicos en los sistemas de IA supone un peligro para nuestros derechos y en ocasiones se producen por el empleo de datos sensibles de las personas que, además de afectar a su intimidad, tiene como consecuencia la discriminación en diversos ámbitos al categorizar a las personas en función de características previamente aprendidas. La IA no puede suponer una amenaza a la dignidad del ser humano, por ser esta la base de todos los demás derechos fundamentales, de modo que ha de ser abordada con marcos normativos que cuenten con un enfoque protector.

3. La disparidad en los enfoques nacionales frente a la regulación de la inteligencia artificial evidencia, sin embargo, que la protección de los derechos humanos no suele ser el objetivo prioritario que inspira al regulador nacional. Los intereses de algunos Estados se centran en aspectos como el fomento de la innovación y el libre desarrollo de la tecnología, como ocurre en Estados Unidos, o en adoptar un modelo basado en el control estatal y la vigilancia, como es el caso de China. El miedo a frenar la innovación y por ello perder posición competitiva en el mercado no justifican poner en riesgo y limitar los derechos humanos de las personas que se ven afectadas por las decisiones algorítmicas o por un uso discriminatorio de la IA. En un contexto globalizado en el que se pueden aprovechar todos los

beneficios que tiene la tecnología se requiere un marco regulatorio uniforme a nivel internacional para garantizar que los derechos humanos quedan amparados.

4. En contraparte, el legislador internacional sí que ha comenzado adoptando instrumentos, como son los Principios de la OCDE y la Recomendación ética de la UNESCO sobre inteligencia artificial, que han servido para sentar las bases de la regulación de la IA a nivel mundial mediante unos valores y pautas centrados, principalmente, en un uso ético y fiable de esta tecnología. Se trata, sin embargo, de instrumentos de *soft law*. A pesar de ello, resultan de gran interés debido a su flexibilidad normativa, que permite una mayor adaptabilidad ante la rápida evolución tecnológica y que ha favorecido su continua actualización, como ha ocurrido con los Principios de la OCDE. Estas iniciativas contribuyen a generar estándares comunes que pueden ser revisados con el paso del tiempo, facilitando su adecuación progresiva a los constantes cambios que se producen en el ámbito de la inteligencia artificial.

   En todo caso, no resultan suficientes para regular la problemática del creciente avance de la IA frente a los derechos humanos, siendo necesaria normativa jurídicamente vinculante que inste a los Estados a cumplir con sus obligaciones en esta materia. En este sentido, el Reglamento de Inteligencia Artificial de la Unión Europea y el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, democracia y Estado de Derecho, si bien proponen principios inspirados en estos instrumentos previos como son el respeto a la dignidad humana y a la privacidad, la no discriminación o la transparencia en la toma de decisiones de los sistemas de IA, son ya normas jurídicamente vinculantes, aunque sea a escala regional.

5. En un análisis exhaustivo del Reglamento de la UE y el Convenio Marco del Consejo de Europa se pueden extraer varias similitudes, además de que ambas propuestas han surgido en Europa. En primer lugar, los dos presentan un enfoque basado en el riesgo, primando los principios de prevención y precaución. Ello implica que desde el primer momento se ha de considerar el riesgo que puede tener el sistema de IA y adoptar las medidas pertinentes en función de tal consideración. Este planteamiento es ampliamente desarrollado por el RIA, ya

que no solo establece una clasificación de estos sistemas en atención al riesgo, sino que además esta clasificación determina el tipo de obligaciones a cumplir, que también especifica el Reglamento.

Las dos normas establecen la obligación de llevar a cabo la gestión de riesgos, lo que supone que las Partes deben asegurar que se adopten medidas como las evaluaciones de impacto y se garanticen la transparencia y la supervisión humana durante todo el ciclo de vida de los sistemas de IA.

Por otro lado, ambos instrumentos excluyen de su ámbito de aplicación los sistemas con fines militares, de defensa o de seguridad nacional. La decisión de esta exclusión responde a la necesidad de preservar las funciones esenciales de los Estados en esta materia. No obstante, la utilización de los sistemas de inteligencia artificial como armas autónomas con fines militares puede suponer un importante riesgo sobre los derechos humanos, de modo que no pueden quedar desamparados en este contexto por la problemática que suscitan estos sistemas en cuanto a las cuestiones de determinación de responsabilidad. Los dos excluyen, además, la regulación de aquellos sistemas que tengan como única finalidad la investigación. Ello pone de manifiesto la intención de no obstaculizar la innovación y el desarrollo tecnológico, aspectos que podrían resultar especialmente sensibles para los Estados.

6. Sin embargo, el Reglamento y el Convenio Marco se diferencian en numerosos otros aspectos. Mientras que el RIA tiene una visión más centrada en el mercado y en la IA como producto; el objeto del Convenio es la protección de los Derechos Humanos, la democracia y el Estado de Derecho durante el ciclo de vida de los sistemas de IA.

   La aplicación del RIA se extiende a los miembros de la Unión Europea y a quienes operen en ella, provocando así un efecto indirecto sobre los terceros, quienes deberán ajustarse a las disposiciones del Reglamento en el desarrollo de sus actividades comerciales con la Unión. Se trata de una norma que es vinculante en todos sus elementos y cuyo efecto directo permite que se generen obligaciones también a particulares. Pese a ello, su entrada en vigor se ha previsto de manera escalonada, dando margen de adaptación en el tiempo a sus destinatarios. La diferencia con el Convenio radica en que este se aplica a Estados, lo que supone que las Partes lo ratifiquen para su entrada en vigor y es flexible en el modo en

que estas lo adapten a sus marcos nacionales, ya que establece una serie de principios y obligaciones generales. La característica más destacable es su vocación universal ya que no solo se aplica a los Estados miembros del Consejo de Europa, sino que prevé la adhesión de terceros Estados.

En atención a esta diferente naturaleza, en el Convenio se produce una distinción en el ámbito de aplicación entre el sector público y el sector privado que en el RIA no se da, ya que las obligaciones del mismo afectan por igual independientemente de quién lleva a cabo las actividades.

Otra diferencia se encuentra en la forma de garantizar el cumplimiento de las obligaciones que imponen estas normas a las Partes. Por un lado, el Convenio Marco propone realizar un seguimiento de las prácticas de los Estados mediante la información periódica a la Conferencia de las Partes sobre cómo se han abordado los riesgos en las actividades desarrolladas durante el ciclo de vida de los sistemas de IA. En cambio, el RIA establece un estricto sistema de sanciones para aquellos responsables que infrinjan las disposiciones del Reglamento. De este modo, el Convenio es más flexible a la hora de dar margen a los Estados para que designen mecanismos nacionales de supervisión y con posterioridad informen debidamente, mientras que el Reglamento opta por imponer multas a cualquiera de sus destinatarios si no se da cumplimiento a lo previsto en la norma.

7. La adopción del Reglamento de la UE y del Convenio Marco del Consejo de Europa suponen un gran avance en cuanto a la protección de derechos humanos en el ámbito de la inteligencia artificial. Su enfoque basado en el riesgo permite prevenir y gestionar las situaciones que puedan poner en peligro los derechos de las personas en este contexto. Su carácter vinculante va a lograr una mayor eficacia en la aplicación en los Estados y una mayor armonización a nivel mundial de la regulación de la IA, especialmente en el caso del Convenio por su vocación universal. Aunque ambos instrumentos se complementan, el Reglamento se enfoca en una regulación más técnica; sin embargo, el Convenio hace especial hincapié en la protección de los derechos humanos al establecer específicamente los principios que todos los sistemas de IA han de respetar, lo que refuerza su potencial como marco de referencia ético para garantizar una IA centrada en el ser humano.

8. En definitiva, la inteligencia artificial es una tecnología en constante evolución y cuyo desarrollo futuro es incierto, de modo que plantea un importante desafío para la protección de los derechos humanos. Un desarrollo y uso éticos y controlados de la IA pueden contribuir a mejorar el bienestar humano, pero todos los riesgos que implica para nuestros derechos han de ser abordados. Ello supone que las respuestas normativas deben, necesariamente, incorporar un enfoque basado en los derechos humanos como eje fundamental de cualquier regulación sobre inteligencia artificial, dado que se trata de una tecnología creada para el ser humano. La creación de marcos jurídicos vinculantes a nivel internacional es urgente en un contexto en el que los desarrollos en inteligencia artificial están condicionados por prioridades económicas o de control estatal, en detrimento de los derechos fundamentales. Las dos propuestas europeas presentan características adecuadas para convertirse en modelo de referencia, aunque queda por ver la aplicación práctica que se hace de ellos para poder valorar su verdadero potencial de protección y sus lagunas.

BIBLIOGRAFÍA

a) Documentación

Naciones Unidas

- Naciones Unidas. (2018, 29 agosto). *Informe del Relator Especial sobre la promoción y protección del derecho a la libertad de opinión y de expresión.* (Doc. A/73/348).
- UNESCO (2021, 23 noviembre). *Recomendación sobre la ética de la inteligencia artificial*. (Doc. SHS/BIO/REC-AIETHICS-2021).

OCDE

- Organización para la Cooperación y el Desarrollo Económico. (2019, 22 mayo). *Recomendación del Consejo sobre Inteligencia Artificial*. (Doc. OECD/LEGAL/0449).
- Organización para la Cooperación y el Desarrollo Económico. (2024, 24 abril). *Informe sobre la implementación de la Recomendación de la OCDE sobre Inteligencia Artificial*. (Doc. C/MIN (2024)17).

Consejo de Europa

- Consejo de Europa. (2024). *Convenio Marco sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho. (CETS Nº 225)*, de 5 de septiembre de 2024.
- Comité sobre Inteligencia Artificial. (2024). *Metodología para la evaluación del riesgo y el impacto de los sistemas de inteligencia artificial desde el punto de vista de los derechos humanos, la democracia y el Estado de Derecho (HUDERIA Methodology). (CAI(2024)16rev2)*, de 28 de noviembre de 2024.

Unión Europea

- High-Level Expert Group on Artificial Intelligence, European Comission (2019). *A Definition of AI: Main Capabilities and Disciplines*, p.6.

- 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. *Diario Oficial de la Unión Europea, L 1689*, de 12.7.2024.

España

- Ministerio de Asuntos Económicos y Transformación Digital, Gobierno de España (2020). *Estrategia Nacional de Inteligencia Artificial*, p.66.

Legislación y documentación extranjera

- Red Iberoamericana de Protección de Datos (2019, 21 junio). *Recomendaciones Generales para el Tratamiento de Datos en la Inteligencia Artificial*,
- Gobierno de Colombia (2021, mayo). *Marco ético para la inteligencia artificial en Colombia*.
- Senado Federal de Brasil. (2023, 3 mayo). *Proyecto de Ley nº 2338/2023, que dispone sobre el desarrollo, fomento y uso ético y responsable de la inteligencia artificial con base en la centralidad de la persona humana*.

b) Monografías, capítulos de libros y artículos de revistas

- Barrio Andrés, M. (2024). Objeto, ámbito de aplicación y sentido del Reglamento Europeo de Inteligencia Artificial. En M. Barrio Andrés (Dir.), *El Reglamento Europeo de Inteligencia Artificial* (pp. 21-48). Tirant lo Blanch.
- Cano Linares, M. Á. (2025). Inteligencia Artificial y derechos humanos: el Convenio Marco del Consejo de Europa. En M. Á. Cano Linares y A. I. Carreras Presencio (Dirs*.), Nuevas Realidades: un enfoque desde el derecho* (pp. 35-51). Dykinson.
- Carrión Bósquez, N. G., Castelo Rivas, W. P., Alcívar Muñóz, M. M., Quiñonez Cedeño, L. P. y Llambo Jami, H. S. (2022). Influencia de la COVID-19 en el clima laboral de trabajadores de la salud en Ecuador. *Revista Información Científica*, *101* (1).
- Casacuberta, D. y Guersenzvaig, A. (2019). Using Dreyfus´ legacy to understand justice in algorithm-based processes. *AI y Society, 34*(2), 313-319.

- Céspedes Gutiérrez, O. Y. y Carrillo Cruz, Y. A. (2024). Democracia en riesgo: la amenaza a las libertades políticas en la era de la inteligencia artificial. *Justicia, 29* (45), 1-13.
- Conde Fuentes, J. (2024). Sistemas de alto riesgo en el Reglamento Europeo de Inteligencia Artificial: La ponderación de los derechos fundamentales en juego. En T. J. Aliste Santos (Coord.), *Eficiencia Procesal I* (pp. 111-129). Atelier.
- Cotino Hueso, L. (2024). El Convenio sobre inteligencia artificial, derechos humanos, democracia y Estado de Derecho del Consejo de Europa. *Revista Administración y Ciudadanía, EGAP, 18* (2), 1-25.
- Demir Kaymak, Z. T. (2024). Efectos de la preparación de los estudiantes de obstetricia y enfermería sobre la inteligencia artificial médica en la ansiedad por la inteligencia artificial. *Science Direct*, 7.
- Díaz Galán, E. C. (2025). 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, 17* (1), 366-390.
- Duitama Pulido, E. A. (2022). Sesgos algorítmicos e inteligencia artificial en el poder judicial. En A. Padilla-Muñoz (Ed.), *El derecho como laboratorio de saberes: meditaciones sobre epistemología* (pp. 41-79). Editorial Universidad del Rosario.
- Fernández Ramírez, M. (2023). Inteligencia artificial, algoritmos predictivos y gestión tecnológica de la Seguridad Social. En Asociación Española de Salud y Seguridad Social (Coord.), *Las transformaciones de la Seguridad Social ante los retos de la era digital. Por una salud y Seguridad Social digna e inclusiva* (pp. 1-22). Laborum.
- Fernández Sánchez, S. (2021). Frank, el algoritmo consciente de Deliveroo. Comentario a la Sentencia del Tribunal de Bolonia 2949/2020, de 31 de diciembre. *Revista de Trabajo y Seguridad Social, CEF, 457*, 179-193.
- Ferrante, E. (2021). Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos? *Nueva sociedad* (294), 27-36.
- Fitria, T.N. (2021). The use technology based on artificial intelligence in english teaching and learning. *ELT Echo: The Journal of English Language Teaching in Foreign Language Context, 6* (2), 213-223.

- Garriga Domínguez, A. (2024). Los derechos ante los sistemas biométricos que incorporan Inteligencia Artificial. *Derechos y Libertades: Revista de Filosofía del Derecho y Derechos Humanos* (51), 117-149.
- Graf, F., Obrecht, L. y Weiner, S. (2022). Erste Erkenntnisse zu Transparenz, Diskriminierung und Manipulation. Rechtliche Rahmenbedingungen für Künstliche Intelligenz in der Schweiz. *Jusletter, 12*, 1-10.
- Grandi, N. M. (2021). Inteligencia Artificial, algoritmos y libertad de expresión. *Universitas,* 34, 177-194.
- Guzmán Huayamave, K. V. (2024). IA y ética: regulaciones y políticas para una sociedad digital. En *II Congreso Internacional Multidisciplinario. Nuevas Perspectivas en Ciencia y Tecnología: El Papel Transformador de la Inteligencia Artificial* (pp. 102-111).
- Herhausen, D., Bernritter, S. F., Ngai, E. W. T., Kumar, A., y Delen, D. (2024). Machine learning in marketing: Recent progress and future research directions. *Journal of Business Research,* 170, 1- 2.
- Hernández Ramos, M. (2024). El marco jurídico regulatorio europeo de la inteligencia artificial. La relación de complementariedad entre el Reglamento de la UE y la Convención Marco del Consejo de Europa. *Revista Española de Derecho Europeo, 92*, 10- 41.
- Jaramillo Verduga, M. J. y Alarcón Dalgo, C. M. Á. (2024). Influencia de la Inteligencia Artificial en el Cuidado de Enfermería y su Reto. *Ciencia Latina Revista Científica Multidisciplinar*, *8* (5), 5-11.
- Kriebitz, A. y Lütge, C. (2024). La inteligencia artificial y sus consecuencias para los derechos humanos: una evaluación desde la ética de la empresa. En E. González-Esteban y J. C. Siurana Aparisi (Eds.), *Inteligencia Artificial: concepto, alcance, retos* (pp. 149-185). Tirant Humanidades.
- Laukyte, M. (2024). Reflexión sobre los derechos fundamentales en la nueva Ley de la Inteligencia Artificial. *Derechos y Libertades: Revista de Filosofía del Derecho y Derechos Humanos* (51), 151-175.
- Lluís Iribarren, A. (2024). *Análisis comparativo de la regulación de la inteligencia artificial en la Unión Europea y Estados Unidos*. [Trabajo Fin de Grado, Universidad Autónoma de Barcelona]. Dipòsit digital de documents de la UAB. https://ddd.uab.cat/record/303428 .

- López Martínez, F. y García Peña J. H. (2024). IA y sesgos: una visión alternativa expresada desde la ética y el derecho. *Informática y Derecho. Revista Iberoamericana de Derecho Informático, 1* (15), 109-121.
- López- Tarruella Martínez, A. (2025). Claroscuros del "efecto Bruselas" del Reglamento de Inteligencia Artificial. *Revista Española de Derecho Internacional, 77* (1), 237-245.
- Márquez Muñoz, L. (2023). La Inteligencia Artificial aplicada al ámbito sanitario: retos en el futuro. En F. Pérez Tortosa (Coord.). *La Justicia en la Sociedad 4.0: Nuevos Retos para el Siglo XXI* (pp. 81-101). Colex.
- Martínez García, D. N., Dalgo Flores, V. M., Herrera López, J. L., Analuisa Jiménez, E. I., y Velasco Acurio, E. F. (2019). Avances de la inteligencia artificial en salud. *Dominio de las Ciencias*, *5* (3), p. 609.
- Martynova, E. (2024). Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law: a Commentary. En L. Belli y W. B. Gaspar (Eds.), *AI From the Global Majority: Official Outcome of the UN IGF Data and Artificial Intelligence Governance Coalition* (pp. 109-119). FGV- Direito Rio.
- Megías Quirós, J. J. (2022). Derechos humanos e Inteligencia Artificial. *Dikaiosyne,* 37, 140-163.
- Mendoza Enríquez, O. A. (2021). El derecho de protección de datos personales en los sistemas de inteligencia artificial. *Revista del Instituto de Ciencias Jurídicas de Puebla, México 15,* (48), 191-192.
- Merino Rus, R. (2021). Democracia y derechos humanos en peligro: una advertencia sobre el impacto de la inteligencia artificial y los algoritmos. *Economistas sin Fronteras, 42*, 32-36.
- Miguez Macho, L. y Torres Carlos, M. (2024). Sistemas de IA prohibidos y sistemas de IA de alto riesgo. En M. Barrio Andrés (Dir.). *El Reglamento Europeo de Inteligencia Artificial* (p. 48-86). Tirant lo Blanch.
- Miranzo Díaz, J. (2024). El Reglamento de Inteligencia Artificial de la Unión Europea: regulación de riesgos y sistemas de estandarización. *A&C- Revista de Direito Administrativo y Constitucional, 96*, 43-78.

- Miró Linares, F. (2018). Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots. *Revista de Derecho Penal y Criminología*, 20, 87-130.
- Montesinos García, A. (2024). Inteligencia Artificial en la Justicia con Perspectiva de Género: Amenazas y Oportunidades, *Actualidad Jurídica Iberoamericana*, 21, 566-597.
- Morandín-Ahuerma, F. (2023). Principios normativos para una ética de la Inteligencia Artificial. *Concytep,* 1, 86-102.
- Muñoz Rodríguez, A. B. (2020). El impacto de la inteligencia artificial en el proceso penal. *Anuario de la Facultad de Derecho*, 36, 695-728.
- Nguyen A., Ngo H. N., Hong, Y., Dang, B. y Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies, 28* (4), 4222-4241.
- Ortiz de Zárate Alcarazo, L. (2023). Sesgos de género en la inteligencia artificial. *Revista de Occidente*, (502), 5-20.
- Pérez-Ugena, M. (2024). Análisis comparado de los distintos enfoques regulatorios de la inteligencia artificial en la Unión Europea, EE. UU, China e Iberoamérica. *Anuario Iberoamericano de Justicia Constitucional*, *28* (1), 146-148.
- Pérez- Ugena Coromina, M. (2024). Sesgo de género (en IA). *Eunomía. Revista en Cultura de la Legalidad*, 26, 311-330.
- Pimentel Cavalcante, A. (2023). Regulación brasileña de la inteligencia artificial. *Revista de la Facultad de Derecho de México, 73* (287), 5-28.
- Presno Linera, M. A. y Meuwese, A. (2024). La regulación de la Inteligencia Artificial en Europa. *Teoría y Realidad Constitucional, 54*, 131-161.
- Reyes Vásquez, P. A (2023). Ética de la Inteligencia Artificial. Recomendación de la UNESCO, noviembre 2021. *Compendium, 26* (50), 1-6.
- Roa Avella, M. P., Sanabria- Moyano, J. E. y Dinas-Hurtado, K. (2022). Uso del algoritmo COMPAS en el proceso penal y los riesgos a los derechos humanos. *Revista Brasileira de Direito Processual Penal*, *8* (1), 275-308.
- Rodríguez Torres, Á. F., Rodríguez Alvear, F. S., Collaguazo Lapo, D. R., y Rodríguez Alvear, J. C. (2024). Diferencias y Aplicaciones de Big Data,

Inteligencia Artificial, Machine Learning y Deep Learning. *Dominio de las Ciencias*, 10 (3), 963-966.

- Romero García-Aranda, C. (2024). La Inteligencia Artificial en la Educación: una visión de la UNESCO. *Dignitas: Revista On-line sobre Derechos Humanos y Relaciones Internacionales,* (7), pp. 14-40.
- Rosa Castillo, A. y Mont Verdaguer, M. (2024). Inteligencia artificial, deepfakes y nueva explotación del cuerpo femenino: implicaciones éticas y sociales. En T. Áranguez y O. Olariu (Coords.), *Los derechos de las mujeres en la sociedad digital* (pp. 153-165). Dykinson.
- San Miguel Caso, C. (2021). La aplicación de la Inteligencia Artificial en el proceso: ¿un nuevo reto para las garantías procesales? *Ius et Scientia, 7* (1), 286-303.
- San Miguel Caso, C. (2023). Inteligencia artificial y algoritmos: la controvertida evolución de la tutela judicial efectiva en el proceso penal. *Estudios Penales y Criminológicos, 44*, 1-23.
- Sánchez Martínez, M. O. (2022). El impacto de la sociedad digital en los derechos humanos. En J. I. Solar Cayón y M. O. Sánchez Martínez (Dirs.), *El impacto de la inteligencia artificial en la teoría y la práctica jurídica* (pp. 115-137). La Ley.
- Sánchez Rosado, J. C., Díez Parra, M. (2022). Impacto de la Inteligencia Artificial en la transformación de la sanidad: beneficios y retos. *Economía Industrial* (423), 129-144.
- Solar Cayón, J. I. (2022). Inteligencia artificial y justicia digital. En F. H. Llano Alonso (Dir.), *Inteligencia artificial y Filosofía del derecho* (pp. 381- 427). Laborum.
- Solar Cayón J. I. (2022). ¿Jueces-robot? Bases para una reflexión realista sobre la aplicación de la inteligencia artificial en la administración de justicia. En J. I. Solar Cayón y M. O. Sánchez Martínez (Dirs.), *El impacto de la inteligencia artificial en la teoría y la práctica jurídica.* La Ley, 245-276.
- Solar Cayón, J. I. (2025). *Inteligencia artificial jurídica e imperio de la ley*. Editorial Tirant Lo Blanch.
- Sourdin, T. (2018). Judge v. Robot? Artificial Intelligence and judicial decision-making. *UNSW Law Journal, 41* (4), p. 1115.

- Uriol, L. M. (2024). Análisis de las recomendaciones de ONU y OCDE sobre la regulación de la Inteligencia Artificial. El estado de situación en Argentina. *Aequitas, 16* (16), 1-19.
- Vázquez Pita, E. (2021). La UNESCO y la gobernanza de la inteligencia artificial en un mundo globalizado. La necesidad de una nueva arquitectura legal. *Anuario de la Facultad de Derecho. Universidad de Extremadura 37*, 273-302.

c) Otros recursos

Páginas web

- Consejo de Europa. Página web. Disponible en: https://www.coe.int/en/web/artificial-intelligence/home .

Prensa en línea, blogs especializados y otros artículos de opinión

- Angulo, S. (2018, 7 febrero). *4 ejemplos de inteligencia artificial para la medicina*. Apps y Software, Enter.co. Disponible en: https://www.enter.co/chips-bits/apps-software/ejemplos-inteligencia-artificial-salud/ .
- Efeminista (2023, 16 mayo). *`Sara´, la chatbot española que asesora a víctimas de violencia machista en el Caribe.* Disponible en: https://efeminista.com/sara-chatbot-espanola-violencia-machista-caribe/ .
- Martín, P. (2023, 23 septiembre). España usará la IA y el "big data" para proteger mejor a las víctimas del machismo. *Diario el Periódico*. Disponible en: https://www.elperiodico.com/es/sociedad/20230923/violencia-genero-machista-inteligencia-artificial-proteccion-big-data-victimas-92338576 .
- Milmo, D. (2025, 11 febrero). EEUU acusa a Europa de "regulación excesiva" y se niega a firmar la declaración de la cumbre de París sobre IA "inclusiva". *ElDiario.es*. Disponible en: https://www.eldiario.es/internacional/eeuu-rechaza-firmar-declaracion-cumbre-paris-ia-inclusiva-hablar-excesiva-regulacion-europa_1_12044385.html .
- Recla, E. y Zurita, C. (2024, 2 julio). *Diferencias y similitudes de la legislación de IA en UE, EEUU y China*. Computing. Disponible en: https://www.computing.es/opinion/legislacion-ia-en-el-mundo-eeuu-china-ue/ .

- Volpicelli, G. (2024, 11 marzo). International AI rights treaty hangs by a thread. *Politico*. Disponible en: https://www.politico.eu/article/council-europe-make-mockery-international-ai-rights-treaty/ .
- Williams, K. (2023, 7 noviembre). *Summary: What Does Biden´s Executive Order on Artificial Intelligence Actually Say?* Electronic Privacy Information Center. Disponible en: https://epic.org/summary-what-does-bidens-executive-order-on-artificial-intelligence-actually-say/ .