FACULTAD DE DERECHO

LA ÉTICA JURÍDICA EN LA ERA DE LA DE LA INTELIGENCIA ARTIFICIAL: DESAFÍOS Y REFLEXIONES FILOSÓFICAS

Autor: Joaquín Gil-Delgado Madrigal

5ºE-3A

Filosofía del Derecho

Tutor: Alberto de Unzurrunzaga Rubio

Madrid Abril, 2024

RESUMEN

Este Trabajo de Fin de Grado explora la intersección de la ética jurídica y la inteligencia artificial (IA), abordando su evolución, desafíos legales y filosóficos actuales y futuros. Analiza cómo la IA está remodelando el panorama legal y los dilemas éticos que surgen, ejemplificados en la toma de decisiones legales automatizadas y la protección de derechos humanos como la privacidad y la no discriminación. El estudio analiza la normativa vigente y se reflexiona en cómo la nueva normativa de IA de la Unión Europea puede liderar la creación de un marco legal ético que proteja los derechos de los consumidores sin comprometer la innovación.

Palabras clave: Inteligencia Artificial, Ética Jurídica, Filosofía del Derecho, Algoritmo, Privacidad, Unión Europea

ABSTRACT

This paper explores the intersection of legal ethics and artificial intelligence (AI), addressing its evolution, current and future legal and philosophical challenges. It examines how AI is reshaping the legal landscape and the ethical dilemmas that arise, exemplified in automated legal decision-making and the protection of human rights such as privacy and non-discrimination. The study analyzes existing legislation and reflects on how the EU's new AI regulations can lead the way in creating an ethical legal framework that protects consumer rights without compromising innovation.

Key words: Artificial Intelligence, Legal Ethics, Legal Philosophy, Algorithm, Privacy, European Union.

SIGLAS Y ACRÓNIMOS

IA	Inteligencia Artificial
ML	Machine Learning
PLN	Procesamiento de Lenguaje Natural
RGPD	Reglamento General de Protección de Datos
AESIA	Agencia Española de Supervisión de la Inteligencia Artificial
AEPD	Agencia Española de Protección de Datos
ENIA	Estrategia Nacional de Inteligencia Artificial
LOPDGDD	Ley Orgánica de Protección de Datos Personales y garantía de los derechos digitales
AI HLEG	Grupo de Alto Nivel sobre Inteligencia Artificial
OCDE	Organización para la Cooperación y el Desarrollo Económicos
GPAI	Global Partnership on Artificial Intelligence
UNICRI	Centro de Inteligencia Artificial y Robótica del Instituto Interregional de las Naciones Unidas para la Investigación del Delito y la Justicia
EIPD	Evaluación de Impacto de Privacidad
CCIA	Computer and Communications Industry Association

ÍNDICE

1. INTRODUCCIÓN

1.1 Contextualización y justificación de la investigación.

Desde el monstruo de Frankenstein creado por Mary Shelley, la humanidad ha albergado la fascinación por crear entidades mecánicas dotadas de inteligencia, particularmente androides que emulan aspectos humanos. No obstante, aquello que parecían elucubraciones de una novela de ciencia ficción, hoy en día se ha convertido en una realidad.

Así, la irrupción de la Inteligencia Artificial (IA) en nuestra sociedad ha marcado un hito sin precedentes en lo que a la evolución tecnológica se refiere. No solo ha transformado y facilitado la realización de tareas cotidianas, sino que ahora ha dado un paso más al realizar acciones que van más allá de la mera automatización y que hace unos años serían inimaginables. La IA también se ha entrometido en el ámbito en el que los seres humanos se relacionan, al cambiar la comunicación y la forma en que interactuamos.

En este contexto, se ha producido una brecha entre la tecnología y el derecho, ya que la rapidez con la que ha evolucionado la ciencia ha superado la capacidad de las legislaciones existentes para abordar los nuevos desafíos planteados, provocando, en ocasiones, lagunas legales y afectando la seguridad jurídica. De hecho, a día de hoy sigue sin existir una ley global que regule este tipo de inteligencias; sin embargo, en el ámbito comunitario, la Unión Europea está avanzando con la Ley de Inteligencia Artificial, considerada la primera legislación integral sobre IA del mundo.

El presente estudio se justifica en la medida en que la intromisión de la IA en la cotidianidad de los seres humanos no solo replantea los límites y alcances de la tecnología, sino que también cuestiona los fundamentos éticos y morales de nuestras leyes. A medida que la IA se vuelve más compleja, surge el interrogante sobre su rol en la toma de decisiones y las repercusiones éticas de sus acciones, lo que pone de manifiesto la insuficiencia de los marcos jurídicos para garantizar que la tecnología actúe de manera justa y equilibrada.

Además, se tratarán temas como la privacidad, la autonomía, la transparencia y la discriminación, con el objetivo de sentar las bases de una ética jurídica que responda adecuadamente a los desafíos planteados.

Por último, este trabajo pretende contribuir al debate sobre cómo los desarrollos tecnológicos, especialmente aquellos relacionados con la IA, están transformando las profesiones jurídicas. Se analizará cómo la inteligencia artificial puede tanto debilitar como fortalecer los fundamentos de la ética y se expondrán caminos para que la tecnología y el derecho evolucionen de manera armónica, asegurando que la IA sirva al bienestar humano y no al contrario.

1.2 Objetivos

El objetivo principal de este Trabajo de fin de Grado es analizar el impacto de la inteligencia artificial en nuestra sociedad desde la perspectiva de la ética jurídica.

Objetivos secundarios:

- Definir de manera precisa y completa el concepto de inteligencia artificial y ética jurídica, exponiendo las diferentes tipologías y principios que los rigen y haciendo hincapié en su relevancia en nuestra sociedad actual.
- Explorar las consecuencias jurídicas de la inteligencia artificial mediante el análisis de la legislación y normativa vigente, con un enfoque particular en la nueva legislación que la Unión Europea ha propuesto.
- Analizar estudios de casos específicos donde la inteligencia artificial ha desempeñado un papel significativo en la toma de decisiones jurídicas, evaluando sus consecuencias éticas y legales.
- Ofrecer recomendaciones y propuestas para abordar los desafíos éticos identificados, incluyendo posibles regulaciones, consideraciones éticas en el diseño de algoritmos y políticas futuras.

1.3 Metodología

En este trabajo se emplea una metodología cualitativa para indagar sobre la ética jurídica en el contexto de la inteligencia artificial. Para lograrlo, se ha realizado una rigurosa revisión de literatura académica y legal, accediendo a bases de datos reconocidas, publicaciones especializadas y recursos digitales, así como al análisis de legislaciones nacionales e internacionales. Esta metodología permite englobar múltiples aspectos del tema y desarrollar un análisis crítico. Así, pretendemos descubrir cómo la IA incide en la sociedad y evaluar sus retos ético-jurídicos para proponer recomendaciones informadas.

1.4 Estructura de trabajo

Este trabajo de fin de grado se puede dividir en ocho partes. El primer capítulo sirve como introducción, estableciendo el escenario de estudio, explicando la relevancia del tema en la actualidad y la necesidad de la investigación. Se detallarán los objetivos del trabajo y la metodología utilizada para abordar la investigación, proporcionando también un esquema general de la estructura completa del trabajo.

En el segundo capítulo se establecerá el marco teórico, abarcando desde los conceptos básicos de la inteligencia artificial, incluyendo su evolución y los diferentes tipos existentes, hasta los principios de la ética jurídica. Se explorará la relación entre ética y derecho y cómo la IA está reconfigurando esta dinámica.

En el tercer capítulo, se analizará la situación actual de la legislación respecto a la IA, identificando los desafíos éticos y legales que plantea su integración en la sociedad y cómo las brechas en la regulación afectan la seguridad jurídica.

En el capítulo cuarto, se comentarán casos concretos donde la IA ha tenido un impacto significativo en el ámbito legal, desde decisiones de libertad condicional hasta la predicción de criminalidad y el uso de chatbots jurídicos, reflejando las implicaciones éticas y legales en cada escenario.

En el quinto, se evaluará el impacto de la IA en los derechos humanos, abordando temas como la privacidad, los posibles sesgos y discriminación, así cómo cuestiones relativas a la transparencia y rendición de cuentas.

En el capítulo sexto, se comentará la propuesta legislativa europea en desarrollo, la conocida Ley de Inteligencia Artificial, analizando sus aspectos más destacados y se presentará un análisis de las críticas principales que ha suscitado.

Por último, en el capítulo final se sintetizará los hallazgos clave del estudio, ofreciendo recomendaciones, limitaciones del estudio y futuras líneas de investigación.

2. MARCO TEÓRICO

2.1 Inteligencia artificial: concepto y evolución histórica.

2.1.1 Evolución histórica

La década de 1940 fue una época marcada por la Segunda Guerra Mundial, durante la cual surgieron múltiples ideas e inventos sobre cómo las máquinas podrían facilitarnos la vida, no solo en el campo de batalla, sino también en el trabajo, agilizando los procesos productivos. De hecho, fue años antes, en 1936, cuando el científico inglés Alan Turing marcó el inicio de la informática teórica con la publicación de su artículo *Números Calculables*, introduciendo el concepto de Máquina de Turing. Esta era una entidad matemática abstracta que formalizó la noción de algoritmo y resultó ser la precursora de las computadoras digitales, permitiendo comprender los límites y las posibilidades de lo que las computadoras pueden y no pueden hacer.

En la década de los 50, los matemáticos Norbert Wiener y John von Neumann, tenían como objetivo unificar la teoría matemática, la electrónica y la automatización en una única teoría. Así consiguieron realizar contribuciones fundamentales que condujeron al desarrollo de lo que hoy en día conocemos como inteligencia artificial. Su enfoque en la teoría de sistemas, la cibernética y la teoría de juegos sentaron las bases de conocimiento

necesarias para que futuros científicos pudieran desarrollar tecnologías más avanzadas en este campo (Wiener, 1950; Burks, 1984).

Sin embargo, no fue hasta 1956, durante la conferencia de Dartmouth en Estados Unidos, donde surgió por primera vez el término inteligencia artificial. En este evento, investigadores como Claude Shannon, Marvin Minsky y John McCarthy participaron y discutieron sobre el desarrollo de programas informáticos con la capacidad de pensar y aprender. Fue McCarthy (1956), quien la definió como *la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes*, estableciendo así el inicio de la IA como disciplina.

En los años 60 y 70, se experimentó con modelos de percepción simbólica y se produjeron avances en redes neuronales. Sin embargo, en 1980 se produce una época conocida como el *Invierno en la IA* que ocasionó una reducción de fondos e interés en la investigación de las inteligencias artificiales, provocado por: expectativas poco realistas generadas por los desarrolladores, altas expectativas de los usuarios finales y una extensa promoción en los medios (Crevier, 1993).

En la década de los 90 surge el *Deep Blue* de IBM que derrota al campeón mundial de ajedrez Garry Kasparov, un hito significativo que demuestra la capacidad de la IA en juegos de estrategia. Deep Blue consiguió que una máquina con suficiente capacidad de cálculo, podría no solo competir sino también superar a un humano en tareas complejas como el ajedrez. Para ello, utilizó una combinación de fuerza bruta (evaluando 200 millones de posiciones por segundo) y algoritmos sofisticados. La IA pasó a ser vista como algo que podría tener aplicaciones prácticas reales y no solo como un tema de investigación, lo que provocó un debate filosófico sobre la naturaleza de la inteligencia humana y la creatividad frente a la inteligencia artificial (Campbell, Hoane, & Hsu, 2002).

Es a partir del siglo XXI cuando surge la época más destacada de las inteligencias artificiales, destacando las IA generativas, como Chat-GPT. Esta tecnología ha revolucionado la manera en que interactuamos con las máquinas, mostrando una capacidad sin precedentes para crear textos, códigos y contenido creativo de manera autónoma, sin apenas intervención humana. Esta autonomía se extiende más allá del ámbito digital, con robots que ahora asumen tareas de gran complejidad, desde procesos

de producción avanzada hasta procedimientos de cirugía asistida, marcando un hito en sectores industriales y médicos.

La influencia de la IA también se ha expandido a múltiples disciplinas científicas, impulsando descubrimientos y eficiencias en áreas tan diversas como la bioinformática, facilitando el análisis de grandes conjuntos de datos genéticos, y la astrofísica, así como en la química computacional. Sin embargo, como se describe en el Plan de Recuperación, Transformación y Resiliencia del Gobierno de España (2023) esta integración extensiva de la IA en facetas clave de nuestra vida ha intensificado la necesidad de enfrentar los dilemas éticos inherentes, los sesgos algorítmicos y su impacto social. Este escenario ha provocado un debate necesario sobre la implementación de regulaciones y el desarrollo de marcos éticos sólidos que guíen el progreso responsable de la inteligencia artificial.

2.1.2 Concepto de Inteligencia Artificial

En este apartado, vamos a tratar de definir el concepto de la inteligencia artificial desde la visión de diferentes autores, capturando las diversas perspectivas y enfoques que se han tomado en la investigación de la IA a lo largo de los años.

En primer lugar, como hemos mencionado anteriormente, cabe destacar una definición ampliamente citada como es la de John McCarthy (2006), para muchos el padre de esta ciencia y que la define como *la ciencia e ingenio de hacer máquinas inteligentes, especialmente programas de cómputo inteligentes*. Esta visión permite una gran variedad de enfoques, desde la programación hasta la robótica, todo bajo el concepto de la IA.

Por otro lado, debemos destacar la definición de los científicos Stuart Russell y Peter Norvig (2020) que describen la IA como *el estudio de agentes que reciben percepciones del ambiente y realizan acciones*. Además, plantean cuatro enfoques que han caracterizado históricamente el campo: pensar humanamente, pensar racionalmente, actuar humanamente y actuar racionalmente. Ambos autores ofrecen una definición funcional que se centra en la capacidad de un agente para operar de manera autónoma al recibir percepciones y realizar acciones. Su clasificación de los enfoques de la IA destaca la diferencia entre simular el proceso de pensamiento y el comportamiento humano frente a alcanzar la racionalidad ideal, lo que refleja una dualidad entre la imitación de la

inteligencia natural y la creación de una inteligencia que aspire a superar las capacidades humanas.

También se debe mencionar a Nils J. Nilsson (1998) que ha ofrecido una visión amplia, considerando la IA como *la actividad dedicada a hacer máquinas inteligentes*, y señala que es una ciencia de la computación que se ocupa de los métodos de cálculo que permiten percibir, razonar y actuar. Nilsson pone énfasis en la IA como una actividad práctica, su definición refleja la ambición de la tecnología de ir más allá de la teoría pura y entrar en el territorio de la creación y la implementación de sistemas que pueden realizar tareas complejas en el mundo real.

La perspectiva de Marvin Minsky (1961) se centra en los aspectos cognitivos de la IA, resaltando actividades que tradicionalmente se han considerado de alto nivel y específicamente humanas. Este la describió como *la construcción de programas de computadora que se dedican a tareas que son, por el momento, realizadas de una manera más satisfactoria por los seres humanos porque requieren procesos mentales de alto nivel como: el aprendizaje perceptual, la organización de la memoria y el razonamiento crítico.* Su definición es importante al enfatizar la importancia de procesos mentales complejos como el aprendizaje, la memoria y el razonamiento crítico. Además, al señalar que estas tareas se realizan *por el momento* de manera más satisfactoria por los seres humanos, Minsky implícitamente sugiere que la IA tiene el potencial de igualar o superar la capacidad humana en estas áreas.

Por último, podemos destacar una definición más actual como la que proporciona el Grupo de Alto Nivel en Inteligencia Artificial (2018), creado por la Comisión Europea para desarrollar la Estrategia Europea en IA que la define como los *sistemas que manifiestan un comportamiento inteligente, al ser capaces de analizar el entorno y realizar acciones, con cierto grado de autonomía, con el fin de alcanzar objetivos específicos.*

Estas definiciones reflejan la naturaleza multidisciplinaria de la IA y cómo su comprensión ha evolucionado a lo largo del tiempo. Cada autor ha contribuido al campo no solo con sus definiciones, sino también con su trabajo, ayudando a dar forma a lo que hoy conocemos como inteligencia artificial.

2.1.3 Tipos de IA

Una vez analizada la evolución histórica y el marco conceptual de la inteligencia artificial, vamos a clasificar de forma no muy exhaustiva sus diferentes tipos, siguiendo la clasificación de los anteriormente citados, Stuart Russell y Peter Norvig (2020) y otras categorizaciones en función de la capacidad o potencia con la que operan que se describen en el Plan de Recuperación, Transformación y Resiliencia del Gobierno de España (2023).

1. Clasificación según Stuart Russell y Peter Norvig

- **Sistemas que piensan como humanos:** estos sistemas buscan imitar el procesamiento de información que ocurre en el cerebro humano, tanto en términos de pensamiento como de comportamiento. Utilizando el aprendizaje a partir de la experiencia y el uso de conocimiento previo para informar decisiones futuras. Se incluyen en esta categoría el modelado cognitivo y las redes neuronales.

- **Sistemas que actúan como humanos:** se enfocan en imitar el comportamiento humano en términos de razonamiento y actuación. Estos sistemas se diseñan para interactuar fluidamente con los usuarios, comprendiendo el lenguaje natural y reconocimiento emocional, y adaptándose a una variedad de contextos. Los chatbots avanzados son ejemplos representativos, capaces de sostener interacciones casi indistinguibles de las humanas.

- **Sistemas que piensan racionalmente:** Se centran en el razonamiento lógico y la inferencia. Aquí se ubican la lógica formal y los sistemas expertos que utilizan conocimiento estructurado para llegar a conclusiones o realizar predicciones sin considerar necesariamente el comportamiento humano. Aquí podríamos encontrar un sistema de diagnóstico médico basado en IA que utiliza conocimientos de medicina y lógica para diagnosticar enfermedades a partir de síntomas específicos.

- **Sistemas que actúan racionalmente:** También conocidos como agentes racionales, estos sistemas están diseñados para operar de manera óptima al

perseguir un conjunto de objetivos predefinidos, adaptando sus acciones al entorno para lograr sus objetivos. Un ejemplo podrían ser los robots de almacén automatizados que navegan para recoger y transportar productos, asegurando la eficiencia y minimizando el tiempo de entrega.

2. Clasificaciones en función de su capacidad o potencia

- **IA Débil**: también llamada IA estrecha. Este tipo de IA está diseñado para realizar trabajos específicos y opera dentro de un conjunto limitado de controles y parámetros. Deben ser programadas para realizar las tareas ya que carece de aprendizaje o adaptación por si mismas. Ejemplos comunes son los asistentes virtuales y los sistemas de recomendación, como Siri de Apple.

- **IA Fuerte:** Una IA fuerte o IA general, tiene la capacidad de comprender y aprender cualquier tarea intelectual que un ser humano pueda. No esta limitada a una única tarea, sino que puede razonar, planificar y tomar decisiones de forma autónoma.

- **IA Superinteligente:** este tipo de IA superaría la inteligencia humana en todos los ámbitos (creatividad, toma de decisiones, razonamiento…). No obstante, esta inteligencia aún no existe y se encuentra en el ámbito teórico y especulativo.

2.2 Ética jurídica

2.2.1 Conceptos fundamentales

La Ética Jurídica es el conjunto de principios y reglas morales que regulan el comportamiento y las relaciones humanas. Como acertadamente la define el profesor Torres-Dulce Lifante, sería *aquella parte de la Ética referida al justo cumplimiento de los deberes de justicia y al recto ejercicio de los derechos*, es decir, la ética jurídica se basa en reflexionar sobre los actos jurídicos y procurar su concordancia con la justicia (Universidad Complutense Madrid, s.f.).

La concepción de justicia ha evolucionado significativamente desde sus orígenes filosóficos, con figuras clave como Platón, con su *República* o Aristóteles con *Ética Nicomáquea*, que establecieron las bases éticas que persisten en la actualidad. En la era de la inteligencia artificial, estos principios éticos fundamentales adquieren nuevas dimensiones. Por ejemplo, al aplicar la justicia a través de sistemas de IA, nos enfrentamos al desafío de cómo estos sistemas pueden incorporar los valores éticos que sostienen la justicia como la integridad y la imparcialidad. Mientras que filósofos como Rawls, en su obra *Una teoría de la justicia*, reflexiona sobre teorías de justicia centradas en la equidad y la libertad, la IA nos obliga a considerar cómo estos conceptos pueden ser programados y garantizados en algoritmos que toman decisiones que afectan a vidas humanas. Con la introducción de la inteligencia artificial, emergen preguntas éticas adicionales. ¿Cómo se pueden alinear los sistemas de IA con los principios éticos del derecho? ¿Quién es responsable de las decisiones tomadas por un sistema autónomo? ¿Cómo se garantiza la transparencia y la rendición de cuentas en los procesos automatizados? Por todo esto, podemos afirmar que estos sistemas desafían nuestras concepciones tradicionales de pensamiento, por lo que se deberá de exigir una revisión de las teorías éticas existentes y posiblemente la creación de nuevas teorías que puedan abordar la especificidad de la inteligencia artificial en la toma de decisiones legales.

Cuando aplicamos estos conceptos en la práctica, una de las primeras tentativas para establecer principios éticos que sirvan de fundamento a la inteligencia artificial se realizó con las conocidas *Tres Leyes de la Robótica*. Estas fueron ideadas por Isaac Asimov, un autor de ciencia ficción. Las leyes constituyen un grupo de normas teóricas que buscan dirigir la conducta ética de los robots. El objetivo es garantizar una coexistencia segura entre robots y humanos. Las leyes son las siguientes: (1) Un robot no hará daño a un ser humano ni permitirá que, por inacción, este sufra daño; (2) Un robot obedecerá las órdenes que reciba de un ser humano, a no ser que las órdenes entren en conflicto con la primera ley; (3) Un robot protegerá su propia existencia en la medida en que dicha protección no entre en conflicto con las leyes primera y segunda.

Mientras que las Tres Leyes de Asimov sirven como un punto de partida conceptual interesante, la regulación de la IA requiere un enfoque más detallado y flexible para abordar tanto los desafíos éticos como técnicos de la interacción humano-IA. Así, en la *Declaración de robótica del Parlamento de la UE de 2017*, en su punto número 13 de

principios éticos, se redactan los cuatro grandes principios éticos que van a gobernar a la IA, estos son:

- **Beneficencia**: las IA deben diseñarse y utilizarse de manera que sus acciones resulten beneficiosas para el conjunto de la sociedad. Esto incluye la promoción del bienestar humano, la mejora de la efectividad y la eficiencia de tareas y procesos, así como contribuir al progreso y al beneficio social. La IA debería estar diseñada para salvaguardar el funcionamiento de las instituciones democráticas y en sostener los principios fundamentales del Estado de Derecho.
- **Principio de no perjuicio o maleficencia**: la IA debe ser desarrollada y gestionada con el objetivo de evitar causar daño a los individuos o a la sociedad. Esto implica no solo evitar daños físicos directos, sino también proteger la privacidad de los datos personales, evitar sesgos que puedan llevar a discriminación, y prevenir el mal uso o abuso de la tecnología.
- **Autonomía**: se refiere al respeto por la autonomía de los seres humanos en relación con las IA. El Grupo Europeo de Ética en Ciencia y Nuevas Tecnologías (2018) menciona: *que la autonomía como derecho a ser libre solo puede atribuirse a los seres humanos, por lo que no cabe atribuir autonomía a los sistemas por inteligentes que sean*. Esto implica que las personas deben tener control sobre si quieren o no interactuar con sistemas de IA y cómo desean que estos sistemas influyan en sus decisiones y vidas. Esta tecnología no debería coartar la libertad de elección de los usuarios, sino empoderarlos y soportar sus decisiones independientes.
- **Justicia**: este principio está vinculado con el principio de beneficencia, se pretende utilizar la IA para garantizar una distribución justa de los recursos, lo que implica superar desigualdades, evitar sesgos y estigmatizaciones y prevenir la aparición de nuevas formas de exclusión social.
- **Transparencia**: este principio, mencionado en el punto 12 de La Declaración, implica que las decisiones y funcionamientos de la inteligencia artificial deberían ser transparentes y comprensibles para los usuarios y otras partes interesadas. Esto significa que las personas deben ser capaces de entender cómo las IA llegan a sus conclusiones o recomendaciones y en qué datos o lógica se basan estas decisiones. La transparencia es crucial para construir confianza y para permitir la rendición de cuentas en caso de errores o problemas.

La ética en la inteligencia artificial es un campo tan importante como su desarrollo técnico. Estos principios éticos son cruciales para garantizar que la evolución de la IA sea responsable y esté en concordancia con los valores humanos y sociales. La correcta implementación de estos fundamentos es esencial para propiciar una tecnología que sea avanzada en términos técnicos y a la vez justa y consciente de su impacto ético y social. En esta línea, las palabras de Margarethe Vestager (2020), vicepresidenta de la Comisión Europea, enfatizó un punto crítico: *La IA no es buena ni mala en sí misma, todo depende de por qué y cómo es utilizada.* Esto nos recuerda que la tecnología es un reflejo de los valores y objetivos de quienes la crean y la implementan. Por lo tanto, la verdadera medida del éxito en el desarrollo de la IA no reside sólo en sus capacidades técnicas, sino también en su alineación con principios éticos sólidos que guíen su uso hacia el mejoramiento de la sociedad.

3. LA INTELIGENCIA ARTIFICIAL EN EL MARCO LEGAL ACTUAL

Como hemos observado, la inteligencia artificial y otras tecnologías análogas están adquiriendo una presencia creciente en nuestra sociedad. Este fenómeno ha motivado a gobiernos y entidades tanto nacionales como internacionales a elaborar una normativa cada vez más sólida, con el propósito de mitigar la brecha entre el marco legal y el avance tecnológico. Al mismo tiempo, se busca promover un ecosistema de inteligencia artificial que proteja los derechos fundamentales y aporte beneficios tangibles a los ciudadanos, al desarrollo empresarial y a la mejora de los servicios públicos. Así, en este apartado, abordaremos el marco normativo de la IA existente en nuestro país, a nivel europeo y otras regulaciones internacionales.

3.1 Legislación a nivel nacional

En España, la regulación de la inteligencia artificial se encuentra en desarrollo. Actualmente, no existe una legislación específica sobre IA, pero hay leyes y otro tipo de

regulaciones e iniciativas que pueden aplicarse a aspectos específicos de esta tecnología. Entre las más relevantes destacamos:

- **Carta de Derechos Digitales:** es un documento que establece los derechos para proteger a los ciudadanos en el ámbito digital. Fue aprobada en 2021 y, aunque no tiene carácter normativo, pretende asegurar que los derechos existentes en el mundo analógico se mantengan en el entorno digital, estableciendo principios para garantizar la privacidad, la seguridad en internet y promover la educación digital, entre otros aspectos. Así, el objetivo de la Carta no es *descubrir nuevos derechos fundamentales sino concretar los más relevantes en el entorno y los espacios digitales.*

- **Agencia Española de Supervisión de la Inteligencia Artificial (AESIA):** es una entidad autónoma que pertenece al Ministerio de Asuntos Económicos y Transformación Digital de España creada en agosto de 2023 y encargada de la supervisión, el asesoramiento, la sensibilización y la formación en cuanto al uso adecuado y el desarrollo de la IA. Se creó para garantizar que se cumpla con la normativa tanto nacional como europea en relación con los sistemas de IA y, en particular, sus algoritmos. La AESIA tiene la capacidad de imponer sanciones significativas si se detecta un incumplimiento de las regulaciones, con multas que pueden alcanzar hasta los 30 millones de euros o, en el caso de las empresas, hasta el 6% de su volumen de negocio total anual mundial del ejercicio financiero anterior, dependiendo de cuál sea la cantidad mayor. En su diseño y funcionamiento, la AESIA toma como modelo algunas de las funciones de la Agencia Española de Protección de Datos (AEPD), pero con un enfoque específico en la IA. La AESIA no es una autoridad administrativa independiente; su Consejo Rector tiene una fuerte dependencia del ejecutivo estatal, y su presidenta, la Secretaria de Estado de Digitalización, es quien propone directamente al director general de la agencia. Con la creación de esta Agencia, España se convierte en el primer país europeo en tener un órgano de estas características.

- **Estrategia Nacional de Inteligencia Artificial (ENIA):** es un marco estratégico destinado a promover el desarrollo y la integración de la inteligencia artificial de

forma inclusiva, sostenible y centrada en los ciudadanos. Forma parte de la Agenda España Digital 2026 y del Plan de Recuperación, Transformación y Resiliencia del país, y busca mejorar la competitividad de España tanto a nivel europeo como internacional. Esta estrategia tiene como objetivos específicos la transformación de España en un líder en la economía del dato, el impulso de la IA como motor de innovación y crecimiento económico, y la preparación del país para las transformaciones socioeconómicas derivadas de esta tecnología. También enfatiza el fortalecimiento de la competitividad mediante la investigación y desarrollo en Tecnologías Habilitadoras Digitales.

- **Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD):** aunque esta ley no regula específicamente la IA, se establece principios y disposiciones que son relevantes para el tratamiento de datos personales en el contexto de la IA. La ley se centra en la protección de datos personales y garantiza los derechos digitales, lo que indirectamente afecta al uso de la IA y los algoritmos en cuanto gestionan o procesan datos personales. La LOPDGDD asegura la transparencia y la seguridad en el tratamiento de datos personales, requisitos que las aplicaciones de IA deben cumplir cuando manejan este tipo de información. Además, promueve los derechos digitales que podrían verse afectados por este tipo de tecnologías, como el derecho a la desconexión digital y la regulación del uso de datos tras la muerte del usuario.

3.2 Legislación a nivel europeo

En Europa, el Parlamento y el Consejo alcanzaron un consenso sobre la primera regulación de inteligencia artificial, originada en la propuesta de la Comisión Europea de 2021. Este reglamento, conocido popularmente como *Ley de IA* (AI Act), instaura un marco normativo pionero a nivel global para el uso de la IA, con el objetivo de regular exhaustivamente esta tecnología en la UE. No obstante, dejando de lado esta normativa que procederemos a analizar más tarde, la Unión Europea ha desarrollado en los últimos años otros reglamentos, programas e iniciativas, entre las que destacamos:

- **Libro Blanco de la Comisión Europea sobre la Inteligencia Artificial:** este es un documento estratégico publicado en febrero de 2020 que propone la creación de un marco normativo de IA que genere *ecosistema de excelencia y confianza* a nivel europeo. En el Libro se tratan temas como el marco regulatorio para sistemas catalogados de alto riesgo, la promoción de la inversión y la innovación, y otras medidas para asegurar que la IA beneficie al público y respete los derechos fundamentales. También se discuten estrategias para aplicar la IA de forma ética y efectiva, centrado en los derechos humanos y los valores europeos, buscando equilibrar la promoción de la innovación tecnológica con la protección de los ciudadanos.

- **El Plan Coordinado de Inteligencia Artificial:** es una estrategia conjunta de los Estados miembros y la Comisión Europea para fomentar el desarrollo y uso de la IA en Europa. Busca optimizar y maximizar el impacto de las inversiones en IA, fomentar la colaboración transfronteriza y evitar la fragmentación en Europa. Para lograrlo, el plan establece cuatro conjuntos de objetivos políticos, apoyados por acciones concretas. También indica un posible mecanismo de financiación y establece un calendario para: crear un entorno favorable para el avance y la integración de la IA en la Unión, hacer de la UE el lugar donde la excelencia prospera desde el laboratorio hasta el mercado, garantizar que la IA beneficie a la ciudadanía, y establecer un liderazgo estratégico en sectores clave.

- **Comunicación de la Comisión Europea al Parlamento Europeo, el Consejo y otros organismos (COM(2019)168):** titulada *Construir Confianza en la Inteligencia Artificial Centrada en el Ser Humano* se enfoca en la creación de un marco de IA ético y confiable. La comunicación, publicada el 8 de abril de 2019, destaca la necesidad de un enfoque que ponga al ser humano en el centro del desarrollo de la IA, reconociendo tanto su potencial para mejorar la sociedad como los desafíos legales y éticos que conlleva su implementación. La estrategia europea en IA y el plan coordinado de la IA subrayan que la confianza es un prerrequisito para un enfoque centrado en el ser humano: la IA no es un fin en sí mismo, sino una herramienta que debe servir a la gente con el objetivo último de aumentar el bienestar humano. La estrategia propone un enfoque triple para impulsar la capacidad tecnológica y la adopción de estas tecnologías en todos los

sectores económicos, prepararse para los cambios socioeconómicos y garantizar un marco ético y jurídico adecuado.

Estas tres iniciativas han sido desarrolladas gracias al trabajo del Grupo de Alto Nivel sobre Inteligencia Artificial (AI HLEG, por sus siglas en inglés) nombrados por la Comisión para asesorar acerca de la estrategia de la UE en materia de IA. Este grupo ha trabajado en estrecha colaboración con la comunidad europea de partes interesadas de inteligencia artificial a través de la llamada AI Alliance.

- **La Alianza Europea de IA (AI Alliance)**: es un evento público abierto que, desde su edición de apertura en 2019, permite la participación tanto en línea como presencial. Esto facilita un foro a nivel alto y de múltiples participantes para discutir y formar el futuro de la inteligencia artificial en Europa a través de un enfoque de múltiples partes interesadas. A día de hoy se han celebrado cuatro asambleas, la última tuvo lugar en noviembre de 2023 y marcó un hito importante en la implementación de la Estrategia Europea de IA.

3.3 Legislación a nivel internacional

Un gran número de países han desarrollado en los últimos años legislaciones encargadas de regular la IA y otras tecnologías que implementen el uso de algoritmos. Aquí enunciaremos las iniciativas diseñadas por órganos internacionales como la ONU, la OCDE o el G7.

- **AI for Good Global Summit:** es una plataforma organizada por la Unión Internacional de Telecomunicaciones de las Naciones Unidas, que busca explorar y promover el uso de la IA para avanzar en los Objetivos de Desarrollo Sostenible. Reuniendo a expertos de diversos sectores incluidos: gobiernos, la industria, la academia y la sociedad civil. Esta cumbre actúa como un foro para el diálogo, la colaboración y la acción, enfocándose en cómo este tipo de tecnologías pueden contribuir a resolver desafíos globales en salud, educación, medio ambiente e igualdad de género, fomentando así un impacto social y económico positivo a nivel mundial.

- **Recomendaciones y Principios Éticos de la OCDE en materia de Inteligencia Artificial:** adoptados en mayo de 2019, constituyen un conjunto de estándares diseñados para ser prácticos y flexibles, capaces de mantener su relevancia a pesar de los rápidos cambios en la tecnología de IA. Aunque no son legalmente obligatorios, tienen la intención de modelar las normativas internacionales y servir como un marco referencial para las legislaciones nacionales. Los principios fueron concebidos por un grupo diverso de 50 expertos en IA, incluyendo representantes gubernamentales, líderes empresariales, y figuras de la sociedad civil, academia y la comunidad científica. Las recomendaciones de la OCDE destacan cinco principios fundamentales y relacionados entre sí para una gobernanza responsable de la IA que se alinea con los valores éticos. En línea con estos principios, la OCDE propone a los gobiernos cinco recomendaciones estratégicas: (1) Facilitar la inversión pública y privada en investigación y desarrollo, con el objetivo de estimular la innovación en inteligencia artificial de manera segura y fiable; (2) Promoción de ecosistemas de IA inclusivos, dotados de infraestructuras tecnológicas avanzadas y sistemas digitales, con plataformas que faciliten la difusión y el intercambio de datos y conocimientos; (3) Asegurar un marco de políticas que abra el camino para el despliegue de sistemas de IA fiable; (4) Capacitar a las personas con habilidades necesarias para interactuar con la IA y apoyar a los trabajadores para una transición justa; (5) Cooperar con países y empresas para avanzar en la administración responsable de IA fiable.

- **Global Partnership on Artificial Intelligence (GPAI):** esta es una iniciativa internacional multilateral que promueve el desarrollo y uso responsable de la IA, respetando los derechos humanos y valores democráticos. La GPAI, lanzada oficialmente en junio de 2020 a partir de una propuesta de Canadá y Francia en la cumbre del G7 en 2018, es una plataforma que conecta a expertos de diversas áreas para fomentar la cooperación internacional en IA. Sus miembros incluyen países que se adhieren a las recomendaciones de la OCDE sobre inteligencia artificial y expertos en la materia. La GPAI se enfoca en cuatro áreas principales: IA Responsable, Gobernanza de Datos, el Futuro del Trabajo e Innovación y Comercialización.

- **Centro de Inteligencia Artificial y Robótica del Instituto Interregional de las Naciones Unidas para la Investigación del Delito y la Justicia (UNICRI):** este fue creado para profundizar el entendimiento sobre la IA y la robótica, especialmente enfocándose en sus impactos en el crimen, el terrorismo y otras amenazas a la seguridad. Este centro también apoya a los Estados Miembros para aprovechar el potencial de estas tecnologías de manera responsable. Parte de su trabajo incluye la publicación de informes, como *Artificial Intelligence and Robotics for Law Enforcement*, que recopila hallazgos, desafíos y recomendaciones sobre la contribución de la IA y la robótica en las funciones policiales y cómo estas tecnologías ya no son una posibilidad futura, sino una realidad actual. Además, se ocupa de identificar amenazas y delitos relacionados con el uso malintencionado de estas tecnologías y destaca la importancia de la coherencia con los derechos humanos y los principios de justicia en el uso de estas tecnologías por parte de las fuerzas del orden.

4. CASOS PRÁCTICOS DE IA EN EL ÁMBITO JURÍDICO

La IA se ha convertido en una herramienta transformadora en un gran número de sectores. Su impacto es particularmente significativo en el terreno legal, donde se han introducido innovaciones disruptivas y se han reformulado los enfoques convencionales en la práctica del derecho, de aquí que nos veamos obligados a preguntarnos: ¿Estamos preparados para delegar la gestión de nuestros derechos esenciales a un algoritmo, a un simple dispositivo electrónico?

La inteligencia artificial en el ámbito jurídico se apoya en herramientas como el procesamiento de lenguaje natural (PLN) y el machine learning (ML). El PLN dota a los sistemas de la capacidad para interpretar y generar lenguaje humano, una habilidad crucial para el análisis de documentos legales complejos. Por otro lado, el machine learning faculta a la IA para descubrir patrones y optimizar su actuación conforme procesa nuevos datos.

El derecho constituye un ámbito susceptible a la aplicación de tecnología, lo cual puede resultar en resultados muy llamativos, así como en desafíos bastante complejos, ya que el despliegue tecnológico posee la facultad de alterar los sistemas legales existentes. Sin embargo, hay que recalcar que el rol de la IA en la esfera legal es de apoyo a los profesionales, no de sustitución. Su finalidad es potenciar su eficacia y mejorar la calidad de los servicios que ofrece.

La Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de IA, conocida como *Ley de Inteligencia Artificial, de 21 de abril de 2021* (la cual se analizará detalladamente en apartados siguientes) clasifica como sistemas de *alto riesgo* aquellos usos de la inteligencia artificial que las entidades gubernamentales podrían implementar para una serie de aplicaciones. Entre estas se incluyen: la realización de valoraciones de riesgo individuales, la utilización de polígrafos y dispositivos análogos, o sistemas que determinen el estado emocional de las personas; la identificación de manipulaciones extremas de contenido digital; la valoración de la credibilidad de evidencias en procesos judiciales; la predicción de comisión o reincidencia de crímenes basándose en perfiles individuales; la evaluación de características personales o antecedentes penales de individuos o colectivos; la creación de perfiles para uso en la detección, investigación o prosecución de delitos penales; y el análisis penal relacionado con sujetos específicos (Garrigues, 2023).

Así, a pesar de las inquietudes presentadas procederemos a explorar los posibles usos más destacados de la IA en el ámbito legal, donde en los últimos años se han producido grandes avances, permitiendo que esta tecnología contribuya en el sector no solo con la automatización de labores, sino también realizando aportaciones profundas que apoyan decisiones legales fundamentadas.

4.1 IA en decisiones de libertad condicional

El artículo 90 del Código Penal de España aborda la libertad condicional, un beneficio penitenciario que permite al condenado la suspensión del resto de su pena y obtener una liberación anticipada. Esto es posible tras cumplir la mayor parte de la sentencia (habitualmente tres cuartas partes), haber alcanzado el tercer grado penitenciario y demostrar buena conducta.

En nuestro ordenamiento jurídico, corresponde al Juez de Vigilancia Penitenciaria realizar una valoración del penado, para ello tendrá en cuenta una serie de variables, que le permitirán adoptar una decisión sobre si el sujeto es merecedor del beneficio. Entre estas se evalúa: la personalidad; los antecedentes; las circunstancias del delito cometido; la relevancia de los bienes jurídicos que pudieran verse afectados por una reiteración en el delito; la conducta del penado durante el cumplimiento de la pena; sus circunstancias familiares y sociales y, los efectos que quepa esperar de la propia suspensión de la ejecución y del cumplimiento de las medidas que le fueren impuestas.

Como se puede observar, evaluar el futuro comportamiento de los reclusos una vez en libertad es un proceso complejo y retador. La práctica indica que el comportamiento en prisión no siempre predice cómo actuará una persona en libertad. Individuos que se adaptan bien en la cárcel pueden tener dificultades para reintegrarse en la sociedad y vivir sin cometer delitos. Por otro lado, aquellos que enfrentan problemas de adaptación en prisión, pero cuyo delito fue aislado, pueden tener un pronóstico más positivo para una reinserción exitosa (Burón, & Diario, 2023).

Es aquí donde la inteligencia artificial puede jugar un papel clave en la mejora del proceso de evaluación para la libertad condicional. A través de algoritmos avanzados, la IA tiene el potencial de analizar datos históricos y tendencias de comportamiento para ofrecer predicciones más precisas sobre el comportamiento futuro de los reclusos. Un adecuado uso de estas herramientas puede ayudar a identificar factores de riesgo y de éxito no evidentes para los humanos, permitiendo una gestión más informada del riesgo y un enfoque más personalizado hacia la rehabilitación y la reinserción social.

La integración de la inteligencia artificial en la toma de decisiones sobre la libertad condicional ofrece varias ventajas significativas. El estudio *An algorithmic assessment of parole decisions*, (Laqueur & Copus, 2022), ha demostrado que, mediante el uso de algoritmos de aprendizaje automático, es posible identificar a reclusos de bajo riesgo con mayor precisión y aumentar la tasa de liberación bajo libertad condicional sin incrementar la tasa de arrestos posteriores. El estudio se centro en datos de 4.168 individuos que fueron liberados bajo libertad condicional en Nueva York entre 2012 y 2015. Para ello, examinó 91 variables para predecir el riesgo de criminalidad, entre las que incluyeron:

edad, sentencia mínima y máxima, tipo de prisión, raza, tiempo en prisión, arrestos previos y otros criterios.

Los resultados indicaron que la IA puede contribuir a reducir la población carcelaria manteniendo la seguridad pública. El uso de IA permite analizar grandes volúmenes de datos y tener en cuenta una amplia variedad de variables que afectan el riesgo de reincidencia, algo que podría ser demasiado complejo para analizar en profundidad mediante métodos tradicionales. Además, el estudio sugiere que la inteligencia artificial podría asistir a mitigar las disparidades raciales en las tasas de liberación bajo libertad condicional, promoviendo así una mayor equidad en el sistema de justicia penal.

Por otro lado, uno de los casos más notorios de la última década involucra a un ciudadano estadounidense, Eric Loomis, quien fue sentenciado en 2013 a seis años de cárcel y cinco de libertad vigilada por la Corte Suprema de Wisconsin, por no detenerse cuando un oficial de policía se lo ordenó mientras estaba al volante de un vehículo que no tenía permiso del propietario para conducir. Durante la sentencia, se utilizó la herramienta de evaluación de riesgo COMPAS, que lo clasificó con alto riesgo de reincidencia. Esta conocida herramienta analiza una serie de factores relacionados con los individuos para predecir su riesgo de volver a delinquir y sus necesidades de rehabilitación. La controversia del Caso Loomis se produjo cuando solicitó el acceso al código fuente del software para conocer las razones de la decisión algorítmica y poder contra-argumentar, pero esto se le denegó porque el algoritmo estaba protegido por la propiedad intelectual al pertenecer a una empresa privada (Campione, 2021).

El caso planteó preguntas importantes sobre la justicia, precisión y transparencia de las evaluaciones de riesgo algorítmicas. Estudios han cuestionado la confiabilidad de herramientas como COMPAS, indicando que tales algoritmos podrían no ser más precisos que los seres humanos y podrían reflejar sesgos inherentes a los datos que procesan. El debate se intensificó en torno a si estas herramientas discriminan injustamente basándose en factores que los individuos no pueden controlar, como el estatus socioeconómico o el código postal, potencialmente exacerbando las desigualdades raciales y sociales.

En este sentido, los investigadores subrayan que la IA no debería reemplazar completamente la toma de decisiones humana, sino más bien servir como una herramienta para identificar y corregir problemas dentro del sistema actual. La adecuada supervisión y uso de esta tecnología ofrece una oportunidad para mejorar la evaluación del riesgo, lo que a su vez puede plantear reformas del sistema penal, ayudando a asegurar que las decisiones de libertad condicional sean más informadas, justas y efectivas. De tal forma, que lo que se busca es lograr un balance entre la aplicación de tecnologías como apoyo al sistema jurídico y la protección efectiva de los derechos fundamentales.

4.2 Programas predictivos de criminalidad

En el apartado anterior, la utilización de la IA estaba destinada a ofrecer determinados beneficios al recluso, permitiéndole cumplir el resto de su condena en libertad. Sin embargo, en este contexto, la tecnología se emplea con el fin de anticipar posibles actos delictivos que podrían ser objeto de juicio y resultar en el encarcelamiento del individuo.

La legislación sugiere que el empleo de pronósticos sobre riesgos de actividad criminal está, en gran medida, restringido según el artículo 5 de la propuesta del Reglamento europeo. Sin embargo, se busca una proporcionalidad y equilibrio en aquellos casos en los que sea imprescindible recurrir a la IA por parte de las autoridades y con tales propósitos, evaluando las consecuencias negativas de su aplicación, las características del contexto que justificarían su uso y, especialmente, el alcance del daño que podría ocurrir si el sistema no se utilizara. Por esto, sin una regulación con unos límites bien definidos se corre el riesgo de la posibilidad de violar los derechos fundamentales de la ciudadanía y, particularmente, de aquellos sujetos bajo investigación criminal (Garrigues, 2023).

En este contexto, las técnicas de Machine Learning ya han demostrado su capacidad para procesar y analizar grandes cantidades de datos, identificando patrones delictivos que permiten prevenir y reducir el crimen. Los modelos de ML se perfilan como herramientas clave para las fuerzas y cuerpos de seguridad del Estado, quienes buscan optimizar sus recursos y focalizar sus esfuerzos en áreas o factores específicos que presentan mayor riesgo (Shermila, Bellarmine, & Santiago, 2018). Además, al examinar los registros de criminalidad, se puede obtener una visión más profunda de la estructura social de las

comunidades, facilitando a los políticos a tomar decisiones informadas para abordar de manera preventiva problemas sociales subyacentes.

Uno de los países pioneros en aplicar este tipo de tecnologías es China. El gobierno chino ha utilizado tradicionalmente los datos personales de sus ciudadanos para monitorear y controlar su conducta, sin importar si se trata de delincuentes o individuos bajo sospecha de actividades que el Estado considera políticamente sensibles. No obstante, el avance de las nuevas tecnologías, que abarcan desde dispositivos móviles y ordenadores hasta el software avanzado de inteligencia artificial, está incrementando estas capacidades de supervisión. En este contexto, es donde se desarrollan softwares como los de la empresa de tecnología Cloud Walk, que está experimentando con un sistema que evalúa la probabilidad de que las personas cometan crímenes basándose en su comportamiento y movimientos, como, por ejemplo, si frecuentan tiendas de armas. Este software notifica a las autoridades cuando detecta que una persona que presenta un riesgo delictivo empieza a representar una amenaza potencial, lo que facilita la intervención policial preventiva.

No obstante, no es necesario irnos al continente asiático para ver este tipo de tecnologías, la compañía estadounidense Clearview IA, ofrece un software de reconocimiento facial a agencias gubernamentales, utilizando una base de datos de más de 3.000 millones de imágenes que se han recopilado de redes sociales y otros lugares de internet. En este sentido, *La Resolución del Parlamento Europeo, de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales (2020/2016(INI))* expresó su *gran preocupación* por el uso de este tipo de IA por las autoridades. En la solicitud se insta a los países miembros a exigir a las autoridades encargadas de hacer cumplir la ley que declaren si están empleando tecnología de reconocimiento facial como la de Clearview AI o la de otros proveedores. Ya que, desde el punto de vista del Comité Europeo de Protección de Datos, el uso de este tipo de tecnologías por parte de las autoridades policiales comunitarias posiblemente no sería compatible con las normativas de protección de datos de la UE. Por ello, se solicita a la Comisión Europea que prohíba el uso de bases de datos privadas de reconocimiento facial en el ámbito de la aplicación de la ley (Parlamento Europeo, 2021).

4.3. Chatbots Jurídicos

Otro ejemplo de usos de IA en el ámbito jurídico que cada vez está ganando más popularidad es el uso de los chatbots. Esta tecnología, está diseñada para interpretar las consultas legales y ofrecer respuestas basadas en una amplia base de datos de información jurídica, optimizando el trabajo en bufetes de abogados y ampliando el acceso a consejos legales básicos para quienes podrían no tener los medios para contratar servicios legales convencionales, democratizando las consultas jurídicas.

Los chatbots en el ámbito legal funcionan como asistentes virtuales programados para simular conversaciones humanas y responder a preguntas frecuentes. La principal ventaja de este tipo de tecnologías es su disponibilidad las 24 horas del día, lo que permite a los despachos ofrecer respuestas inmediatas y, por consiguiente, mejorar su accesibilidad y eficiencia. En un reciente artículo de Henry J. (2024), se analizó cómo los 100 despachos de abogados más reconocidos de Estados Unidos, según la clasificación de la revista jurídica *The American Lawyer*, están implementando la inteligencia artificial. Según el reportaje, las aplicaciones más comunes de la IA en estos despachos incluyen la producción de materiales de marketing y biografías de abogados, la síntesis de documentos y la generación de transcripciones, así como la redacción de documentos legales y la investigación jurídica.

Sin embargo, emergen preocupaciones éticas sustanciales en cuanto a la calidad del asesoramiento proporcionado y los límites de la toma de decisiones automatizada. Los chatbots, por su diseño, no pueden emular la comprensión contextual profunda ni la empatía de un abogado humano, lo que es crucial para la representación legal efectiva. Así entre las preocupaciones más comúnmente señaladas en el estudio, destacan: el sesgo, la fijación de precios de servicios mejorados por la IA, la falta de transparencia, la seguridad y confidencialidad de los datos del cliente y del despacho, la precisión y el riesgo de que los abogados cometan errores con la nueva tecnología. En este sentido, Paul Philip, Director Ejecutivo de la SRA (2023), organismo regulador de los abogados en Inglaterra y Gales, dijo: *Los despachos deben asegurarse de entender y mitigar los riesgos de la IA, de la misma manera que un abogado siempre debe supervisar adecuadamente a un empleado con menos experiencia, deberían estar supervisando el*

uso de la IA. Deben asegurarse de que les esté ayudando a prestar servicios legales con los altos estándares que sus clientes esperan.

La creciente integración de los chatbots en el sector jurídico debe acompañarse, por tanto, de una reflexión crítica sobre estas cuestiones y del desarrollo de estándares éticos y prácticas recomendadas que aseguren su alineación con los principios del derecho (Susskind & Susskind, 2015). Para maximizar su potencial, los bufetes deben centrarse en la mejora continua de estos sistemas, con un énfasis especial en la seguridad y la personalización de la interacción. La adaptación y aceptación de tecnologías emergentes como los chatbots serán fundamentales para los despachos que deseen permanecer competitivos y al servicio de las necesidades cambiantes de sus clientes.

5. LA INTELIGENCIA ARTIFICIAL Y LOS DERECHOS HUMANOS

Como se ha expuesto a lo largo del trabajo, la IA se presenta como una fuerza transformadora de los diferentes sectores de la sociedad. En el ámbito económico, este tipo de tecnologías suponen una fuerza de prosperidad, con el potencial de añadir 15,7 billones de dólares al PIB mundial y aumentar este en un 14% para 2030. Sin embargo, esta promesa económica viene acompañada de un gran desafío, ya que la misma tecnología que podría impulsar la riqueza global, también podría desplazar hasta un 30% de la fuerza laboral. Con un tercio de las empresas ya implementando soluciones de IA que reemplazan tareas humanas, y más de la mitad de la población expresando preocupaciones sobre los sesgos y errores en el contenido generado por IA, surge la necesidad de cuestionarnos sobre la coexistencia entre el avance tecnológico y la protección de los derechos humanos (Authority Hacker, 2024).

Por otro lado, Aristóteles nos recuerda en su exploración del *telos* (el fin o propósito último de las cosas) que cada herramienta debe servir al bienestar humano, lo que nos ofrece una perspectiva prudente para examinar el avance de la IA. En el contexto actual, esto nos insta a interrogarnos sobre el verdadero fin de esta tecnología. Al explorar el *telos* de la IA, nos enfrentamos a la obligación ética de dirigirla hacia un desarrollo que

no solo optimice procesos, sino que también enriquezca y respete la identidad moral de nuestra sociedad. La cautela aquí no es solo una opción; es una necesidad, para garantizar que el desarrollo de la IA permanezca fiel al propósito ético de servir a la humanidad, en lugar de convertirse en una fuerza que podría deshumanizarnos y atentar contra la dignidad humana. En este sentido se pronunció el famoso escritor y filósofo israelí, Yuval Noah Harari, en su obra *Homo Deus: Breve historia del mañana*, donde mencionó que la novedad de la inteligencia artificial consiste en el peligro de que los seres humanos puedan perder valor *porque la inteligencia se está desconectando de la conciencia*. (Harari, 2016, p. 341)

Así, en este apartado procederemos a analizar aquellos derechos humanos que pueden ser más vulnerables a los avances tecnológicos descontrolados. En concreto, nos centraremos en los riesgos que suponen para nuestra privacidad, que se ve amenazada por la gran cantidad de datos personales que se pueden deducir de nuestras acciones en línea; los posibles sesgos y discriminaciones que puedan acaecer de estas tecnologías diseñadas por hombres y por último la transparencia y rendición de cuentas, que, a pesar de no ser un derecho humano como tal, su ausencia o limitación puede afectar a estos derechos.

5.1 Derecho a la privacidad

Cuando consideramos el impacto de la inteligencia artificial en los derechos humanos, el derecho a la privacidad surge como una preocupante cuestión. Este derecho, crucial en el espacio digital, afirma que las personas deben tener el control sobre su información personal, asegurando la protección de su espacio privado a menos que decidan compartirlo explícitamente. Anteriormente, la privacidad de uno podía sentirse segura simplemente con cerrar la puerta de su casa. Sin embargo, con los avances tecnológicos actuales, las fronteras de nuestra privacidad se han vuelto difusas y difíciles de definir.

En nuestro mundo cada vez más interconectado, la privacidad no solo se trata de la seguridad de los datos personales, sino que también abarca la libertad de tomar decisiones sin coacción externa. Es esencial proteger los datos personales contra la recopilación y el uso indebido sin consentimiento, lo que resalta la necesidad de marcos legales y tecnológicos robustos para preservar tales derechos.

Por lo tanto, el derecho a la privacidad y la intimidad refuerza la premisa de que cada individuo tiene el poder de determinar qué información personal comparte y con quién la comparte. Esto pone de manifiesto la relación intrínseca entre la autonomía personal, la confidencialidad de la información y el control sobre la propia identidad digital en la era de la inteligencia artificial.

Como explica en su obra Zuboff (2020), vivimos en lo que se denomina el *capitalismo de vigilancia*, una etapa novedosa en la economía política caracterizada por un patrón emergente de acumulación de riqueza digital. En esta era, la recopilación de datos personales trasciende la mejora de productos y servicios y adquiere un valor comercial significativo debido a su capacidad para predecir comportamientos y tendencias. Por medio de los sistemas de IA, estos datos se transforman en mercancías que se negocian en lo que la autora describe como *mercados de futuros conductuales*. Esta dinámica perpetúa la erosión de la privacidad, que se ve agravada por nuestra creciente dependencia de un entorno digital que constantemente nos monitorea y modula nuestro comportamiento. Esta situación nos lleva a una resignación generalizada en la que, bajo la creencia de no tener nada que ocultar, terminamos aceptando una falsa disyuntiva impuesta y renunciamos a nuestro derecho a la privacidad sin plena conciencia del trueque que estamos haciendo (Zuboff, 2020).

Los numerosos escándalos ocurridos en los últimos años ponen de manifiesto esta necesidad de protección real de nuestra privacidad frente a la comercialización constante de nuestros datos personales. El caso de Cambridge Analytica es un ejemplo destacado: esta empresa de análisis de datos se vio envuelta en 2018 en una gran controversia al revelarse que había recopilado información de millones de perfiles de Facebook sin consentimiento de los usuarios. Utilizaron esta información para construir perfiles psicográficos detallados con el fin de influir en el voto electoral, lo que desató un debate global sobre la privacidad y el uso de datos personales en línea. Este escándalo fue un punto de inflexión que subrayó la importancia de implementar medidas más estrictas para la protección de datos en el ámbito digital (Heawood, 2018).

Dada la trascendencia y la potencial manipulación de tal magnitud de datos, en nuestro país, la Agencia Española de Protección de Datos (AEPD, 2020) ha desarrollado una guía

para el cumplimiento del Reglamento General de Protección de Datos (RGPD) en el contexto de la IA, buscando despejar incertidumbres de usuarios y profesionales y proporcionar una base de seguridad jurídica para aquellos involucrados en el procesamiento de datos a través del aprendizaje automático.

En concreto, en la guía se prevé la posibilidad de implementar una Evaluación de Impacto de Privacidad (EIPD). Esto es un proceso para tratamientos que presentan altos riesgos para los derechos y libertades individuales, especialmente cuando involucran la elaboración de perfiles y toma de decisiones automatizadas. Este proceso debe realizarse antes de iniciar el tratamiento de los datos y está orientado a integrar la protección de la privacidad en el diseño del sistema de IA, promoviendo medidas como la supresión, agregación, ocultamiento y segregación de datos para disminuir riesgos a la privacidad. También subraya la importancia de informar y proporcionar control a los interesados sobre el tratamiento de sus datos. Si después de la EIPD persiste un alto riesgo, se debe consultar a la autoridad de control. El proceso debe incluir la adopción de medidas de seguridad y de gobernanza de datos para demostrar el cumplimiento con los principios y obligaciones legales del RGPD (AEPD, 2020, p.31).

Por lo tanto, es necesario una postura crítica y proactiva para asegurar que la evolución de la IA sea compatible con los derechos humanos y las libertades fundamentales. En años recientes, hemos observado una mejoría en la cooperación internacional, desarrollando diversas iniciativas para asegurar una protección sólida de estos derechos. Un enfoque clave ha sido la incorporación de la privacidad desde la etapa inicial en el diseño de sistemas de IA, convirtiéndola en un pilar fundamental durante el desarrollo tecnológico. También, ha sido esencial aumentar la conciencia sobre la privacidad y proporcionar educación a los usuarios respecto a sus derechos y la utilización de sus datos personales, con el objetivo de reducir el impacto negativo sobre su privacidad. Por último, la creación de sistemas de auditoría especializados que evalúen los sistemas de IA bajo criterios de ética, así como la formación de comités independientes, son medidas necesarias para supervisar adecuadamente la implementación de esta tecnología.

5.2 Derecho a la no discriminación

El artículo 14 de la Constitución de 1978 proclama el derecho a la igualdad y a la no discriminación, citando como motivos especialmente rechazables el lugar de nacimiento, la raza, el sexo, la religión u opinión, y prohibiendo la discriminación por cualquier otra circunstancia personal o social.

En el ámbito digital esta discriminación también puede producirse y más aún con el avance de la IA y otras tecnologías con capacidad para tomar decisiones de forma autónoma basadas en una muestra de datos. Es precisamente en los datos donde se producen gran parte de los problemas, la causa de la discriminación puede ocurrir debido a que los datos recopilados y analizados pueden reflejar prejuicios existentes que favorecen a un grupo sobre otro, lo que podría llevar a prácticas arbitrarias basadas en género, raza, religión o estatus socioeconómico, entre otros factores.

Esto también se plasma en el artículo 23 de la *Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación* que subraya el compromiso de las administraciones públicas por garantizar que la IA se utilice de manera justa y transparente, evitando cualquier forma de discriminación en la toma de decisiones automatizadas. Poniendo énfasis en la implementación de algoritmos que minimicen los sesgos y se evalúen rigurosamente para prevenir impactos discriminatorios.

No obstante, existen distintos escenarios en los que la toma de decisiones basada en algoritmos puede resultar en actos discriminatorios. Pero nos centraremos en aquellas que ocurren por prejuicios presentes en los datos usados para entrenar sistemas de IA, particularmente en la esfera del aprendizaje automático o machine learning. Aquí, podemos diferenciar dos tipos de discriminación: aquella que produce la IA de forma voluntaria y la que surge de forma involuntaria (Miquel Sanchis, 2023)

La discriminación involuntaria ocurre cuando los sistemas de inteligencia artificial, basados en conjuntos de datos que contienen sesgos, perpetúan y reflejan estas inclinaciones en sus resultados. El algoritmo, sin tener intención de discriminar, puede producir decisiones sesgadas si el aprendizaje proviene de datos que contienen prejuicios preexistentes. Esto no solo lleva a una discriminación no intencionada, sino que también

puede resultar en una representación desproporcionada de algunos grupos en detrimento de otros en la muestra utilizada. Un ejemplo donde se puede ilustrar este posible sesgo por parte de un sistema de IA sería en el ámbito de los seguros de salud. Si la IA se entrena con datos históricos en los que la mayoría de los asegurados con cobertura amplia son personas de alto nivel socioeconómico, el sistema podría aprender incorrectamente que las características asociadas con dichos niveles socioeconómicos son indicadores de clientes deseables. Esto resultaría en que individuos de niveles socioeconómicos más bajos podrían recibir cotizaciones más altas o ser excluidos de ciertas coberturas, no por su riesgo de salud individual, sino por el perfil socioeconómico inferido a partir de los datos sesgados.

Así, existen diferentes técnicas utilizadas para mitigar los sesgos y mejorar la equidad en los modelos de IA (Bruha,1999):

- **Preprocessing**: esta fase ocurre antes de entrenar el modelo de IA. Implica preparar y limpiar los datos para reducir los sesgos existentes. Las técnicas de preprocessing pueden incluir el balanceo de conjuntos de datos (para que no haya predominancia de una clase sobre otra), la transformación de variables, la eliminación o corrección de datos erróneos o sesgados y la selección de características que no reflejen sesgos discriminatorios.

- **In-processing**: durante el entrenamiento del modelo, el in-processing se refiere a las técnicas integradas directamente en el algoritmo para mitigar los sesgos. Esto puede implicar modificar el algoritmo de aprendizaje para que priorice la equidad y la no discriminación. Algunos enfoques incluyen la adición de términos de regularización que penalizan el sesgo en la función de pérdida del modelo o el ajuste de parámetros que controlan la equidad durante el entrenamiento.

- **Postprocessing**: una vez el modelo ha sido entrenado, el postprocessing implica ajustar las predicciones del modelo para garantizar resultados justos. Esto puede incluir técnicas como la recalibración de las probabilidades de salida del modelo o la modificación de las decisiones de clasificación para asegurar la representación equitativa de diferentes grupos. El postprocessing es útil

especialmente cuando no se desea o no se puede cambiar el proceso de aprendizaje o los datos subyacentes.

También existe la posibilidad de que la discriminación por parte de algoritmos ocurra de forma voluntaria. Esta ocurre cuando, los encargados de implementar sistemas de IA y los propios algoritmos podrían estar programados con prejuicios intencionales y deliberados durante la recolección de datos. Es complicado demostrar la premeditación detrás de estas prácticas, sobre todo cuando ocurre en contextos corporativos por motivos económicos que buscan justificar dichas prácticas (Miquel Sanchis, 2023).

5.3 El derecho a la transparencia y rendición de cuentas

Si bien la transparencia y la rendición de cuentas no se consideran derechos humanos *per se,* su ausencia o carencia puede tener un impacto significativo en la protección y el ejercicio de muchos de estos derechos. Es por esto que, para fomentar y mantener la confianza de los usuarios en los sistemas de inteligencia artificial, es esencial asegurar la transparencia en los procesos. Esto significa que los procedimientos deben ser claros, los objetivos y capacidades del sistema deben ser comunicados de manera abierta y las decisiones que toma el sistema deben ser comprensibles tanto para los afectados directos como para los indirectos.

La transparencia en la utilización de la Inteligencia Artificial (IA) se ha convertido en un pilar fundamental en varias legislaciones, incluyendo la propuesta del Parlamento Europeo y el Consejo para regular la IA. El título IV del mencionado Reglamento detalla las obligaciones de transparencia para ciertos sistemas de IA. Específicamente, el artículo 52 obliga a los proveedores a garantizar que los usuarios reconozcan cuando interactúan con un sistema de IA, salvo excepciones justificadas legalmente. Adicionalmente, los sistemas capaces de detectar emociones o realizar categorización biométrica deben informar a los usuarios sobre su uso, a menos que formen parte de investigaciones penales autorizadas. Se subraya también la necesidad de identificar claramente los contenidos generados o alterados por IA, como los *ultrafalsificados*, excepto cuando su uso esté permitido para ciertos fines legales siempre y cuando no se violen derechos de terceros.

Estas disposiciones buscan asegurar la transparencia y proteger los derechos fundamentales en el contexto de la IA.

Sin embargo, la necesidad de transparencia en los sistemas de inteligencia artificial debe ir más allá de la teoría y aplicarse de manera concreta en la realidad. Es crucial que la claridad en la creación de algoritmos sirva de base para una implementación efectiva que responda a este imperativo.

En el libro *Towards transparency by design for artificial intelligence* (Felzmann *et al*, 2020) se establecen una serie de principios a seguir en la fase de diseño de estas tecnologías, con el objetivo de garantizar su transparencia. Los principios son los siguientes:

- **Anticiparse es mejor que reaccionar (Principio 1):** la planificación adelantada para que estos sistemas sean transparentes y sus decisiones sean accesibles a los afectados es fundamental. Existen metodologías de diseño establecidas que pueden guiar este proceso desde las etapas tempranas del desarrollo de la IA.
- **Pensar en la transparencia como un proceso integrador (Principio 2):** se enfatiza que la transparencia debe abordar la complejidad del proceso de toma de decisiones y cómo los distintos aspectos del sistema cumplen con los estándares establecidos para dichas decisiones. Dependiendo del propósito de la información y de la audiencia, estos estándares de decisión pueden presentarse como principios generales o como estándares más específicos y detallados. Una visión integradora también implica la conciencia del impacto práctico que puede tener el uso de estándares de toma de decisiones inadecuados o insuficientemente realizados.
- **Comunicar teniendo en cuenta a la audiencia (Principio 3):** destaca la necesidad de considerar la audiencia que va a recibir la información y cómo la interpretará, lo que implica que la forma y el contenido de la información proporcionada deben adecuarse a quienes la van a usar.
- **Explicar qué datos se utilizan y cómo se procesan (Principio 4) y explicar los criterios de toma de decisiones y su justificación (Principio 5):** se reconocen las limitaciones técnicas que pueden dificultar la explicación en la toma de decisiones de sistemas de IA complejos y se subraya la necesidad de proporcionar un esquema descriptivo claro de los datos usados y los procesos de tratamiento de

datos, incluyendo las etapas que pueden ser inspeccionadas así como el sistema normativo en el que se sustenta.

- **Explicar los riesgos y las medidas para mitigarlos (Principio 6):** este principio se concentra en hacer transparentes los nuevos riesgos asociados con la operación de sistemas de IA.
- **Garantizar la inspeccionabilidad y auditabilidad (Principio 7):** los sistemas de IA deben permitir la inspección de la toma de decisiones del sistema a través de auditorías.
- **Ser receptivos a las consultas y preocupaciones de las partes interesadas (Principio 8):** va más allá de simplemente permitir la inspección y auditoría de decisiones individuales, enfatizando la necesidad de estar abierto al escrutinio y al diálogo con periodistas, organizaciones civiles, administradores públicos y el público general.
- **Informar diligentemente sobre el sistema (Principio 9):** este principio especifica que los diseñadores de sistemas de IA deben hacer transparentes sus actividades mediante informes detallados.

La transparencia debe de ir de la mano de la rendición de cuentas. En el preámbulo de la propuesta de Reglamento sobre Inteligencia Artificial, destaca la importancia de identificar como de alto riesgo aquellos sistemas de IA que se emplean en ámbitos de aplicación de la ley. Esto se debe a la necesidad de que estas tecnologías sean altamente precisas, fiables y transparentes para prevenir consecuencias negativas en los individuos y mantener su confianza. Más adelante, en el artículo 17 del Reglamento se menciona que los proveedores de sistemas de inteligencia artificial considerados de alto riesgo deben implementar un marco de rendición de cuentas, definiendo los deberes de los directivos y el personal en relación con todas las facetas del mandato.

6. LEGISLACIÓN Y POLÍTICA FUTURA: LEY DE INTELIGENCIA ARTIFICIAL DE LA UE

En abril de 2021, la Comisión Europea alcanzó un hito legislativo al presentar el primer marco legal de la UE en inteligencia artificial, reflejando un paso adelante en la Estrategia para el Mercado Único Digital basada en el artículo 114 del TFUE. Esta legislación, respaldada ampliamente por el Parlamento Europeo (84 votos a favor, siete en contra y 12 abstenciones), busca *garantizar la seguridad y el respeto de derechos*, además de *impulsar la innovación*. La propuesta se enmarca dentro de los esfuerzos por mantener a Europa a la vanguardia en la regulación técnica y legal de la IA, en un contexto global donde otras potencias ya están avanzando en la misma dirección, con la visión de que la UE no quede relegada a un papel pasivo en la economía digital global.

El Reglamento excluye su aplicación a áreas que no están bajo la jurisdicción del Derecho de la Unión Europea, así como a las competencias de los Estados miembros en lo que respecta a la seguridad nacional. Tampoco se extiende a los sistemas de IA que se utilizan estrictamente para propósitos militares o de defensa, ni a aquellos utilizados en contextos de investigación e innovación. El ámbito de aplicación incluye: proveedores que operen o implementen sistemas de IA en la Unión, sin importar su ubicación; proveedores y usuarios en terceros países cuyos sistemas se utilicen dentro de la Unión; usuarios en la UE; así como los representantes autorizados, importadores y distribuidores de dichos sistemas.

Como hemos ido anunciando a lo largo del trabajo, en este apartado se analizará aquellos aspectos más relevantes de la Ley y las opiniones de los críticos.

6.1 Clasificación de riesgos

Uno de los aspectos más llamativos de la legislación es la clasificación de los diferentes riesgos de los sistemas de IA. La ley dispone una serie de responsabilidades específicas para los proveedores y usuarios de sistemas de inteligencia artificial, las cuales varían dependiendo del grado de riesgo asociado con el uso de la IA. Así, se identifican cuatro niveles de riesgo: riesgo mínimo, riesgo limitado, riesgo alto y riesgo inaceptable.

- **Riesgo Mínimo:** La mayoría de las aplicaciones de IA caerán bajo este nivel, donde los sistemas no presentan un riesgo significativo para los derechos o la seguridad. Ejemplos podrían incluir sistemas de IA que ofrecen recomendaciones de películas o música a los usuarios.

- **Riesgo Limitado:** En esta categoría se incluyen sistemas de IA que tienen riesgos potenciales, pero estos riesgos son menos graves o más fácilmente mitigables. El riesgo limitado se refiere a los sistemas de IA con obligaciones específicas de transparencia. Un ejemplo sería un asistente virtual que requiere una cierta transparencia en su funcionamiento, pero no representa una amenaza grave para los derechos o seguridad. Cuando interactúan con chatbots impulsados por IA, es fundamental que los usuarios sepan que están comunicándose con un sistema automatizado, permitiéndoles así decidir conscientemente si desean seguir con la interacción o no. Los proveedores deben garantizar que los contenidos creados por sistemas de IA sean claramente identificados como tales. Asimismo, cualquier texto generado por IA que se difunda con el fin de informar al público sobre temas de interés general deberá estar marcado explícitamente como generado por inteligencia artificial. Esta norma se extiende al contenido de audio y video, especialmente en el caso de las creaciones hiperrealistas o *deepfakes*.

- **Riesgo alto:** regulados en Título III de la Propuesta, los sistemas de IA clasificados aquí son aquellos que tienen un potencial significativo para afectar negativamente a la seguridad o a los derechos fundamentales. Entre los sistemas de IA que se consideran de alto riesgo están aquellos utilizados en infraestructuras críticas, como el transporte; en el ámbito educativo o de formación, que afectan las oportunidades educativas y la trayectoria profesional; en componentes de seguridad de productos, como la IA en cirugías robóticas; en el ámbito laboral, incluyendo el software que selecciona currículums en procesos de contratación; en servicios públicos y privados fundamentales, como los sistemas de calificación crediticia que impactan la capacidad de obtener préstamos; en la aplicación de la ley que afecta los derechos fundamentales, como la evaluación de pruebas; en la gestión de migraciones y control de fronteras, como la verificación de documentos; y en la administración de justicia y procesos democráticos, como la aplicación de leyes a hechos específicos.

En el artículo 9, se enuncia un sistema de gestión de riesgos asociado a la IA de alto riesgo. Este será un proceso repetitivo a lo largo de toda la vida útil de la tecnología, que incluye: a) La identificación y análisis de todos los riesgos posibles asociados al uso del sistema de IA; b) La evaluación y valoración de los riesgos potenciales tanto en usos previstos como en mal uso razonablemente anticipado; c) La consideración de riesgos adicionales basados en los datos obtenidos del seguimiento posventa según se menciona en otro artículo relevante; d) La implementación de medidas apropiadas para manejar los riesgos identificados.

- **Riesgo inaceptable:** son aquellos sistemas de IA que se consideran un peligro para la seguridad o los derechos fundamentales de las personas y, por ende, estarán prohibidos. Esto incluye sistemas que puedan manipular el comportamiento de grupos vulnerables, como juguetes inteligentes que podrían incitar a los niños a adoptar conductas peligrosas; sistemas de puntuación social (*social scoring*) que evalúan a las personas basándose en criterios como su situación socioeconómica o rasgos personales; tecnologías de reconocimiento biométrico en tiempo real y a distancia, como el reconocimiento facial; sistemas policiales predictivos basados en perfiles, localización o comportamientos delictivos anteriores; sistemas para detectar emociones utilizados por la seguridad pública, en la administración de fronteras, espacios laborales y entornos educativos. También se incluyen métodos que influyen en la percepción inconsciente de las personas para cambiar su comportamiento de manera que podría causarles daño físico o psicológico a sí mismos o a otros. Sin embargo, se contemplan excepciones, como en el caso de sistemas de identificación biométrica cuando operen a posteriori, permitidos para investigar delitos graves, siempre que exista aprobación judicial previa.

6.2 Transparencia y códigos de conducta

Los sistemas de IA no considerados de alto riesgo únicamente necesitan cumplir con exigencias limitadas en cuanto a transparencia, como la necesidad de informar a las personas cuando están interactuando con IA. Para los sistemas de alto riesgo, se requieren normativas más estrictas relacionadas con la calidad de los datos, documentación,

trazabilidad y vigilancia humana para mitigar posibles riesgos para la seguridad y los derechos fundamentales. Estas medidas buscan asegurar la transparencia sin afectar de forma excesiva los derechos de propiedad intelectual, limitándose a la divulgación de la *información mínima* que permite a los individuos obtener una compensación efectiva.

En concreto, los sistemas de alto riesgo deben ofrecer un nivel de transparencia que permita a los usuarios entender y utilizar correctamente las salidas del sistema, contrarrestando cualquier opacidad que pueda complicar su comprensión y uso. Esta información abarcará desde la identificación del proveedor hasta los detalles operativos del sistema, incluyendo su propósito, precisión, y seguridad. Además, se describirán las capacidades y posibles limitaciones del sistema, así como los riesgos asociados y la manera en que se ha validado su funcionamiento. Se especificarán los datos usados para entrenar, validar y probar el sistema, y se informará sobre las medidas adoptadas para facilitar la supervisión humana del sistema. Por último, se detallará la vida útil esperada del sistema y las recomendaciones para su mantenimiento y actualización de software, asegurando así su correcto funcionamiento a lo largo del tiempo.

Por otro lado, en el Título IX se establecen las bases para crear códigos de conducta que animen a los proveedores de sistemas de IA de riesgo no elevado a adherirse voluntariamente a las normativas destinadas a los sistemas de IA de alto riesgo, descritas en el Título III. Estos proveedores podrían desarrollar sus propios códigos que no solo se adhieran a las regulaciones obligatorias, sino que también abarquen compromisos adicionales como la sostenibilidad ambiental, la accesibilidad para personas con discapacidad, la inclusión de partes interesadas en el proceso de creación de IA y la promoción de la diversidad en los equipos de desarrollo.

6.3 Impulso a la innovación

El Título V del Reglamento enuncia *espacios controlados de pruebas para la IA*, conocidos como *Sandboxes*, creados con el objetivo de apoyar la innovación en este ámbito. Serán las autoridades nacionales las encargadas de fomentar un ambiente regulado que promueva el desarrollo, la experimentación y la verificación de nuevas

soluciones de IA durante un tiempo restringido antes de su lanzamiento comercial, siguiendo un esquema detallado.

La participación en *sandboxes*, estará limitada en el tiempo y dependerá del alcance del proyecto. Esta participación deberá apegarse a un plan establecido y, mientras los involucrados sigan las indicaciones de las autoridades y se adhieran al plan, estarán exentos de penalizaciones administrativas por posibles infracciones legales concernientes al sistema en prueba dentro del *sandbox*. Sin embargo, esto no afectará la capacidad de las autoridades para ejercer sus funciones de supervisión ni eximirá a los participantes de la responsabilidad por daños ocasionados durante su participación.

En estos espacios, respetando las normas impuestas, se permitirá el uso de datos personales originalmente recopilados para otros fines si son esenciales para el desarrollo de sistemas con un significativo interés público, como aquellos relacionados con la salud, el medio ambiente, la energía sostenible, la movilidad o la seguridad de infraestructuras críticas.

Las condiciones para las pruebas en entornos reales serán acordadas entre las autoridades nacionales y los participantes, con el objetivo de proteger la seguridad, la salud, los derechos fundamentales, las normativas actuales del Reglamento, así como de otras leyes pertinentes de la Unión Europea y de los Estados miembros, todo dentro del contexto de un entorno de prueba regulado.

6.4 La oficina de IA de la Comisión Europea

En el Reglamento se menciona la voluntad de crear una entidad centralizada a nivel europeo, conocida como el *Comité Europeo de Inteligencia Artificial*, que, junto con autoridades designadas por cada estado miembro, supervisará la regulación de los sistemas de IA. Individuos y grupos podrán presentar reclamaciones ante estas autoridades si consideran que hay una infracción de la normativa. Estas autoridades tendrán la obligación de mantener informado al denunciante sobre el estado y el resultado de su queja y proporcionar información sobre la opción de procedimientos judiciales si fuera necesario.

Además, se establecerá una base de datos pública a nivel de la UE que recogerá información sobre los sistemas de IA de alto riesgo, facilitada por los proveedores de dichos sistemas. Esta base de datos permitirá una supervisión más exhaustiva al proporcionar a las autoridades y al público la capacidad de verificar el cumplimiento de los sistemas con los requisitos establecidos.

6.5 Sanciones

Según lo que se establece en el Reglamento, cada estado miembro será responsable de definir y aplicar un sistema de sanciones, que incluirá multas administrativas, para las violaciones a dicha normativa. Se tomarán todas las medidas necesarias para asegurar su implementación efectiva. Las sanciones serán establecidas de forma que resulten justas, proporcionales y suficientemente severas como para tener un efecto disuasorio. Especial consideración se dará a los intereses económicos de los pequeños proveedores y startups, asegurando que las sanciones no comprometan su viabilidad económica. En el texto, se enuncian las siguientes:

- Hasta 30M€, o 6% del volumen anual de negocio para empresas (3% si son PYMEs) por incumplir las prohibiciones del Reglamento.
- Hasta 20M€, o 4% del volumen anual de negocio (2% para PYMEs) por incumplir las obligaciones del Reglamento a proveedores, importadores, distribuidores, usuarios.
- Hasta 10M€, o 2% del volumen anual de negocio (1% para PYMEs) por suministrar información incorrecta a entidades notificadas o autoridades nacionales competentes.
- Hasta 500K€ a entidades de la Unión por incumplimiento de prohibiciones o 250K€ por incumplimiento de obligaciones.

6.6 Críticas

Se podría sintetizar la situación internacional de la tecnología en el famoso dicho: *Estados Unidos inventa, China copia y Europa regula*. En este escenario, Estados Unidos continúa liderando el avance tecnológico, con empresas destacadas como Google,

Microsoft y OpenAI, que están al frente de la innovación en inteligencia artificial, incluyendo el desarrollo de algoritmos de aprendizaje automático y modelos de lenguaje natural. Por otro lado, se percibe a China como una potencia en rápido crecimiento tecnológico, frecuentemente asociada con la adaptación y réplica de innovaciones. En Europa, el enfoque es distinto: existe una tendencia proactiva y cautelosa hacia la regulación de tecnologías emergentes, lo que se refleja en la reciente legislación de IA. Dicha ley busca crear un entorno ético y legal que garantice la seguridad y el respeto de los derechos fundamentales en la aplicación de la IA, buscando un equilibrio entre la innovación y la protección de los ciudadanos.

Así, una parte de la industria tecnológica europea está marcada por una creciente preocupación respecto a la Ley de Inteligencia Artificial y es justamente la percepción de excesiva regulación la que concentra la mayor parte de las críticas. Se argumenta que la rigurosidad del nuevo marco legal podría frenar la innovación, llevando incluso a un éxodo de empresas y talentos del sector de la IA hacia regiones con normativas menos restrictivas. Estas inquietudes emergen en un contexto donde el sector de la IA ya representa un movimiento significativo en la economía del continente, con una cifra que asciende a 33.200 millones de dólares anuales, de acuerdo a la consultora IDC (2023).

El acuerdo, alcanzado el 8 de diciembre, ha sido recibido con escepticismo por parte de entidades como la Computer and Communications Industry Association (CCIA), que lo califica de *oportunidad perdida*. Este sentimiento es compartido por otras voces dentro de la industria que critican la desviación del texto legislativo del enfoque basado en riesgos, propuesto inicialmente por la Comisión Europea, que buscaba un equilibrio entre innovación y regulación (El Español, 2023).

Las sanciones por incumplimiento de la ley, que podrían ascender hasta 35 millones de euros o el 7% del volumen de negocio global, añaden otro factor de preocupación en cuanto a las consecuencias económicas. La verdadera prueba para el Reglamento y su impacto en la economía europea vendrá con su implementación en los próximos años. Habrá que observar si Europa logra consolidarse como un referente de garantismo tecnológico sin pagar un alto coste de oportunidad, o si, como temen algunos sectores de la industria, esta legislación frenará el desarrollo de una tecnología que es crucial en la actualidad.

7. CONCULSIONES

Como hemos observado a lo largo del trabajo, la inteligencia artificial está transformando la sociedad a un ritmo sin precedentes, presentando tanto oportunidades como desafíos en múltiples esferas, incluida la jurídica. En el trabajo de fin de grado se ha explorado en profundidad cómo la ética interactúa con la avanzada era de la IA, abordando cuestiones filosóficas, legales y prácticas. En este apartado de conclusiones, resumiremos los hallazgos principales y cómo se vinculan con los objetivos presentados al principio del documento.

I. Primero, hemos observado que la IA, con su extensa capacidad para procesar y analizar datos, ofrece herramientas revolucionarias para el ámbito legal. Sin embargo, esta tecnología plantea preguntas fundamentales sobre la equidad, la transparencia y la responsabilidad. Los casos prácticos examinados, como el uso de IA en decisiones de libertad condicional y programas predictivos de criminalidad, ilustran tanto el potencial como los peligros inherentes a la dependencia de algoritmos en procesos legales críticos. Siendo necesaria, la supervisión humana en todo momento, de tal forma que la tecnología se encuentre al servicio de las personas y no al revés.

II. La interacción entre la IA y los humanos ha revelado áreas de preocupación, especialmente en lo que respecta a derechos fundamentales, como la privacidad, la no discriminación y la transparencia. La capacidad de la IA para facilitar nuestro día a día no debe privar la necesidad de proteger estos derechos humanos. Por lo tanto, la regulación juega un papel crucial en la mediación de los efectos de la IA en la sociedad.

III. La Unión Europea con la Ley de Inteligencia Artificial, representa un paso adelante en la dirección correcta, estableciendo un marco para la clasificación de riesgos, la promoción de la innovación y la imposición de sanciones por mal uso. Este marco legal, en su intento de ser exhaustivo, enfrenta el desafío de mantenerse al día con la evolución tecnológica, asegurando que las provisiones sean lo suficientemente flexibles para adaptarse a los nuevos inventos, pero también lo suficientemente firmes para prevenir violaciones de los derechos

fundamentales, una tarea compleja dada la velocidad a la que emergen las nuevas aplicaciones de IA.

IV. Europa se enfrenta a un reto complejo. Por un lado, su nueva Ley de IA pretende salvaguardar los derechos de los ciudadanos y consumidores, pero por otro, podría restringir la innovación y fomentar la fuga de empresas y talento hacia regiones con regulaciones más flexibles. La Unión Europea debe esforzarse por convertirse en un referente en el uso ético de la tecnología, pero evitando incurrir en un coste de oportunidad excesivo.

Este trabajo reconoce la IA como un espejo que refleja nuestras virtudes y defectos como sociedad. La ética jurídica no solo debe guiar el desarrollo de la IA, sino también permitirnos comprender y cuestionar los principios que valoramos en nuestra sociedad.

En última instancia, la IA, como cualquier herramienta creada por la humanidad, tiene el potencial de ser tan noble o tan nefasta como los fines a los que la destinamos. Por esto, es tarea de la ética jurídicia garantizar que la dirección que tome la IA se corresponda con ideales de justicia y bienestar que vayan más allá de lo tecnológico, en consonancia con los valores más distinguidos del ser humano.

Sin embargo, la verdadera respuesta a cómo se materializará en el futuro no la proporcionará solamente la reflexión filosófica; la economía, ese gigante de pies de barro, jugará un papel determinante. Aunque nuestros interrogantes existenciales y el curso de la IA pueden ser influenciados por las dinámicas económicas, el capital no siempre se destina a la iluminación del conocimiento. Así, mientras buscamos asegurar que la IA sirva a los valores humanos más elevados, no podemos ignorar la influencia de las realidades económicas que pueden, en última instancia, determinar su trayectoria.

Dentro de las limitaciones inherentes al estudio se incluye el dinamismo de la disciplina, que puede hacer que los descubrimientos relevantes hoy día se tornen obsoletos en un futuro cercano. Además, la extensa variedad de aplicaciones posibles de la IA conduce a un análisis que, por su amplitud, quizás no alcance la profundidad técnica necesaria en determinados campos. Estas limitaciones subrayan la importancia de una investigación

continua y especializada para seguir el ritmo del progreso tecnológico y sus implicaciones en la sociedad y el derecho.

Las futuras líneas de investigación en el campo de la inteligencia artificial pordrían centrarse en profundizar sobre la protección de datos y la privacidad, evaluando cómo tecnologías emergentes mencionadas en el texto, como el reconocimiento facial y la vigilancia automatizada pueden ser reguladas para proteger las libertades individuales. Es crucial también investigar las consecuencias a largo plazo de la IA en el derecho, incluyendo su impacto en la práctica y teoría jurídicas. Además, realizar comparaciones internacionales para entender cómo las distintas legislaciones que regulan la IA afectan al desarrollo de estas tecnologías y a los derechos de los ciudadanos.

8. BIBLIOGRAFÍA

Agencia Española de Protección de Datos. (2020). *Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción.* Recuperado de https://www.aepd.es/sites/default/files/2020-02/adecuacion-rgpd-ia.pdf

Authority Hacker. (2024). *AI Survey: How 3,812 Digital Marketers Are Using AI in 2024.* Recuperado de https://www.authorityhacker.com/ai-survey/

Bruha, I. (1999). Pre-and post-processing in machine learning and data mining. In *Advanced Course on Artificial Intelligence* (pp. 258-266). Berlin, Heidelberg: Springer Berlin Heidelberg.

Burks, A. W. (1984). *John von Neumann and the Origins of Modern Computing.* MIT Press.

Burón, J. N., & Diario, L. A. (2023). La inteligencia artificial al servicio de la ejecución penal. Posibles utilidades. *Diario La Ley*, (10330), 4.

Campbell, M., Hoane, A. J., & Hsu, F. H. (2002). *Deep Blue.* Artificial Intelligence, 134(1-2), 57-83.

Campione, T. R. (2021). *Recopilar y vigilar: algunas consideraciones filosófico-jurídicas sobre inteligencia artificial.* Sociología y Tecnociencia, 11.

Canle Fernández, E. (2021, 4 de febrero). Los antecedentes de la inteligencia artificial. Tokio School. Recuperado de https://www.tokioschool.com/noticias/antecedentes-inteligencia-artificial/

Comisión Europea. (2018). COM(2018) 795 final. *Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones.* EUR-Lex. Recuperado de https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52018DC0795

Crevier, D. (1993). *AI: The Tumultuous History of the Search for Artificial Intelligence.* New York, NY: Basic Books.

El Español. (2023). Ley europea de inteligencia artificial amenaza con frenar un sector que mueve millones de dólares al año. *El Español.* Recuperado de https://www.elespanol.com/invertia/disruptores-innovadores/politica-digital/europa/20231210/ley-europea-inteligencia-artificial-amenaza-frenar-sector-mueve-millones-dolares-ano/815918489_0.html

España. (1978). *Constitución Española, Artículo 14: Derecho a la igualdad de trato y no discriminación en la educación no formal.* Recuperado de https://www.boe.es/buscar/doc.php?id=BOE-A-2022-11589

España. (2022). Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación. *BOE-A-2022-11589.* Boletín Oficial del Estado. Recuperado de https://www.boe.es/buscar/act.php?id=BOE-A-2022-11589

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2020). Towards transparency by design for artificial intelligence. *Science and engineering ethics*, 26(6), 3333-3361.

Garrigues. (2023). Inteligencia artificial aplicada a la investigación criminal: todo un reto para la defensa. Recuperado de https://www.garrigues.com/es_ES/garrigues-digital/inteligencia-artificial-aplicada-investigacion-criminal-todo-reto-defensa

Grupo Europeo de Ética en Ciencia y Nuevas Tecnologías. (2018). *Statement on Artificial Intelligence, Robotics and "Autonomous" Systems.* Comisión Europea, Dirección General de Investigación e Innovación, 11-12.

Hajian, S., & Domingo-Ferrer, J. (2013). A Methodology for Direct and Indirect Discrimination Prevention in Data Mining. *IEEE Transactions on Knowledge and Data Engineering*, 25(7), 1445- 1459. p. 1445.

Harari, Y. N. (2016), *Homo Deus: A Brief History of Tomorrow*, London: Vintage, 2015; trad. cast. *Homo Deus. Breve historia del porvenir,* Barcelona: Debate.

Heawood, J. (2018). Pseudo-public political speech: Democratic implications of the Cambridge Analytica scandal. *Information polity*, 23(4), 429-434.

Henry, J. (2024). We Asked Every Am Law 100 Firm How They're Using Gen AI: Here's What We Learned. *The American Lawyer*. Recuperado de https://www.law.com/americanlawyer/2024/01/29/we-asked-every-am-law-100-firm-how-theyre-using-gen-ai-heres-what-we-learned/

High-Level Expert Group on Artificial Intelligence. (2018). A definition of AI: Main capabilities and disciplines. Recuperado de https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf

IDC. (2023). *Worldwide AI and Generative AI Spending Guide.* Recuperado en abril de, de https://www.idc.com/getdoc.jsp?containerId=IDC_P33198

Instituto de Ciencias de la Computación (ICC), Universidad de Buenos Aires. (s.f.). ¿Qué es la inteligencia artificial? Recuperado de https://icc.fcen.uba.ar/que-es-la-inteligencia-artificial/

Laqueur, H. S., & Copus, R. W. (2022). An algorithmic assessment of parole decisions. *Journal of Quantitative Criminology*, 1-38.

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine,* 27(4), 12-14.

Miquel Sanchis, J. (2023). La primera regulación positiva sobre la Inteligencia Artificial en España.

Minsky, M. (1961). Steps toward Artificial Intelligence. *Proceedings of the IRE,* 49(1).

Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis.* San Francisco, CA: Morgan Kaufmann Publishers.

Parlamento Europeo. (2021). Resolución del Parlamento Europeo del 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales (2020/2016(INI)) [TA-9-2021-0405]. Recuperado de https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_ES.pdf

Parlamento Europeo. (2023). Ley de IA de la UE: Primera normativa sobre inteligencia artificial. Recuperado de https://www.europarl.europa.eu/topics/es/article/20230601STO93804/ley-de-ia-de-la-ue-primera-normativa-sobre-inteligencia-artificial

Plan de Recuperación, Transformación y Resiliencia del Gobierno de España. (2023). ¿Qué es la inteligencia artificial (IA)? Recuperado de https://planderecuperacion.gob.es/noticias/que-es-inteligencia-artificial-ia-prtr

Rodríguez, N. (Fecha no especificada). Historia de la inteligencia artificial. Medium. Recuperado de https://medium.com/@natisr/historia-de-la-inteligencia-artificial-63277f78fe2c

Rouhiainen, L. (2018). *Inteligencia artificial.* Recuperado de https://planetadelibrosec0.cdnstatics.com/libros_contenido_extra/40/39308_Inteligencia_artificial.pdf

Russell, S. J., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach (4th ed.).* Upper Saddle River, NJ: Prentice Hall.

Samoili, S., Cobo, M. L., Gómez, E., De Prato, G., Martínez-Plumed, F., & Delipetrev, B. (2020). AI Watch. *Defining Artificial Intelligence.* Towards an operational definition and taxonomy of artificial intelligence.

Shermila, A. M., Bellarmine, A. B., & Santiago, N. (2018, May). Crime data analysis and prediction of perpetrator identity using machine learning approach. In *2018 2nd international conference on trends in electronics and informatics (ICOEI)* (pp. 107-114). IEEE.

Smarandache, F. (2022). *Collected Papers (on Physics, Artificial Intelligence, Health Issues, Decision Making, Economics, Statistics), Volume XI.*

Solicitors Regulation Authority. (2023). Report looks at pros and cons of AI in law firms. Recuperado de https://www.sra.org.uk/sra/news/press/2023-press-releases/risk-outlook-ai/#:~:text=Accuracy%20and%20bias%20problems%20%E2%80%93%20these,in%20computers%20than%20in%20humans.

Susskind, R. E., & Susskind, D. (2015). *The future of the professions: How technology will transform the work of human experts.* Oxford University Press, USA.

Unión Europea. (2017). Resolución legislativa del Parlamento Europeo del 16 de febrero de 2017 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo relativo al respeto de la vida privada y a la protección de los datos personales en las comunicaciones electrónicas (COM(2017)0010 – C8-0009/2017 – 2017/03(COD)) [CELEX:52017IP0051]. Recuperado de https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52017IP0051&from=EN

Universidad Complutense Madrid. (s.f.). Ética jurídica. Metodología de la investigación en ética aplicada. Recuperado de https://www.ucm.es/metodologia-investigacion-etica-aplicada/etica-juridica

Vestager, M. [@vestager]. (2020, February 19). Artificial intelligence is not good or bad in itself: It all depends on why and how it is used. Let's enable the best possible use and control the risks that AI may pose to our values - no harm, no discrimination! #EUshapingDigital [Tweet]. Twitter. Recuperado de https://twitter.com/vestager/status/1230087490415087616?lang=es

Wiener, N. (1948). *Cybernetics or Control and Communication in the Animal and the Machine.* MIT Press.

Wiener, N. (1950). *The Human Use of Human Beings: Cybernetics and Society.* Houghton Mifflin.

Zuboff, S. (2020). *Capitalismo de la vigilancia.* Política exterior, 34(194), 7-12.