FACULTAD DE DERECHO

PROBLEMAS ÉTICOS Y JURÍDICOS DE LA INTELIGENCIA ARTIFICIAL

Desde la perspectiva kantiana

Autor: Paula Albarrán Muro

4º, E-5 FIPE

Área: Filosofía del Derecho

Madrid

Marzo 2025

RESUMEN

El desarrollo de la Inteligencia Artificial (IA) durante los últimos años plantea importantes desafíos éticos y jurídicos, por lo que es necesario llevar a cabo un profundo análisis de las normativas existentes. Este trabajo estudia estos problemas desde la perspectiva de la ética de Kant, analizando en qué medida se adecuan a ella las normas existentes así como la responsabilidad jurídica de la IA y los agentes que participan de la misma. Para ello, se examinan cuestiones como la responsabilidad jurídica en caso de daños, la autonomía moral de los sistemas, la protección de datos y la transparencia en la toma de decisiones automatizadas.

Por medio de la Teoría Tridimensional del Derecho, el análisis concluye que es fundamental crear una normativa internacional que se base en principios éticos universales acordes con la ética kantiana para garantizar un uso justo y responsable de la IA. Por ello, se propone la creación de un Tratado Internacional suscrito por las principales potencias tecnológicas, en el que se establezcan unos principios éticos universales así como un marco normativo común que incluya la implementación de mecanismos independientes de supervisión y auditoría para evitar cualquier discriminación y asegurar que las decisiones de los sistemas automatizados sean justas.

PALABRAS CLAVES

Inteligencia Artificial, ética kantiana, derechos humanos, responsabilidad jurídica, autonomía moral, daños, privacidad de datos, transparencia, armonización.

ABSTRACT

The development of Artificial Intelligence (AI) in recent years poses significant ethical and legal challenges, making it necessary to conduct a thorough analysis of existing regulations. This work studies these problems from the perspective of Kantian ethics, analyzing to what extent the existing norms conform to it, as well as the legal responsibility of AI and the agents involved. To this end, issues such as legal liability in case of damages, the moral autonomy of systems, data protection, and transparency in automated decision-making are examined.

Through the Three-Dimensional Theory of Law, the analysis concludes that it is essential to create an international regulation based on universal ethical principles in accordance with Kantian ethics to ensure a fair and responsible use of AI. Therefore, it is proposed to create an International Treaty signed by the main technological powers, in which universal ethical principles are established as well as a common regulatory framework that includes the implementation of independent supervision and audit mechanisms to avoid any discrimination and ensure that the decisions of automated systems are fair.

KEY WORDS

Artificial Intelligence, Kantian ethics, human rights, legal responsibility, moral autonomy, damages, data privacy, transparency, harmonization.

ÍNDICE

INTRODUCCIÓN……7

CAPÍTULO I: ASPECTOS FORMALES Y METODOLÓGICOS……8

1. RELEVANCIA DEL TEMA……...8
2. METODOLOGÍA DE INVESTIGACIÓN…….8
3. ESTRUCTURA DEL TRABAJO……...9

CAPÍTULO II: INTRODUCCIÓN A LA ÉTICA KANTIANA ……10

1. CONTEXTO FILOSÓFICO E INTELECTUAL DE KANT……10
2. PRINCIPIOS FUNDAMENTALES DE LA ÉTICA KANTIANA……..…10
    a. **2.1 Autonomía y racionalidad**…….10
    b. **2.2 Imperativo categórico**……10
    c. **2.3 Libertad y responsabilidad** ……..11
    d. **2.4 El deber y la buena voluntad**……12
3. DIFERENCIAS ENTRE OTRAS TEORÍAS MORALES Y LA ÉTICA DEONTOLÓGICA DE KANT……..……12
4. IMPORTANCIA DE LA ÉTICA KANTIANA EN EL PANORAMA ACTUAL DE LA IA ……...…...13

CAPÍTULO III: CONCEPTUALIZACIÓN DE LA INTELIGENCIA ARTIFICIAL……...….15

1. DEFINICIÓN Y TIPOS DE IA…….15
2. EVOLUCIÓN HISTÓRICA Y CONTEXTO TECNOLÓGICO……...…...17
3. IA EN LA SOCIEDAD ACTUAL: APLICACIONES Y POTENCIAL ……..…18

CAPÍTULO IV: FUNDAMENTOS ÉTICOS EN EL USO DE IA…… 20

1. PRINCIPIOS ÉTICOS APLICABLES A LA IA…….20
2. DILEMAS MORALES EN LA AUTONOMÍA DE LA IA…….22
    a. **2.1 Problemas con los vehículos autónomos**……..…22

b. **2.2 Problemas en el ámbito sanitario**……..……23
c. **2.3 Problemas en el ámbito jurídico** ……..…… 25
3. RESPONSABILIDAD: ÉTICA: USUARIOS, DESARROLLADORES Y REGULADORES …… 26

CAPÍTULO V: DESAFÍOS JURÍDICOS EN LA REGULACIÓN DE LA IA…….27
1. AMBIGÜEDAD Y RIGIDEZ NORMATIVA……...….27
2. IDENTIDAD Y PERSONIFICACIÓN DE LA IA ……..…….28
3. DERECHOS DE PROPIEDAD INTELECTUAL Y LA IA ……...… 29
4. RESPONSABILIDAD CIVIL Y PENAL: ¿QUIÉN RESPONDE POR LOS ACTOS DE LA IA? 30

CAPÍTULO VI: IMPACTO DE LA IA EN LOS DERECHOS HUMANOS……...…… 32
1. PRIVACIDAD Y PROTECCIÓN DE DATOS PERSONALES ……...……32
2. DISCRIMINACIÓN Y SESGOS ALGORÍTMICOS ……...……...33
3. DERECHO A LA TRANSPARENCIA Y A LA EXPLICABILIDAD ……...……34

CAPÍTULO VII: IA Y EL DERECHO INTERNACIONAL…….…… 36
1. RETOS DE LA REGULACIÓN GLOBAL DE LA IA ……...36
2. INCERTIDUMBRE LEGAL POR FALTA DE UN MARCO NORMATIVO COMÚN 36
a. **2.1 Normativa vigente sobre IA**……..……36
i. 2.1.1 Normativa sobre IA en China …….…….36
ii. 2.1.2 Normativa sobre IA en Estados Unidos……37
iii. 2.1.3 Normativa sobre IA en la Unión Europea ……...38
b. **2.2 Falta de armonización y conflictos normativos**……..……39
c. **2.3 Propuesta de armonización regulatoria** ……..……40
3. INICIATIVAS INTERNACIONALES Y COOPERACIÓN INTERGUBERNAMENTAL ……40
a. 3.1 Iniciativas de la OCDE……..40
b. 3.2 Iniciativas de la ONU…….41
c. 3.3 Iniciativas del G20…….42

   d. **3.4 *Antiqua et Nova***……42
4. ARMONIZACIÓN NORMATIVA Y CONFLICTOS JURÍDICOS INTERNACIONALES ……..... 43

CAPÍTULO VIII: PROPUESTAS DE SOLUCIONES ÉTICO-JURÍDICAS…….. 44

1. PRINCIPIOS PROPUESTOS PARA UNA IA RESPONSABLE ……..….44
2. NORMATIVAS SUGERIDAS PARA UNA REGULACIÓN EFECTIVA……...45
3. ROL DE LAS ORGANIZACIONES Y LA SOCIEDAD CIVIL ……...…….46
4. RÉGIMEN DE RESPONSABILIDAD CIVIL Y PENAL……47
   a. **4.1 Responsabilidad de los desarrolladores de IA**…….47
   b. **4.2 Responsabilidad de los usuarios de IA**…….47
   c. **4.3 Posibles sanciones civiles y penales**……..48

CAPÍTULO IX: CONCLUSIONES……..……..49

1. REFLEXIONES FINALES SOBRE LOS PROBLEMAS ÉTICOS Y JURÍDICOS DE LA IA …….…...49
2. FUTURAS LÍNEAS DE INVESTIGACIÓN Y DESARROLLO EN IA ……50

BIBLIOGRAFÍA……..……...53

INTRODUCCIÓN

Este Trabajo de Fin de Grado se enmarca dentro del área de la Filosofía del Derecho y aborda los problemas éticos y jurídicos causados por el uso y desarrollo de la Inteligencia Artificial (IA) desde la perspectiva de la ética de Kant. La rápida evolución tecnológica ha derivado en la progresiva incorporación de la IA en distintos ámbitos de la sociedad, ocasionando retos significativos tanto desde el punto de vista de la ética filosófica como del normativo. Este trabajo se basa en la Teoría Tridimensional del Derecho, desarrollada por Miguel Reale, filósofo y jurista, para analizar los problemas de la IA en tres dimensiones distintas: materia, forma y finalidad, resaltando además la función clave de los agentes jurídicos como cuarto factor necesario para la ordenación y regulación de la IA.

En primer lugar, la materia de estudio es la IA y sus aplicaciones en diferentes sectores, desde el ámbito laboral hasta el sanitario o incluso el judicial, incluyendo la gestión de datos y la seguridad. La IA se ha configurado como un instrumento indispensable, pero su utilización también implica riesgos como la ausencia de transparencia en la toma de decisiones, sesgos algorítmicos o la vulneración de derechos fundamentales.

En segundo lugar, la forma se refiere al estudio de las normas jurídicas existentes en la regulación del uso de la IA. Pese al progreso en la legislación, como la Ley de IA de la Unión Europea, existen carencias normativas que obstaculizan una regulación efectiva y armonizada a nível global. Existen problemas de interpretación y aplicación en relación a las normas actuales que generan inseguridad jurídica y vacíos legales que pueden emplearse con fines inmorales. En este sentido es fundamental examinar la diferencia entre lo unívoco y lo equívoco en las normas: en tanto algunas disposiciones pueden ser claras y aplicables, otras son ambiguas o inadecuadas para hacer frente a los desafíos que proyecta la IA, lo que exige de una reforma urgente del marco jurídico.

Por último, la finalidad de este trabajo es asegurar el respeto a los derechos humanos en el desarrollo y uso de la IA. En sintonía con la ética de Kant, se busca que la IA no emplee a los individuos como instrumentos para lograr un fin, sino que se respete su dignidad y autonomía. Con esa finalidad se tratará de proponer una solución a los problemas éticos y jurídicos que se identifiquen a lo largo de este trabajo.

CAPÍTULO I: ASPECTOS FORMALES Y METODOLÓGICOS

1. RELEVANCIA DEL TEMA

Este tema es especialmente relevante debido a las aplicaciones de la IA en la vida cotidiana y los problemas que plantea. El mismo Sam Altman, CEO de Open AI, advertía de que con la IA "*debemos tener el mismo cuidado que con las armas nucleares*"[1]. Con el desarrollo de tecnologías como vehículos autónomos, algoritmos de recomendación en plataformas como instagram o tik tok o con sistemas de reconocimiento facial que se encuentran en diversos teléfonos móviles, es necesario establecer un marco normativo claro, conciso, transparente y ético que promueva el bienestar social. Pues como decía Idoia Salazar, cofundadora y presidenta del Observatorio del Impacto Social y Ético de la Inteligencia Artificial (OdiseIA), "*es una tecnología de gran impacto que necesita ser regulada para conseguir un impacto positivo*"[2]. Este trabajo explorará los problemas jurídicos y éticos derivados del uso y desarrollo de la IA desde un enfoque kantiano, analizando su adaptación a la ética del mismo y buscando soluciones normativas.

La ética de Kant fundamentada en el respeto a la dignidad humana y en principios universales brinda un marco sólido para llevar a cabo un análisis de los problemas que plantea la IA en la actualidad. En su ética el filósofo destacó el imperativo de fijar unos principios universales que garanticen el respeto a la dignidad humana. El filósofo sostuvo que los individuos no deben ser tratados como medios para lograr otros fines, sino como fines en sí mismos. Si se aplica este principio a la IA se manifiesta la necesidad de diseñar sistemas que promuevan el bienestar de todas las personas y el respeto a sus derechos fundamentales.

2. METODOLOGÍA DE INVESTIGACIÓN

[1] Montero Carretero, L, "*El fin de la era del todo vale para la IA"*, ABC Empresa, domingo 28 de mayo de 2023, pp.3
[2] Montero Carretero, L, "El fin de la era del todo vale para la IA", ABC Empresa, domingo 28 de mayo de 2023, pp.3

Para el desarrollo de este trabajo se emplea una metodología de investigación basada tanto en el análisis de la doctrina y normas relativas a la IA, como en el análisis de obras fundamentales de Immanuel Kant con el fin de proporcionar soluciones meditadas y contrastadas a los problemas morales involucrados. En añadidura, se revisan fuentes secundarias como informes técnicos de Comisiones Europeas, literatura académica, marcos regulatorios, principalmente los de la Unión Europea, así como casos de especial relevancia en este ámbito.

El enfoque del estudio es descriptivo-análitico, centrándose en identificar problemas éticos y jurídicos concretos y analizando el marco regulatorio en el que se encuentran, evaluando las soluciones propuestas actualmente y proponiendo nuevos recursos de actuación. Se trata de alcanzar una comprensión global combinando para ello tanto la teoría como la práctica.

3. ESTRUCTURA DEL TRABAJO

Este trabajo se divide en nueve apartados y una bibliografía en los cuales se analizará qué es la IA, cuáles son sus fundamentos y las normas que la regulan, qué desafíos presenta y qué impacto tiene en la sociedad actual, combinando todo ello con la ética kantiana. Para ello se estudiarán casos de especial relevancia en el ámbito y a partir de ellos se tratará de aportar soluciones ético-jurídicas a los problemas planteados.

CAPÍTULO II: INTRODUCCIÓN A LA ÉTICA KANTIANA

1. CONTEXTO FILOSÓFICO E INTELECTUAL DE KANT

Immanuel Kant (1724-1804) es uno de los filósofos modernos más relevantes. En su obra trató de sintetizar empirismo y racionalismo en un sistema filosófico que fijase los límites y fundamentos del conocimiento, la estética y la moral. Todo esto lo recogió principalmente en sus obras "Crítica a la razón pura" y "Fundamentación de la metafísica de las costumbres" en las que expresa su deseo de establecer unos principios morales de carácter universal que se basen en la razón.

2. PRINCIPIOS FUNDAMENTALES DE LA ÉTICA KANTIANA

La ética kantiana tiene como base el imperativo categórico que consiste en una ley universal sobre la moral que debe dirigir toda acción humana y cuyos principios fundamentales son:

○ 2.1 Autonomía y razón

"Una voluntad que no puede ser determinada más que a través de estímulos sensibles, es decir, patológicamente, es una voluntad animal (arbitrium brutum). La que es, en cambio, independiente de tales estímulos y puede, por tanto, ser determinada a través de motivos sólo representables por la razón, se llama voluntad libre (arbitrium liberum), y todo cuanto se relaciona con esta última, sea como fundamento, sea como consecuencia, recibe el nombre de práctico."[3]

Kant considera que los seres humanos son capaces de legislar moralmente para sí mismos actuando conforme a principios universales. Considera que la moral se basa en la razón pura y no en los sentimientos o la experiencia. La autonomía moral, conforme a Kant, consiste en la capacidad de la persona de no actuar por inclinaciones personales sino por deber.

[3] Kant, I. (1781). *Crítica de la razón pura* (A. González, Trans.). Editorial Losada, Buenos Aires, 2006, pp. 455-456.

- **2.2 El imperativo categórico**

Kant manifiesta que el imperativo categórico es el principio fundamental por el que deben regirse los deberes morales de todos los hombres. Explica que se trata de un imperativo ya que el deber limita la voluntad e impone la acción moral, y es categórico porque para que la acción sea moral debe ser un fin en sí misma y no un medio para lograr otro fin.

Este imperativo categórico se manifiesta a través de distintas formulaciones:

1. "*Obra sólo según aquella máxima por la cual puedas querer que se convierta en una ley universal*"[4] Con esta formulación Kant explica que nuestro actuar debe poder tomarse como ley universal aplicable a toda la humanidad en todos los contextos. En caso de no ser así, esa forma de actuar resultaría moralmente inválida.
2. "*Obra de tal modo que uses a la humanidad, tanto en tu persona como en la de cualquier otro, siempre como un fin y nunca como un medio*"[5] Esta es una de las formulaciones más importantes de Kant e invita a respetar siempre la dignidad del ser humano considerándolo un fin en sí mismo y no un medio para lograr otro fin.
3. "*Obra como si mediante tus máximas fueras siempre un legislador en un reino de los fines*"[6] Con esto el filósofo quiere decir que todo individuo debe considerarse legislador y destinatario de las normas morales por lo que todos sus actos serán contribuciones a dicho orden moral.
4. "*Obra como si fueras, por medio de tus máximas, un miembro legislador en un reino universal de fines*" [7] Con esto Kant desarrolla la formulación anterior explicando que todo ser racional debe actuar como si integrase un sistema moral ideal en el cual cada persona es tratada con dignidad y respeto.

Con cada una de estas formulaciones Kant trata de asegurar una moralidad universal y objetiva que se distancie de intereses particulares y del relativismo.

---

[4]Kant, I. (1785). *Fundamentación de la metafísica de las costumbres* (A. González, Trans.). Editorial Losada, Buenos Aires, 2002, p. 421.
[5]Kant, I. (1785). *Fundamentación de la metafísica de las costumbres* (A. González, Trans.). Editorial Losada, Buenos Aires, 2002, p. 429.
[6] Kant, I. (1785). *Fundamentación de la metafísica de las costumbres* (A. González, Trans.). Editorial Losada, Buenos Aires, 2002, p. 433.
[7] Kant, I. (1785). *Fundamentación de la metafísica de las costumbres* (A. González, Trans.). Editorial Losada, Buenos Aires, 2002, p. 438.

- **2.3 Libertad y responsabilidad**

"*La libertad y la voluntad propia bajo leyes morales son una misma cosa.*"[8]

Kant en su concepción de la moralidad, considera que la libertad y la responsabilidad son inseparables. Estima que actuar libremente no supone actuar conforme a caprichos personales sin límites, sino actuar conforme a la razón y bajo principios universales. Sostiene que la verdadera libertad se halla en cumplir con el imperativo categórico, por el cual debemos actuar de manera que nuestras acciones puedan convertirse en ley universal. En otras palabras, que nuestras acciones puedan ser adoptadas por todas las personas sin contradicción. Al actuar de manera libre y conforme a la razón, nos hacemos responsables de nuestras acciones, ya que decidimos actuar según la ley moral. Si alguna acción no puede universalizarse, entonces es inmoral.

- **2.4 El deber y la buena voluntad**

"No consideraremos los mandamientos como obligatorios por ser mandamientos de Dios, sino que los consideraremos mandamientos de Dios por constituir para nosotros una obligación interna."[9]

El concepto del deber es clave en la ética kantiana, dado que la moralidad depende del cumplimiento del deber por respeto a la ley moral y no de consecuencias o inclinaciones.

Lo único que puede ser bueno irrestringiblemente, conforme a kant, es la buena voluntad que es aquella que actúa según el deber y por respeto a la ley moral en vez de por intereses ajenos o inclinaciones personales.

3. DIFERENCIAS ENTRE OTRAS TEORÍAS MORALES Y LA ÉTICA DEONTOLÓGICA DE KANT

[8]Kant, I. (1788). *Crítica de la razón práctica* (A. González, Trans.). Editorial Losada, Buenos Aires, 2002, p. A33.
[9]Kant, I. (1781). *Crítica de la razón pura* (A. González, Trans.). Editorial Losada, Buenos Aires, 2006, pp. 463.

La ética deontológica de Kant se aleja de otras grandes teorías acerca de la moral como pueden ser el utilitarismo de Bentham o la ética de Aristóteles. Sandel en su obra "*Justice: What's the Right Thing to Do?"*[10] comparó las diferencias entre estas teorías .

La doctrina utilitarista de Bentham considera que la justicia se basa en lograr el bienestar del máximo número de personas y, por tanto, la moralidad consiste en alcanzar la felicidad del máximo número de personas. Por su lado Kant sostiene que la moralidad depende de su conformidad con el deber y no de las consecuencias de un acto y considera que la doctrina utilitarista atenta contra la autonomía individual y la dignidad humana al implicar que algunos individuos tengan que hacer sacrificios personales para el bienestar general.

Si se aplica el utilitarismo de Bentham al uso y desarrollo de IA como principio que lo guíe, se concluiría que la IA debiera maximizar los beneficios sociales, lo que supondría sacrificios individuales.

Por su lado, la ética aristotélica se basa en la eudaimonía (felicidad) como objetivo de la vida moral. Considera que la moralidad consiste en cultivar virtudes y buscar la excelencia en la vida del ser humano. Kant por su parte considera que las virtudes son deberes racionales independientes del bienestar personal y considera que la moralidad proviene del cumplimiento del deber y de aplicar el imperativo categórico.

Se distancia también de teorías contractualistas que basan la moralidad en pactos sociales como Rousseau o Hobbes, considerando que la moral se fundamenta en la razón pura y no en acuerdos entre individuos.

En conclusión, la ética de Kant se diferencia de otras teorías por su empeño en la universalidad, la autonomía moral y la racionalidad, que la distancia de teorías basadas en acuerdos sociales, la experiencia y las consecuencias.

4. IMPORTANCIA DE LA ÉTICA KANTIANA EN EL PANORAMA ACTUAL DE LA IA

[10]Sandel, M. J., *Justice: What's the Right Thing to Do?*, Farrar, Straus and Giroux, Nueva York, 2009.

La ética kantiana es relevante en el panorama actual tanto en lo relativo a los derechos humanos como a la inteligencia artificial debido a su énfasis en la dignidad humana y la autonomía moral que sientan las bases para principios éticos que puedan regir en un marco normativo regulatorio de la misma.

La aplicación de la ética kantiana al campo de la inteligencia artificial implica que la misma deba ser usada y diseñada de forma que respete la autonomía y dignidad humana. Esto quiere decir que debe operar conforme a principios universales, como ejemplo los algoritmos que toman decisiones deben asegurar la transparencia e imparcialidad en su funcionamiento. Deben tratar a las personas como fines en sí mismas y no como meros medios para lograr otro fin, por lo que la recabación de datos personales debe hacerse conforme a los derechos fundamentales de las personas. Más allá y dado que la IA carece de autonomía moral, no debe considerarse como un ser moral, pero quienes la usen y programen si, por lo que ellos serán responsables de asegurar que su uso se haga conforme a principios éticos universales.

En conclusión, la ética kantiana brinda un marco estable para abordar problemas morales en el ámbito de la IA, aportando principios fundamentales que guíen el desarrollo y uso de la misma, como son: la universalidad, la dignidad humana y la autonomía.

CAPÍTULO III: CONCEPTUALIZACIÓN DE LA INTELIGENCIA ARTIFICIAL (IA)

1. DEFINICIÓN Y TIPOS DE IA

El grupo de expertos de alto nivel sobre inteligencia artificial de la Comisión Europea define la Inteligencia Artificial como "*sistemas diseñados por humanos que, dado un objetivo complejo, actúan en el mundo físico o digital percibiendo su entorno, interpretando los datos estructurados o no estructurados recopilados, razonando sobre el conocimiento derivado de estos datos y decidiendo las mejores acciones a tomar (según parámetros predefinidos) para lograr el objetivo dado. Los sistemas de IA también pueden diseñarse para aprender a adaptar su comportamiento analizando cómo el entorno se ve afectado por sus acciones anteriores. Como disciplina científica, la IA incluye varios enfoques y técnicas, como el aprendizaje automático (del cual el aprendizaje profundo y el aprendizaje por refuerzo son ejemplos específicos), el razonamiento automático (que incluye planificación, programación, representación y razonamiento del conocimiento, búsqueda y optimización) y la robótica (que incluye control, percepción, sensores y actuadores, así como la integración de todas las demás técnicas en sistemas ciberfísicos)*".[11]

Existen distintos tipos de IA que se clasifican en base a su nível de autonomía y su capacidad de aprendizaje. Podemos distinguir:

- IA reactiva, que se refiere a IA que responde a entradas concretas y carece de memoria o aprendizaje como puede ser Deep Blue, un ordenador diseñado para ganar a los hombres en el juego del ajedrez, que analiza los posibles movimientos y escoge la mejor opción sin aprender de otras partidas. En "Crítica a la razón pura" Kant señala que la razón es limitada y que el conocimiento se construye partiendo de categorías innatas en la mente de las personas. La IA reactiva no reflexiona ni razona acerca de sus decisiones, se limita a seguir una serie de cálculos predefinidos, siendo incapaz de trascender su programación. Desde la perspectiva de la ética kantiana, esta IA carece de autonomía

[11]European Commission, *High-Level Expert Group on Artificial Intelligence: Definition of Artificial Intelligence*, disponible en https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf; última consulta 31/03/2025.

moral, dado que no funciona por deber o reflexiona sobre principios morales limitándose a ejecutar respuestas predeterminadas.

- IA con memoria limitada que utiliza datos anteriores para decisiones próximas al igual que los sistemas de aprendizaje automático, como por ejemplo los vehículos autónomos que se pueden adaptar a las normas de circulación y a las condiciones del tráfico y en base a ello tomar decisiones sobre cómo realizar su trabajo con mayor efectividad. Para Kant, la razón humana no se limita a aprender de la experiencia, como la IA con memoria limitada, sino que también la interpreta y la somete a juicios morales basados en principios universales. En otras palabras, la IA con memoria limitada carece de razón crítica como la que poseen los seres humanos. Esta IA puede predecir y optimizar su comportamiento, pero no tiene la capacidad de realizar juicios normativos ni actuar por deber.
- IA de teoría de la mente que tiene la capacidad de interpretar las emociones de los seres humanos y actuar conforme a ellas. Es un gran avance para la simulación de la inteligencia emocional, pero en la actualidad permanece en investigación. La inteligencia emocional permite a los seres humanos actuar de forma empática y comprender las motivaciones ajenas, lo que supone respetar la dignidad de los demás y que es la base de la teoría moral kantiana. No obstante, una IA con inteligencia emocional actúa conforme a parámetros programados por seres humanos y no puede lograr una comprensión autónoma del deber moral. Kant considera que la ética no puede limitarse a simples interpretaciones de emociones, sino que precisa una voluntad libre con capacidad para reconocer y actuar conforme a imperativos categóricos.
- IA autoconsciente, que es el nível más alto de desarrollo de IA en la actualidad y que posee conciencia propia, distanciándose del resto de tipos de IA que solo poseen inteligencia y aprendizaje. Conforme a la ética kantiana una IA autoconsciente carecería de voluntad propia aunque pudiera reconocer su existencia y cambiar su comportamiento en base a ello. Esto se debe a que no podría actuar por deber ni formular principios universales, dado que su funcionamiento seguiría partiendo de parámetros predeterminados por seres humanos. Puesto que sigue sujeta a los límites de su programación inicial, no puede legislarse a sí misma conforme a la razón, por lo que carece de autonomía moral.

En suma, desde la perspectiva de la ética kantiana, todo tipo de IA carece de autonomía moral, dado que ninguna posee voluntad propia ni capacidad para obrar por deber, que son elementos claves en la ética del filósofo. La IA es determinada por reglas externas y no es capaz de alcanzar la autonomía práctica, mientras que la razón humana se rige por principios universales y es capaz de trascender su experiencia para formular juicios éticos. A pesar de los avances en inteligencia emocional, la diferencia entre la inteligencia emocional y artificial sigue siendo clara por la falta de la IA de racionalidad práctica y su incapacidad para actuar según principios morales incondicionales.

2. EVOLUCIÓN HISTÓRICA Y CONTEXTO TECNOLÓGICO

Alan Turing fue la primera persona en plantearse la posibilidad de la existencia de la IA al escribir en 1950 en un papel que se podía crear una máquina capaz de pensar, introduciendo así el Test de Turing como criterio evaluador de la IA. Posteriormente, en 1956 en la Conferencia de Dartmouth se formalizó el campo de la IA de la mano de personas a quienes en adelante se les conocerá como los Padres de la IA. Algunas figuras que destacan son McCarthy que planteó la posibilidad de usar la lógica para resolver problemas y Weizenbaum que creó uno de los primeros Chatbots al que llamó Eliza.

A pesar de unos primeros años de auge en el desarrollo de la IA, la misma sufrió un primer invierno como consecuencia de expectativas no alcanzadas, pero resurgió en la década de los 80 de la mano de los sistemas expertos que condujeron a un boom de la IA entre 1980 y 1987. Desde entonces surgieron descubrimientos como la Paradoja de Moravek o la Nouvelle IA que promovieron el desarrollo de la IA hasta nuestros días.[12]

Este avance alcanzó especial relevancia con la llegada de Chat GPT, considerado actualmente como uno de los impulsores de la Cuarta Revolución Industrial.

[12] European Union, *AI and Human Rights*, curso digital, disponible en https://help.elearning.ext.coe.int/user/view.php?id=210138&course=7331

Este progreso histórico de la IA puede enmarcarse en la idea kantiana de la minoría y mayoría de edad. Kant consideraba que la minoría de edad intelectual tenía lugar cuando el hombre no pensaba de forma autónoma sino que se dejaba guiar por otros. En cambio, cuando el ser humano se atrevía a pensar por sí mismo alcanzaba la mayoría de edad expresada en su célebre lema "sapere aude" (atrévete a pensar). Esto puede entenderse de manera similar sobre la evolución de la IA que en un primer momento se encontraba en su minoría de edad ya que dependía totalmente de reglas estrictas y carecía de capacidad de adaptación. No obstante, con los avances en modelos de lenguaje y aprendizaje automático, está empezando a desarrollar una mayor autonomía operativa, acercándose a una aparente mayoría de edad tecnológica, peor que no alcanza la autonomía moral.

3. IA EN LA SOCIEDAD ACTUAL: APLICACIONES Y POTENCIAL

En la actualidad la IA se ha involucrado en la mayoría de ámbitos de la vida cotidiana, llevando a cabo una transformación de los sectores más importantes que con su aplicación han visto incrementada su eficiencia, la toma de decisiones y la experiencia de los usuarios. Las principales áreas en las que se ha aplicado la IA son:

- Finanzas: en la actualidad la IA se ha aplicado al campo de las finanzas con la finalidad de detectar fraudes, predecir distintas tendencias del mercado y evaluar riesgos de créditos aportando soluciones más dinámicas y precisas.
- Justicia: en el área de la justicia se utiliza la IA para buscar jurisprudencia en despachos de abogados así como algoritmos de IA que permiten evaluar los riesgos delictivos así como asignar recursos judiciales.
- Salud: se han desarrollado instrumentos de IA capaces de analizar imágenes médicas que pueden detectar anomalías, como tumores, que permiten a los médicos realizar un diagnóstico con mayor agilidad y precisión.
- Consumo y servicios: en este campo destacan asistentes virtuales como pueden ser Alexa o Siri que interactúan de manera personalizada con los usuarios y sistemas de recomendación en línea que facilitan la experiencia de comprar en plataformas digitales.

- Transporte: en este campo destacan los vehículos autónomos que a través de redes neuronales y sistemas profundos de aprendizaje permiten mejorar la seguridad vial y mejorar la movilidad urbana.
- Creación artística: se aplica en el ámbito artístico tanto para literatura, como la pintura, periodismo o la música.

Todos estos descubrimientos y desarrollos de la IA permiten un avance en los distintos campos pero al mismo tiempo plantean graves problemas éticos así como de responsabilidad civil que plantean serias preocupaciones. Estos problemas suscitan la necesidad urgente de desarollar marcos regulatorios que proporcionen un equilibrio entre la innovación de la IA y la responsabilidad ética. De acuerdo con la ética kantiana la IA deberá emplearse de forma que se respete la autonomía moral humana, evitando que se transforme en un instrumento de discriminación o control que considere a las personas como simples medios para lograr un fin.

CAPÍTULO IV: FUNDAMENTOS ÉTICOS EN EL USO DE IA

1. PRINCIPIOS ÉTICOS APLICABLES A LA IA

La UNESCO publicó la "Recomendación sobre la Ética de la Inteligencia Artificial"[13] en la que establece los principios éticos fundamentales que deben guiar el desarrollo y uso de la IA. Estos principios se alinean con la ética kantiana al buscar un equilibrio entre la dignidad humana y la tecnología. Los principales son:

- Transparencia y explicabilidad: conforme a este principio los usuarios y reguladores deben ser capaces de comprender los sistemas de IA y los motivos detrás de sus decisiones. La transparencia en este sentido es fundamental para que los usuarios puedan supervisar las decisiones automatizadas que afectan a sus derechos y libertades.

  En la ética de Kant se establece que las decisiones deben ser justificadas y comprensibles de manera racional. De esta forma, los sistemas de IA tienen que ser transparentes, para comprobar que no toman decisiones sesgadas o arbitrarias.

- Proporcionalidad e inocuidad: los sistemas de IA deben diseñarse de manera que no causen ningún daño, asegurando que sus impactos sean en todo caso proporcionales al objetivo buscado. Se deben priorizar el bienestar, la salud y la seguridad de los individuos.

  Conforme al principio de beneficencia de Kant, todas las acciones deben buscar el bienestar general y ser moralmente correctas, lo que se refleja en el diseño de sistemas de IA que da prioridad al bienestar y la protección frente a posibles daños.

- Rendición de cuentas y responsabilidad: conforme a este principio, los diseñadores, operadores y usuarios de sistemas de IA, deberán responder de los daños causados por la misma. Esto incluye rendir cuentas por las decisiones tomadas por sistemas automatizados cuando vulneran los derechos de un individuo.

[13] UNESCO, *Recomendación sobre la ética de la inteligencia artificial*, 2021 (disponible en https://unesdoc.unesco.org/

En su ética, Kant estableció que para que una acción fuese moral esta debía poder tomarse como ley universal para todos y en todas circunstancias, y que en caso de no serlo, cada individuo debía ser responsable de sus decisiones y acciones. Si se aplica esto al contexto de la IA, quienes desarrollan los sistemas de IA son responsables de los efectos de la misma, debiendo asegurarse de que sus decisiones sean conformes a principios universales de respeto de la dignidad humana y justicia.

- Privacidad y protección de datos: la recopilación y uso de datos debe hacerse conforme a principios éticos y claros, manejándose de manera responsable y respetando el derecho a la privacidad de los individuos.

  Para Kant la privacidad y la autonomía individual eran claves para el respeto de la dignidad humana. Conforme a ello, la recopilación de datos debe hacerse de manera que se respete la autonomía y libertad de cada individuo. Por tanto, los sistemas de IA deben diseñarse de manera que no vulneren la privacidad de sus usuarios.

- No discriminación: se debe garantizar que los sistemas de IA no hayan sido diseñados con ningún sesgo o discriminación para asegurar un trato justo y equitativo a todas las personas, indistintamente de su sexo, raza, religión o cualquier otro rasgo personal.

  En su ética, Kant enfatizó el principio de que toda persona debe ser tratada como un fin y nunca como un medio, respetando su dignidad humana, lo que aplicado a la IA implica que no se refuercen desigualdades mediante la reproducción de sesgos, garantizando decisiones equitativas y justas.

Lograr el equilibrio entre todos estos efectos no siempre resulta fácil, dado que es posible que entren en conflicto unos con otros en ciertas situaciones. A modo de ejemplo, la privacidad puede entrar en conflicto con la transparencia si el describir un algoritmo implica revelar información sensible. Por ende, es trascendental que los reglamentos sobre IA definan pautas concretas para ordenar y priorizar la ética en sus diseños y despliegues.

Desde la perspectiva de la ética kantiana, toda regulación o desarrollo de la IA debe asegurar que los hombres sean considerados no como simples medios para lograr otros propósitos sino como

fines en sí mismos, garantizando que el desarrollo de la tecnología se lleve a cabo respetando la dignidad humana y la justicia universal.

2. DILEMAS MORALES EN LA AUTONOMÍA DE LA IA

Existen complejos dilemas morales relacionados con la autonomía de los sistemas de IA. Estos se dan especialmente en circunstancias en las que la toma de decisiones puede tener repercusiones graves.

○ 2.1 Problemas con los vehículos autónomos

Este es el caso de los vehículos autónomos, es decir, que carecen de conductor. En situaciones de emergencia estos vehículos deben priorizar la vida de los hombres. La cuestión que se plantea es cómo deben programarse las prioridades morales en los algoritmos de manera que se respeten los principios éticos. Desde la óptica de la ética kantiana, toda decisión debe tomarse respetando el principio de que los individuos deben tratarse como fines en sí mismos y no como meros medios.

En 2018 se llevó a cabo una prueba de un uber de conducción autónoma que contaba con la presencia de un back up driver en la parte de atrás. Este vehículo atropelló a Elaine Herzberg provocándole heridas que causaron su muerte. En este caso el back up driver era el encargado de tomar el control del vehículo en caso de que el mismo tomara una decisión incorrecta que pusiera en peligro la vida de cualquier ciudadano. La investigación concluyó que el back up driver no estaba prestando atención en el momento del accidente.

Este fue el primer atropello mortal causado por un vehículo autónomo y un caso perfecto para el estudio de los dilemas de responsabilidad derivados de los daños causados por IA.

El back up driver sostuvo que el vehículo cumplía con los límites de seguridad y que la mujer se abalanzó sobre la carretera sin que le diera tiempo de actuar y frenar el vehículo. Deefendió que evitar dicho accidente habría sido complicado aunque el vehículo hubiera sido conducido por una persona física. Este caso esta pendiente de fallo.

Este caso plantea uno de los dilemas legales más complejos de la IA más allá de la ética. Lo que se plantea es quién es responsable civil o penalmente de este homicidio imprudente. Por un lado podría ser el error humano cometido por el back up driver que no evitó el accidente por una distracción. Por otro lado, podría ser quién diseñó el sistema automatizado que no aseguró la seguridad de los peatones o la propia compañía de Uber, a pesar de haber quedado liberada de toda responsabilidad criminal por los fiscales.

A continuación se incluye un fragmento de la reconstrucción del accidente:

> El sistema de conducción de Uber - que en ese momento llevaba 19 minutos controlando el coche - registró un vehículo que circulaba por delante a 5,6 segundos de distancia, pero no alertó a Vásquez. A continuación, el ordenador anuló su evaluación inicial: no sabía qué objeto era. A continuación, volvió a cambiar la clasificación a un vehículo, y luego osciló entre vehículo y otra cosa. A 2,6 segundos del objeto, el sistema lo identificó como "bicicleta". A 1,5 segundos, volvió a considerarlo otra cosa. Luego volvió a bicicleta. El sistema generó un plan para intentar esquivar lo que fuera, pero decidió que no podía. Entonces, a 0,2 segundos del impacto, el coche emitió un sonido para alertar a Vásquez de que el vehículo iba a reducir la velocidad. Dos centésimas de segundos antes del impacto, a 39 Mp/h, Vásquez agarró el volante, lo que sacó al coche de la autonomía y lo puso en modo manual. Era demasiado tarde. La bicicleta destrozada dejó una estela de 7 metros en la acera. Una persona yacía gravemente herida en la calzada (Smiley, 2022). [14]

Uber explicó que ellos mismos habían desactivado el sistema de seguridad de Volvo (el modelo del coche) para evitar cualquier interferencia con su propia tecnología. No obstante, su sistema de IA nunca identificó a Elaine como una persona, cambiando su calificación y recalculando su trayectoria. Además, Uber había programado el coche para que retrasara las frenadas bruscas durante un segundo evitando falsas alarmas, lo que supuso que el coche no reaccionara a tiempo. Volvo llevó a cabo simulaciones en la que demostró que su sistema de frenado habría evitado el accidente en 17 de cada 20 casos o al menos reducido los daños.

El problema ético que se plantea es quién es responsable Uber o el back up driver, o quien permitiera que se llevasen a cabo este tipo de pruebas en lugares públicos donde es difícil garantizar la seguridad de los ciudadanos.

[14]Smiley, A., “Uber's Self-Driving Car Didn't Know Pedestrians Could Jaywalk”, *The Atlantic*, 2022 (disponible en https://www.theatlantic.com.

- **2.2 Problemas en el ámbito sanitario**

Otro de los dilemas que plantea la autonomía de la IA es su aplicación en el sector de la salud. Esta ha revolucionado este campo haciendo más sencillos los diagnósticos médicos, personalizando la atención a los pacientes y optimizando algunos tratamientos. Esto se lleva a cabo mediante algoritmos que analizan imágenes y detectan enfermedades con mayor precisión. Sin embargo, el uso de IA en medicina plantea diversos dilemas éticos. Estos son:

1. Autonomía del paciente: puede verse comprometida si los profesionales sanitarios se apoyan de manera excesiva en los sistemas automatizados de IA sin llevar a cabo un análisis crítico de los resultados. En la ética de kant se destaca la importancia de respetar la autonomía individual y por ello el uso de IA en la salud no debe en ningún caso comprometer la misma mediante la falta de reflexión ética de los médicos y no basada únicamente en cálculos de algoritmos.[15]
2. Privacidad de los datos médicos: para desarrollar sistemas de IA de reconocimiento o diagnóstico de enfermedades se requiere primero el análisis de numerosas imágenes para entrenar a los sistemas. Esto puede suponer riesgos de violacion de confidencialidad de los pacientes. La IA debe garantizar la protección de los datos médicos de los pacientes ajustándose a protocolos exigentes de seguridad y consentimiento informado.
3. Responsabilidad médica: si la IA lleva a cabo un diagnóstico erróneo de una enfermedad o una recomendación inadecuada de un tratamiento, se plantea el dilema de quien tiene la responsabilidad de ese error: quien ha desarrollado el software, el hospital que decidió implementarlo o el médico que confió en la IA.

Un ejemplo del caso es el de la supercomputadora Watson de IBM, programada para recomendar tratamientos oncológicos que en numerosas ocasiones ha cometido errores recomendando tratamientos inseguros o incorrectos.

El origen de estos problemas reside en el entrenamiento que recibió Watson en el que en vez de emplear datos de pacientes reales el software se llevó a cabo con limitados casos de cáncer hipotéticos y sintéticos. Más allá de esto, las recomendaciones se hicieron en base a la

[15] Lazcoz Moratinos, G., "Sistemas de Inteligencia Artificial en la asistencia sanitaria: cómo garantizar la supervisión humana desde la normativa de protección de datos", *Centro de Investigación Biomédica en Red (CIBERER - ISCIII) Instituto de Investigación Sanitaria Fundación Jiménez Díaz (IIS-FJD)*

experiencia de ciertos especialistas del Memorial Sloan Kettering Cancer Center, en vez de conforme a guías más extensas.

Pese a estos defectos, los altos mandos de IBM promovieron públicamente a Watson for Oncology como instrumento entrenado con casos de pacientes reales y haber sido reconocido por médicos mundialmente. No obstante, las evaluaciones internas y las críticas de clientes combaten estas declaraciones, afirmando que las recomendaciones frecuentemente inconcretas hacen surgir dudas sobre la tecnología y su proceso de creación.[16]

Este caso refleja la importancia de crear sistemas automatizados transparentes, explicables y auditables en el campo de la salud para evitar tratamientos erróneos que puedan causar daños a los pacientes. Además, surge de nuevo el dilema de quién responde por los daños causados por este diagnóstico erróneo, el médico que se fió de la IA, quien la desarrolló o quien permitió que se implementase en el Hospital. [17]

- **2.3 Problemas en el ámbito jurídico**

Otro de los dilemas éticos que surgen en relación a la autonomía de la IA es su aplicación en el sector justicia tanto para analizar jurisprudencia como para evaluar los riesgos de un sujeto a reincidir en crímenes o asistir en la toma de decisiones judiciales. No obstante, la automatización de este sector proyecta diversas preocupaciones jurídico-éticas.

Un ejemplo actual es un sistema, COMPAS, desarrollado en Estados Unidos que se emplea para predecir la probabilidad de que un criminal reincida. Ha sido demostrado por diversos estudios que este sistema de IA manifiesta sesgos raciales que afectan a las decisiones de sentencias de libertad condicional, calificando a personas de origen afroamericano como más probables a reincidir que a cualquier otro grupo.[18]

---

[16] Strickland, E., "IBM Watson, Heal Thyself: How AI Fails in Health Care", *IEEE Spectrum*, 2019 (disponible en https://spectrum.ieee.org.

[17]*Salud y Fármacos*, "La regulación de la inteligencia artificial en la asistencia sanitaria: aspectos éticos y legales", *Salud y Fármacos*, noviembre 2018 (disponible en https://www.saludyfarmacos.org/lang/es/boletin-farmacos/boletines/nov201804/20_la/.

[18]Angwin, J., Larson, J., Mattu, S., & Kirchner, L., "Machine Bias", *ProPublica*, 2016 (disponible en https://www.propublica.org

Si se aplica la ética kantiana a este contexto se puede ver claramente los problemas de esta IA que atenta contra la dignidad humana y que al contrario de lo que defendía el filósofo no trata a todos los individuos por igual y priva a los ciudadanos de juicios transparentes y éticos.

Con el fin de lograr un uso ético de este tipo de IA es necesario que estos sistemas sean totalmente auditables, con algoritmos explicables y cuyas decisiones finales esten en manos de un juez humano para evitar la deshumanización de estos procesos.

3. RESPONSABILIDAD ÉTICA: USUARIOS, DESARROLLADORES Y REGULADORES

En el uso de la IA la responsabilidad ética se reparte entre los usuarios, los desarrolladores y los reguladores. Los primeros deben usar la IA de forma responsable, no empleandola para fines que vulneren los derechos humanos. En cuanto a los segundos, estos deben asegurarse de que los sistemas sean justos, seguros y acordes a principios éticos universales. En cuanto a los reguladores, su función es establecer un marco legal claro que asegure el cumplimiento de las normas éticas. Conforme a la ética de Kant, se deben instaurar leyes universales que salvaguarden la dignidad humana y la autonomía moral.

Estos tres usuarios deberán colaborar con el fin de minimizar los riesgos propios del empleo de la IA, garantizando que el desarrollo de la tecnología esté alineado con los derechos fundamentales y los valores éticos.

La responsabilidad ética es compartida por los desarrolladores, que deben crear sistemas seguros y justos; los usuarios, que deben emplear la IA de manera responsable; y los reguladores, que tienen la tarea de establecer normas claras y eficaces. La colaboración entre estos actores es esencial para mitigar los riesgos y maximizar los beneficios de la IA, asegurando su alineación con los valores democráticos y los derechos humanos.

CAPÍTULO V: DESAFÍOS JURÍDICOS EN LA REGULACIÓN DE LA IA

El avance de la IA ha generado distintos problemas jurídicos que exigen de una regulación clara y efectiva para asegurar un uso responsable de la misma. Los problemas más relevantes que han surgido son la ambigüedad y rigidez de las normas, la atribución de derechos a las IA, la determinación de la responsabilidad legal, el riesgo de discriminación y sesgo y los efectos negativos que pueda tener en derechos fundamentales como la seguridad o la privacidad.

En su ética Kant destacó la necesidad de establecer parámetros éticos claros y universales que en este caso implican establecer un marco regulatorio de la IA que se base en principios universales que deben respetar la autonomía moral de las personas y la dignidad humana. Conviene recordar que Kant sostenía que las leyes debían realizarse considerando a los individuos como fines en sí mismos y no como simples medios para lograr un fin. Por ello, todo marco normativo relativo a la IA debe dar prioridad a la protección de los derechos humanos, así como evitar que los sistemas automatizados instrumentalicen a los individuos. En este contexto de los derechos humanos, conviene recordar que la primera y segunda generación se centraba en derechos individuales y políticos, mientras que la tercera generación comprende derechos colectivos como el medio ambiente o la paz. Esta última ha colapsado en muchos puntos por lo que surgió la Agenda 2030 y los ODS para organizar una respuesta global. La IA podría enmarcarse dentro de esta tercera generación de derechos humanos debido a su impacto en el desarrollo y bienestar general, pero su continua evolución plantea si aún podría regularse de manera ética.

1. AMBIGÜEDAD Y RIGIDEZ NORMATIVA

Uno de los problemas existentes en la regulación de la IA se encuentra en las normas que la comprenden. Desde la perspectiva de la interpretación jurídica, se pueden diferenciar normas unívocas que son las que tienen una interpretación clara, única y determinada, y normas equívocas que tienen interpretaciones diversas o que provocan ambigüedad[19]. En la regulación de la IA se pueden observar ambos problemas.

[19]Kelsen, H., *Teoría pura del derecho* (2ª ed.), Ediciones Jurídicas Euroamérica, Buenos Aires, 2003

En primer lugar, existen normas equívocas que utilizan términos vagos como “transparencia algorítmica” o “IA confiable” como en el Reglamento de la UE sobre IA en el que hace referencia a la necesidad de que la IA sea “transparente” y “confiable” pero no dispone criterios concretos para evaluar estos requisitos. Estas normas son equívocas porque no definen con claridad qué es una IA “transparente” y “confiable” lo que genera incertidumbre jurídica porque cada operador jurídico puede realizar una interpretación distintas de lo que se considera aceptable en esos términos.

En segundo lugar, existen también normas unívocas en el contexto de la regulación de la IA en el establecimiento de requisitos categóricos o estrictos. Un ejemplo de ello es la clasificación de los niveles de riesgo de la IA que se lleva a cabo en la Ley de Ia de la Unión Europea. En ella se establece un tratamiento claro para sistemas de IA de “alto riesgo”, imponiendo requisitos concretos como una evaluación del impacto o documentos técnicos. Estas normas son unívocas porque se definen de manera detallada y no dejan margen para la interpretación. Esto puede conllevar problemas dado que su rigidez impide la adaptación a los avances tecnológicos.

Para conseguir una regulación efectiva de la IA, se requiere una perspectiva equilibrada que armonice la claridad de las normas unívocas con la flexibilidad indispensable para adaptarse a los avances tecnológicos. Para ello, deberá llevarse a cabo una revisión del lenguaje legal contenido en las leyes de IA para evitar términos ambiguos e incluir mecanismos de actualización de las regulaciones.

2. IDENTIDAD Y PERSONIFICACIÓN DE LA IA

Algunos expertos han planteado que los sistemas de IA avanzados podrían considerarse sujetos de derecho, defendiendo que se les podría atribuir personalidad jurídica o identidad, pero otros debaten que únicamente pueden ser considerados herramientas creadas por humanos.

Este debate jurídico ha llegado a algunos países en los que han explorado esta idea de conceder a la IA un estatus legal parecido al de una persona jurídica, lo que supondría adjudicarles derechos y obligaciones. No obstante, esto plantea diversos desafíos, debido a que ninguna IA posea

voluntad propia, autonomía moral o intención consciente que son elementos esenciales para atribuir responsabilidad legal .

Si se analiza este debate desde la perspectiva de Kant surgen distintos problemas para atribuir personalidad jurídica a la IA, debido a que el filósofo sostenía que la autonomía y la moralidad son cualidades exclusivas del ser humano. Para Kant la dignidad humana reside en la capacidad de razonar y actuar acorde a principios morales universales, cosa que las máquinas no pueden realizar. Por ello, si se tratara a alguna IA como un sujeto de derecho esto podría conllevar el riesgo de deshumanización, trasladando la responsabilidad por actos a entidades que carecen de conciencia y moralidad.

Por todo lo anterior, es esencial que el marco regulatorio de la IA sea claro en cuanto a la incapacidad de la IA para obtener un estándar de persona en el marco jurídico, y que por ello cualquier responsabilidad resultante de su uso debe recaer en usuarios, operadores o desarrolladores.

3. DERECHOS DE PROPIEDAD INTELECTUAL Y LA IA

Con el crecimiento de la IA han surgido debates sobre quién posee el título de las creaciones producidas por sistemas automatizados, es decir, si la autoría de una obra puede pertenecer a una IA o si por el contrario los derechos de autoría pertenecen a los diseñadores y programadores del sistema.

Kant, que defendía la importancia de la autonomía del individuo, consideraba a la razón y la creatividad como una expresión más de esta, lo que significa que solo los seres humanos pueden ser considerados autores de una obra. Esto se manifiesta de forma que la propiedad intelectual está siempre vinculada a la creatividad humana en el marco jurídico tradicional. Por ello, la autoría de cualquier obra generada por IA deberá atribuirse a los individuos responsables de su desarrollo y supervisión.

Desde una perspectiva legal, algunos ordenamientos jurídicos han empezado a implementar normativas para esclarecer este asunto. Un ejemplo de ello es la Unión Europea[20] donde se establece que a las obras producidas por IA no se les puede atribuir derechos de autor salvo que haya una intervención humana considerable en su proceso creativo. En contraste, países como Estados Unidos aún están discutiendo si las IA pueden ser protegidas por las leyes de propiedad intelectual.

La situación se hace más compleja en áreas como la literatura, la música y el arte digital, donde sistemas de IA han producido obras creativas sin intervención humana clara. Esto levanta preguntas sobre la necesidad de modificar la legislación vigente para combatir estos nuevos retos tecnológicos.

Es vital que las normativas continúen avanzando para establecer criterios precisos acerca de la propiedad intelectual en la época de la IA, asegurando una armonía entre la salvaguarda de la creatividad humana y el fomento de la innovación tecnológica.

4. RESPONSABILIDAD CIVIL Y PENAL: ¿QUIÉN RESPONDE POR LOS ACTOS DE LA IA?

El principal problema jurídico de la inteligencia artificial reside en sobre quién recae la responsabilidad cuando un sistema automatizado causa daños. La principal cuestión consiste en determinar quién es el responsable legal: el propio sistema de IA (si se le reconociese algún tipo de personalidad legal), el usuario que lo utiliza, la empresa que lo comercializa o el programador.

Desde la óptica de Kant, la responsabilidad está ligada a la capacidad de actuar acorde a principios racionales y a la autonomía moral. La IA no posee voluntad ni intencionalidad por lo que no puede considerarse responsable de sus actos, debiendo entonces recaer sobre las personas encargadas del diseño, implementación y uso de estos sistemas.

En la actualidad, los marcos legales abordan el tema desde distintos enfoques:

[20] **Reglamento (UE) 2021/0106 del Parlamento Europeo y del Consejo, de 21 de abril de 2021, sobre la inteligencia artificial (Ley de IA)** (DOUE L 122, 18 de mayo de 2021).

Responsabilidad del usuario: en los casos en los que este haga un uso indebido de la IA que cause daños y por tanto podría ser considerado responsable de las repercusiones negativas de dicho uso.

Responsabilidad del desarrollador: podrá ser considerado responsable cuando se cause un daño como consecuencia de un error cometido por éste en el diseño de un sistema de IA.

Responsabilidad compartida: este planteamiento es cada vez más común y consiste en repartir la responsabilidad entre los agentes involucrados en el desarrollo y aplicación de la IA.

Modelos de seguro obligatorio: ciertos reguladores han optado por planteamientos parecidos a los de seguros de automóviles, en los cuales los fabricantes de IA se verían obligados a contratar pólizas que cubran los posibles daños que puedan causar sus sistemas.

Los problemas en este campo no son solo de carácter civil sino que también afectan al derecho penal. Una IA puede tomar decisiones discriminatorias, producir contenido difamatorio o contribuir de forma indirecta a la comisión de delitos. En estos casos, se plantea la cuestión de cómo atribuirle responsabilidad legal. Se debe desarrollar un marco regulatorio concreto y claro para este tipo de situaciones sin dispensar a los humanos de su obligación de supervisión.

En todo caso, las leyes y reglamentos deben asegurar que quienes sufran daños por IA puedan acceder a mecanismos de reparación y que se instituyan estándares de supervisión y seguridad aptos para prevenir riesgos.

Para concluir, los problemas jurídicos que plantea la IA precisan de una regulación normativa que combine la promoción de la innovación con la protección de los derechos fundamentales. Para ello, se puede tomar como base la filosofía de Kant, por la cual las leyes de IA deberán respetar la dignidad y autonomía humana en todo caso.

CAPÍTULO VI: IMPACTO DE LA IA EN LOS DERECHOS HUMANOS

1. PRIVACIDAD Y PROTECCIÓN DE DATOS PERSONALES

La forma de gestionar los datos personales ha sido revolucionada por la IA, en concreto por cosas como los algoritmos de publicidad personalizada o los sistemas de reconocimiento facial. La IA recaba datos personales de sus usuarios para poder llevar a cabo estas acciones. Esto implica un grave riesgo de vulneración de derechos fundamentales como la privacidad o el control de la información personal en caso de una gestión inadecuada de los mismos. Esto quedó reflejado con el bloqueo temporal de Chat GPT en Italia al descubrirse que el chatbot de Open AI no cumplía con las leyes de protección de datos del consumidor.[21]

La Unión Europea ha establecido en el Reglamento General de Protección de Datos (RGPD) los principios esenciales para asegurar la privacidad en la era actual. Los más importantes son el artículo 17 que establece el derecho de supresión o de olvido por el cual una persona puede requerir que se eliminen sus datos personales si ya no son necesarios para su finalidad, si se retira el consentimiento jurídico, si se opone él mismo al tratamiento de sus datos, si se han tratado de manera ilícita o si deben suprimirse para el cumplimiento de obligaciones del DUE. El artículo 32, que establece la necesidad de prestar consentimiento informado, es decir, que sea fruto de una manifestación libre de la voluntad, informada, concreta e inequívoca de la persona que acepta el tratamiento de sus datos personales. Esto se observa en las "cookies" de las páginas web que exigen que se marque una casilla aceptando o rechazando el tratamiento de los datos personales o en escritos pidiendo que se firme una declaración del tratamiento de los datos. Por último, el principio de minimización de datos recogido en el artículo 5 que establece que los datos deben ser "adecuados, pertinentes y limitados a lo necesario en relación con los fines para los que son tratados"[22] No obstante, son muchos los gobiernos y empresas que enfrentan desafíos a la hora de aplicar estos principios en el uso de IA.

---

[21] Montero Carretero, L, "Los límites de los algoritmos", *ABC Empresa* , 28 de mayo de 2023

[22] **Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos** (DOUE L 119, 4 de mayo de 2016).

Desde la óptica kantiana, respetar la privacidad está estrechamente ligado con la autonomía moral de los individuos. Kant defendía que todo individuo debía ser tratado como un fin en sí mismo y no como un medio para lograr otro propósito. La recopilación de datos sin consentimiento o un uso indebido vulneran este principio ya que tratan al ser humano como un medio para lograr información sobre un tema concreto.

2. DISCRIMINACIÓN Y SESGOS ALGORÍTMICOS

Los sistemas de IA se programan a partir de datos históricos que pueden estar sesgados, lo que implicaría que las soluciones propuestas por estos serían también sesgadas. La presencia de sesgos en algunos algoritmos prolongan y amplían discriminaciones presentes en la sociedad. Es el caso de la desigualdad en las concesiones de créditos, donde ha quedado demostrado que algunos sistemas automatizados conceden tasas de interés superiores a las minorías como consecuencia de patrones históricos de discriminación bancaria. Otro ejemplo es la discriminación existentes en los contratos laborales en los cuales ha quedado demostrado que algunos algoritmos que empresas emplean para escoger candidatos tienden a descartar a algunos grupos como consecuencia de sesgos existentes en los datos usados para entrenar dicha IA. Por ejemplo, en 2018 Amazon fue criticado por que sus sistemas de IA para reclutamiento laboral penalizan de manera automática los currículums que incluían términos vinculados con mujeres. Además, existen estudios que han desvelado que algoritmos empleados en plataformas de publicidad en línea pueden enseñar anuncios a un público concreto en exclusiva. [23]

Por último, los riesgos que existen en la justicia penal; Ciertos países emplean IA para predecir las posibilidades de reincidencia de algunos delincuentes y en los cuales han existido prejuicios para algunos colectivos como consecuencia de datos históricos usados en el entrenamiento de la IA como es el caso COMPAS mencionado anteriormente.

Desde la óptica kantiana, la presencia de sesgos en IA supone una vulneración del principio ético de tratar a las personas como fines en sí mismos ya que implica tratar a las personas como

[23] *Laboratoria*, "10 casos donde la inteligencia artificial jugó en contra de la diversidad", *Laboratoria*, disponible en https://www.laboratoria.la/post/10-casos-donde-la-inteligencia-artificial-jugo-en-contra-de-la-diversidad

medios para mantener un sistema dispar. Para evitar que esto suceda se deberían crear mecanismos que aseguren la equidad en los algoritmos. Garantizando que se respete la dignidad de todos los hombres en las decisiones automatizadas.

3. DERECHO A LA TRANSPARENCIA Y EXPLICABILIDAD

Uno de los desafíos que enfrenta el uso de la IA consiste en la ausencia de transparencia en la toma de decisiones. En muchas ocasiones ni los desarrolladores de los modelos de IA tienen la capacidad de explicar cómo alcanzan algunos resultados. Esto causa problemas en áreas sensibles. Entre ellas, las decisiones financieras, ya que algunos bancos podrían no conceder un préstamo a alguien como consecuencia de los resultados de un algoritmo sin aportar ninguna justificación concreta al cliente. Otro ámbito sería el de los sistemas de vigilancia en los que los sistemas de reconocimiento facial podrían tomar por sospechosos a personas sin ninguna revisión humana apropiada. También pueden surgir problemas en el campo de la medicina ya que existen sistemas de IA capaces de detectar enfermedades con gran precisión, pero que no exista forma de justificar este diagnóstico lo que puede complicar la decisión de los médicos de confiar en la IA al no conocer los motivos detrás de dicho diagnóstico.

En 2020 Nature Medicine[24] publicó un estudio en el que reveló que un algoritmo de diagnóstico de cáncer de piel desarrollado por Google Health lograba mejores resultados que médicos humanos en ciertas pruebas, pero, más adelante se desveló que el algoritmo se fundamentaba en factores irrelevantes, como que aparecieran reglas de medición en las imágenes, en vez de en características médicas relevantes. Este caso no se asocia con la falta de precisión del modelo, sino con su dependencia en variables no trascendentales para el diagnóstico real. Refleja la falta de explicabilidad existente en numerosos sistemas de IA, que imposibilita a los usuarios comprender las razones detrás de sus decisiones.

Kant sostenía la importancia de la claridad y la razón en los actos morales, lo que supondría que los sistemas de IA deberían ser comprensibles y explicables, es decir, ser totalmente

[24] Nature Medicine, "Evaluation of a deep learning algorithm for detection of skin cancer in a clinical setting", *Nature Medicine*, vol. 26, n. 9, 2020, pp. 1344-1350 (disponible en https://www.nature.com/articles/s41591-020-1004-1

transparentes para los afectados, de forma que se garantice la racionalidad y justicia de toda decisión.

Para resolver estos problemas algunos organismos internacionales han formulado algunos principios de IA auditable y explicable. Un ejemplo de ello son las directrices que ha desarrollado la Comisión Europea en relación a la necesidad de que los sistemas de IA sean comprensibles tanto para los usuarios como para los reguladores.

En suma, se requiere un enfoque no regulador ético riguroso de la IA debido a su gran impacto en los derechos humanos, que podría hacerse en base a la ética de Kant que promueve la equidad, la transparencia y el respeto a la autonomía individual en el diseño y en la aplicación de la IA.

CAPÍTULO VII: IA Y EL DERECHO INTERNACIONAL

1. RETOS DE LA REGULACIÓN GLOBAL DE LA IA

En el marco regulatorio internacional existen numerosos desafíos en el ámbito de la IA como consecuencia de la falta de un marco normativo común. Distintos países han escogido diversos modos de enfocar los problemas causados por el uso de IA lo que genera problemas para su aplicación y cumplimiento. Regiones como Estados Unidos o China han optado por una regulación más flexible y que se centra principalmente en el crecimiento de la economía y la innovación de la tecnología, mientras que la Unión Europea ha optado por un enfoque más restrictivo basado en derechos fundamentales y principios éticos.

El mayor problema para la regulación internacional efectiva de la IA reside en su naturaleza transfronteriza. Los sistemas de IA son desarrollados por empresas multinacionales que actúan en distintas jurisdicciones, lo que dificulta la aplicación de la normativa local. Por ello, resulta necesario el establecimiento de estándares a nivel internacional que aseguren la ética, la seguridad y el respeto a los derechos humanos al aplicar la IA.

En línea con el pensamiento de Kant, un reglamento global debería basarse en los principios universales de autonomía y justicia. Kant respaldaba la creación de una legislación moral común que asegure el respeto a la dignidad humano en todo contexto. Por ello, debería diseñarse el marco regulatorio de la IA en base a normas universales que garanticen un empleo ético de la misma en cualquier lugar del mundo.

2. INCERTIDUMBRE LEGAL POR FALTA DE UN MARCO NORMATIVO COMÚN
    - **2.1 Normativa vigente sobre IA**
        i. *2.1.1 Normativa sobre IA en China*

China se ha decantado por una regulación centralizada y muy restrictiva cuyas principales normativas son la normativa sobre IA en la seguridad pública, la Ley de Protección de Información Personal (PIPL) y la Regulación de Algoritmos de IA.

China ha introducido IA en sus sistemas de seguridad pública como el análisis de datos masivos para la vigilancia o sistemas de reconocimiento facial. La normativa existente que afecta a esto establece el uso obligatorio de IA en seguridad ciudadana que se puede usar para identificar a individuos en zonas públicas sin que sea precisa una autorización judicial. También se establecen mecanismos de supervisión en plataformas digitales así como en redes sociales con el fin de reconocer posibles riesgos o discrepancias políticas. Por último, se ha mandado recabar datos biométricos de los ciudadanos con el objetivo de renovar los sistemas de reconocimiento facial en lugares públicos.[25]

Por otro lado, la PIPL[26], aprobada en 2021, es la legislación China principal en lo relativo al uso de datos personales en IA. Está inspirada en el GDPR de la UE y decreta que las empresas solo pueden recabar datos personales de los usuarios con su consentimiento previo explícito y que sólo podrán exportar estos datos a otros países si lo aprueba el gobierno. En caso de infringir cualquiera de estos dos requisitos, se impondrá una multa del 5% de los ingresos anuales d ela empresa infractora.

En 2022 China aprobó una tercera medida consistente en la Regulación de Algoritmos de IA llevada a cabo por la Administración del Ciberespacio de China[27]. Esta medida obliga a las empresas tecnológicas que emplean IA a registrar sus modelos de algoritmos con el gobierno aportando detalles sobre sus criterios para la toma de decisiones y sobre su funcionamiento. También se establece como requisito que las empresas implementen mecanismos de supervisión y auditoría regulares que aseguren el cumplimiento de las normas. Por último, se prohíbe el uso de algoritmos para manipular la opinión pública a través de información sesgada.

[25] **China AI Governance Initiative**, 2021 (disponible en https://www.ChinaAIgovernance.org
[26] **Ley de Protección de la Información Personal (PIPL)**, aprobada el 20 de agosto de 2021 por la Asamblea Popular Nacional de China (disponible en https://www.npc.gov.cn
[27] **Cyberspace Administration of China**, *Regulaciones de Gestión de Servicios de Recomendación Algorítmica de Información en Internet*, 2022 (disponible en https://www.cac.gov.cn

Con estas medidas China busca proteger la estabilidad social y la seguridad nacional. A pesar de ello, estas medidas han sido criticadas al considerar que vulnera los derechos humanos al permitir la vigilancia masiva mediante el uso de IA y la ausencia de transparencia en el tratamiento de los datos personales.

ii. 2.1.2 Normativa sobre IA en Estados Unidos

Estados Unidos ha optado por una regulación de la IA descentralizada y más flexible que se basa en la autorregulación y en el establecimiento de pautas no vinculantes en vez de leyes federales unificadas.

Las iniciativas más destacadas son Blueprint for an AI Bill of Rights que consiste en un escrito realizado por la Casa Blanca en el que se establecen principios éticos a seguir en el uso de IA, pero que carecen de carácter vinculante como la transparencia, la equidad o la seguridad.

Otra iniciativa destacada es la Executive Order on AI (2023) que sienta las pautas a seguir en el uso de IA en economía, sanidad y defensa, encomendando la regulación a agencias del gobierno; a saber: La Comisión Federal de Comercio (FTC) se ocupa de la supervisión del uso de la IA en privacidad de datos y publicidad. La Administración de Alimentos y Medicamentos (FDA) se encarga de la regulación de la IA en diagnósticos y dispositivos médicos. Por último, la Comisión de Bolsa y Valores (SEC) se ocupa de la evaluación del impacto de IA en el sector de las finanzas.

Este enfoque fomenta la competitividad e innovación tecnológica, pero implica problemas a la hora de determinar la responsabilidad legal y la protección del consumidor. La ausencia de un marco único crea inseguridad jurídica en los casos de errores cometidos por IA como puede ser en discriminaciones algorítmicas o en diagnósticos médicos.

iii. 2.1.3 Normativa sobre Ia en la Unión Europea

La Unión Europea por su parte ha optado por un enfoque basado en valores democráticos y derechos fundamentales. La Comisión Europea presentó en 2021 la Ley de Inteligencia Artificial, una regulación pionera en el ámbito de la IA que clasifica en cuatro niveles de riesgo los sistemas de IA. En primer lugar, el nível de riesgo inaceptable que son aquellos sistemas que

están prohibidos como son el uso de IA para vigilancia masiva o para la manipulación subliminal. En segundo lugar, el nível de alto riesgo que está compuesto por sistemas de IA empleados en sectores críticos como son el transporte, la sanidad o la justicia. Los sistemas clasificados en este nível requieren certificados y auditorías previas a su implementación. En tercer lugar, el nível de riesgo limitado en el que se incluyen sistemas como chatbots que no precisan de una regulación estricta pero si de un funcionamiento transparente. Por último, el nível de riesgo mínimo compuesto por la IA que no presenta riesgos importantes y puede emplearse con libertad, por ejemplo los algoritmos de recomendación de páginas de streaming.

Esta Ley ha influenciado otros marcos normativos como el del G7 o los debates de la ONU. No obstante, algunos sectores critican que esta regulación limita la competitividad e innovación tecnológica de la UE con respecto a China o Estados Unidos.

Desde la perspectiva de Kant, la legislación europea establece principios morales fundamentales como la autonomía y dignidad humana así como la equidad. Además, la Ley de IA cumple con la maxima kantiana de tratar al hombre siempre como un fin y nunca como un medio al asegurar que el desarrollo de la tecnología no implique riesgos desmesurados para los ciudadanos ni vulnere sus derechos fundamentales.

- **2.2 Falta de armonización y conflictos normativos**

La heterogeneidad en los marcos normativos de China, EEUU y UE crea inseguridad legal y limita la cooperación internacional en el ámbito de IA. Uno de los principales problemas surge en la transferencia de datos. La UE limita severamente la exportación de datos a países con regulaciones divergentes al GDPR, lo que repercute en empresas chinas y estadounidenses que para entrenar sus sistemas de IA requieren de datos europeos. Otro problema consiste en que las empresas actúan globalmente debiendo ajustar sus productos a regulaciones divergentes, lo que incrementa los costes. Esto impacta principalmente a empresas tecnológicas con sede en China o EE y que buscan operar en la UE como consecuencia de los severos requisitos de la Ley de IA. Por último, surge el problema de la determinación de responsabilidad en los casos de errores cometidos por IA. La falta de una regulación internacional clara y común dificulta la rendición de cuentas.

Un ejemplo claro de problemas derivados por la falta de armonización de la ley de IA internacionalmente es el caso de Tik Tok. En 2020, EEUU trató de prohibir la aplicación de Tik Tok en base a riesgos para la seguridad nacional, alegando que la empresa matriz, ByteDance, podía estar forzada a transferir datos al gobierno de China conforme a la Ley de Seguridad Nacional de China. Al mismo tiempo, la UE inició una investigación por supuestas vulneraciones del GDPR en la recopilación de datos de menores.[28]

ByteDance desmintió exportar información al gobierno de China, pero la administración de EEUU estimó como única solución posible la compra de las operaciones de Tik Tik en EEUU por parte de una empresa nacional. Al final, ByteDance logró evitar la prohibición mediante el refuerzo de su política de privacidad y trasladando una parte del almacenamiento de datos a servidores en EEUU.

Este caso refleja la incertidumbre generada para gobiernos y empresas por la ausencia de armonización regulatoria, mostrando la importancia de llevar a cabo acuerdos internacionales sobre la supervisión de la IA y el uso de datos.

- **2.3 Propuesta de armonización regulatoria**

Para poner fin a las discrepancias regulatorias y asegurar un desarrollo seguro y ético de la IA se plantea crear un Tratado Internacional de IA en el que se expongan principios globales y mecanismos de cooperación regulatoria.

3. INICIATIVAS INTERNACIONALES Y COOPERACIÓN INTERGUBERNAMENTAL

Muchas organizaciones internacionales han buscado establecer un marco regulatorio de IA común con el objetivo de garantizar un uso y desarrollo ético y seguro de la misma. Las iniciativas más destacadas son:

- **3.1 Iniciativas de la OCDE**

[28] **Supreme Court of the United States**, Nos. 24-656 y 24-657, *TikTok Inc., et al., Petitioners 24–656 v. Merrick B. Garland*, 17 de enero de 2025.

La Organización para la Cooperación y el Desarrollo Económicos (OCDE) es un pilar clave en la determinación de principios éticos y jurídicos para el uso y desarrollo de la IA.

En 2019 estableció los "Principios de la OCDE sobre IA"[29] un documento pionero en el abordaje de la regulación de IA que establecía como principios fundamentales para el uso y desarrollo de la misma la transparencia y explicabilidad de sus algoritmos, la responsabilidad tanto de desarrolladores como de operarios, la robustez y seguridad para minimizar los riesgos, el respeto a los derechos humanos y el crecimiento inclusivo y bienestar. Todos los países miembros de la OCDE y otros asociados como Brasil o Argentina adoptaron estos principios que sentaron la base para futuras normativas internacionales y nacionales reguladoras de IA.

En 2023 la OCDE actualizó sus directrices, consolidando la necesidad de cooperación internacional y armonizaciones de las regulaciones relativas a IA en un contexto globalizado. Destacó la importancia de establecer principios comunes en ámbitos como la protección de datos, la privacidad y la gobernanza de algoritmos. Estableció la necesidad de crear un marco normativo común para garantizar una IA confiable y la igualdad en el acceso a la misma.

Todos estos principios entran dentro del marco de la ética kantiana al establecer normas universales que se pueden aplicar a todo individuo, que respetan la autonomía moral y la dignidad humana, favoreciendo una regulación equitativa y justa de carácter internacional.

- **3.2 Iniciativas de la ONU**

La UNESCO adoptó en 2021 la Recomendación sobre la Ética de la Inteligencia Artificial[30] como marco ético global que garantiza un desarrollo sostenible y el respeto a los derechos humanos. Establece como principios el respeto a los derechos humanos, especialmente a la igualdad de oportunidades, la no discriminación y la dignidad. La supervisión y gobernanza recomendando crear mecanismos independientes que controlen y evalúen el impacto de la IA en la sociedad. La equidad y diversidad, asegurando que se represente a todos los géneros y culturas en el desarrollo de IA, a la par que evitar cualquier sesgo en los algoritmos.

[29] **OCDE**, *Recomendación sobre la inteligencia artificial*, 2019 (disponible en https://www.oecd.org
[30] **UNESCO**, *Recommendation on the Ethics of Artificial Intelligence*, 2021 (disponible en https://unesdoc.unesco.org

Otra iniciativa de la ONU en el ámbito de la IA fue la Cumbre Global sobre Inteligencia Artificial para el Bien Común en 2022 en la que se debatieron estrategias para limitar la brecha tecnológica entre países en desarrollo y desarrollados.

La iniciativa de la ONU se alinea con la ética kantiana al tratar de establecer unos principios universales. Kant sostenía la necesidad de crear leyes que se basaran en principios éticos aplicables globalmente. La recomendación de la UNESCO cumple esto al proponer principios globales para la regulación de la IA.

- **3.3 Iniciativas del G20**

El G20 adoptó los "Principios de IA del G20"[31] en 2019 que regulan la IA desde un enfoque económico y de innovación centrado en la colaboración público-privada y la competitividad económica. Los principios son: desarrollo de infraestructuras digitales para el uso equitativo de IA, fomento de la innovación e inversión en IA, y la promoción de estándares internacionales en IA.

El G20 da prioridad al aspecto económico por encima de una regulación estricta como hacen la ONU y la UE, pero aún así reconoce la necesidad de asegurar la equidad y disminuir al mínimo los riesgos de la IA en la privacidad y el mercado laboral.

Este enfoque contrasta con la ética kantiana al dar prioridad a la innovación que puede llevar a tratar a los hombres como medios en vez de fines en sí mismos. Kant consideraba que el avance tecnológico debía alinearse con principios morales universales, lo que supone que la IA deba desarrollarse conforme a principios éticos y no exclusivamente en base a la innovación.

- **3.4 *Antiqua et Nova***

"*Antiqua et Nova*"[32] es una reflexión de los dicasterios vaticanos de Doctrina de la Fe y de Cultura en combinación con las palabras del papa Francisco sobre la regulación de la IA desde

[31] Pérez Bes, F., "El G20 busca un equilibrio entre regulación e innovación para la IA", *Demócrata*, 30 de noviembre de 2024, disponible en https://www.democrata.es/analisis-y-opinion/el-g20-busca-un-equilibrio-entre-regulacion-e-innovacion-para-la-ia/

[32] **Vaticano**, *Antiqua et Nova*, 2025 (disponible en https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_sp.html

un punto de vista ético y humanista. En ella se establece la necesidad de establecer un marco regulatorio y ético de la IA que se base en el respeto a la dignidad humana y la búsqueda del bien común. Señala como principios clave la responsabilidad, la transparencia y la inclusión, promoviendo una IA que respete los derechos humanos y que no instrumentalice a las personas, alineándose con la ética kantiana y su imperativo de considerar a las personas siempre como fines en sí mismas. Además, el documento enfatiza la necesidad de armonizar las regulaciones internacionales para evitar desigualdades lo que se alinea con la idea kantiana de establecer un marco normativo universal que se base en la razón y la autonomía moral.

En este documento se enfatiza la necesidad de armonizar las normas internacionales para evitar desigualdades jurídicas y garantizar que la IA esté al servicio del desarrollo humano.

Todas estas iniciativas son un reflejo de la urgencia de establecer un marco regulatorio de IA global que tenga como base principios ético-jurídicos comunes. No obstante, la implementación de este dependerá de la voluntad de cada Estado, así como del compromiso de las empresas de tecnología.

4. ARMONIZACIÓN NORMATIVA Y CONFLICTOS JURÍDICOS INTERNACIONALES

Cada país ha optado por un enfoque distinto en la regulación de la IA. La Unión Europea se ha focalizado en la ética y los derechos fundamentales. Mientras que China ha dado prioridad a la supervisión de la IA por parte del Estado y al control gubernamental, y Estados Unidos ha optado por un enfoque menos centralizado que se basa en la autorregulación del sector privado.

La ausencia de armonización entre los distintos Estados crea una incertidumbre de carácter jurídico así como posibles conflictos normativos. Esto sucede dado que las empresas internacionales operan en distintas jurisdicciones lo que las obliga a cumplir requisitos regulatorios que se contradicen entre sí dificultando el comercio internacional y la innovación.

Si aplicamos la filosofía kantiana a este problema la solución sería establecer un modelo que se base en un “contrato social global” que asegure que la leyes regulatorias de la IA respeten la dignidad y autonomía humana en todo el mundo. Para lograr esto se deberían realizar tratados

internacionales y acuerdos multilaterales que logren una regulación efectiva y coherente de la IA en el plano internacional.

CAPÍTULO VIII: PROPUESTAS DE SOLUCIÓN ÉTICO-JURÍDICAS

Los problemas éticos y jurídicos derivados del uso y desarrollo de la IA que han sido analizados en este trabajo reflejan la necesidad de buscar soluciones estructuradas y globales. Por ello, se propone como posible herramienta de solución de estos problemas la elaboración de un Tratado Internacional que suscriban las principales potencias en IA (China, EEUU, UE, Rusia e India) y a ser posible el resto de potencias en las que la IA está evolucionando. Este tratado deberá basarse en principios éticos universales alineados con la ética kantiana, destacando el respeto a la dignidad humana, el bienestar social y la autonomía moral.

Los apartados que deben incluirse en este tratado son los siguientes:

1. PRINCIPIOS PROPUESTOS PARA UNA IA RESPONSABLE

Con el fin de lograr un uso justo y ético de la IA es necesario el desarrollo de un marco normativo que tenga por base principios universales. Kant sostenía que en todo caso debía respetarse la dignidad y autonomía humana y promoverse el bienestar social. Por ello, el diseño y aplicación de IA debe respetar ambos y guiarse por principios éticos universales como:

- No discriminación y equidad:se deben diseñar los sistemas de IA con mecanismos que eviten cualquier sesgo o trato discriminatorio por parte de los algoritmos y prevenir la creación o perpetuación de desigualdades sociales. Esto implica incluir mecanismos que garanticen decisiones justas e imparciales.
- Explicabilidad y transparencia: Para prevenir decisiones arbitrarias o sesgos en los algoritmos de IA, se debe garantizar la transparencia de los mismos debiendo estos ser auditables y comprensibles. De esta manera ciudadanos, usuarios y reguladores pueden comprender cómo se toman las decisiones y evitar que tengan carácter arbitrario o sesgado.
- Respeto por la autonomía humana: la IA no debe reemplazar la autonomía moral de los ciudadanos ni su capacidad para juzgar sino limitarse a asistir y complementar la toma de las decisiones.

- Privacidad y seguridad: el uso de los datos recabados por la IA debe respetar los principios ético-jurídicos garantizando su protección y evitando un uso indebido. Esto incluye la implementación de políticas estrictas de seguridad de datos.
- Responsabilidad y rendición de cuentas: los males causados por el uso de IA serán responsabilidad de quienes desarrollaron la misma, así como de los operadores.

Todos estos principios se alinean con el Imperativo Categórico de Kant al establecer que cualquier norma que se aplique a la IA debe poder tomar carácter universal y ser justa para cada individuo, así como garantizar el respeto a la dignidad y autonomía humana.

2. NORMATIVAS SUGERIDAS PARA UNA REGULACIÓN EFECTIVA

Los desafíos ético-jurídicos que existen en relación a la IA hacen necesario el desarrollo de una regulación que, sin limitar la innovación tecnológica, proteja la dignidad y autonomía humana. Por ello, se proponen estas medidas de regulación:

- Sanciones por uso indebido de la IA: se deben establecer sanciones de responsabilidad civil y penal para quienes lleven a cabo un uso indebido de la IA, ya sea usándola de manera discriminatoria o negligente, garantizando que cualquier daño causado por el uso indebido de IA sea debidamente penalizado.
- Requisito de explicabilidad: se debe garantizar que las decisiones tomadas por sistemas de IA sean explicables. Esto implica que las decisiones sean comprensibles y auditables por reguladores y usuarios.
- Auditoría y supervisión de los algoritmos: para asegurar la justicia, equidad y transparencia de los algoritmos de IA se deberán crear organismos independientes que se encarguen de su evaluación y auditoría. Esto incluye la revisión no solo de los algoritmos existentes sino también la evaluación de sus impactos.
- Protección de los datos personales: se deben establecer normas más severas en relación a la recopilación, almacenamiento y uso de datos personales en los sistemas de IA. Las leyes deben asegurar que los datos sean empleados de forma ética y que no vulnere el derecho a lr privacidad.

Estas regulaciones serían acordes a la perspectiva kantiana que defiende que las normas deben ajustarse a principios de justicia y moralidad, evitando que el uso de IA tome al hombre como un mero medio político o comercial sin ninguna consideración ética, y lo tome como un fin en sí mismo.

3. ROL DE LAS ORGANIZACIONES Y LA SOCIEDAD CIVIL

La sociedad y las organizaciones deben implicarse en el desarrollo e implementación de la IA, no siendo esto tarea exclusiva de corporaciones tecnológicas y gobiernos. Estos deben ocuparse del control del uso de la IA y de su supervisión. Entre las acciones que deben llevar a cabo destacan:

- Participación en la toma de decisiones: Kant defendía que debía tomarse al ser humano como un fin en sí mismo por ello deben tenerse en cuenta sus intereses en todos los aspectos de la vida lo que incluye el uso de la IA. Por ello, los ciudadanos deben participar en la toma de decisiones sobre el desarrollo de la IA, aportando su perspectiva tanto en debates públicos como en consultas.
- Fomento de la educación en ética de la IA: se deben impulsar programas educativos que enseñen a los futuros usuarios y desarrolladores de IA los principios éticos que deben primar en el uso de las nuevas tecnologías para que llegado el momento comprendan sus implicaciones morales y sean capaces de hacer un uso y desarrollo de la misma responsable.
- Monitoreo y denuncia de abusos: se deben poner a disposición de los ciudadanos mecanismos de vigilancia y denuncia de malos usos de la IA en distintos ámbitos como pueden ser seguridad, justicia o empleo.
- Creación de espacios de diálogo interdisciplinarios: para poder diseñar normativas más justas y equilibradas es necesario crear un espacio en el que puedan interactuar juristas, filósofos, desarrolladores de IA y miembros de la sociedad civil.

Todas estas propuestas refuerzan la noción kantiana de que cada individuo es responsable de actuar acorde a principios de equidad y justicia, no dependiendo la moralidad únicamente de las instituciones.

4. RÉGIMEN DE RESPONSABILIDAD CIVIL Y PENAL

Se debe establecer un régimen común de responsabilidad civil y penal por los daños causados por el uso y desarrollo de la IA con el fin de evitar inseguridad jurídica por falta de armonización entre las regulaciones internacionales. La existencia de un marco normativo único promovería una mayor confianza en la IA garantizando que los actores involucrados asuman responsabilidades de manera clara y equitativa. Este régimen debe establecer la responsabilidad de los usuarios y de los desarrolladores, teniendo en cuenta las implicaciones jurídicas y éticas de sus correspondientes responsabilidades.

○ 4.1 Responsabilidad de los desarrolladores de IA

Una propuesta de régimen de responsabilidad común sería establecer que los desarrolladores de sistemas de IA sean los responsables por la presencia de sesgos algorítmicos y discriminaciones. Serán responsables cuando los sistemas de IA produzcan resultados erróneos o injustos como consecuencia de la calidad de los datos con los que se haya entrenado o por las decisiones de sueño que favorezcan a unos grupos por encima de otros. Esto es importante ya que existen ámbitos como la justicia penal, el crédito o el empleo en el que la presencia de un error o un sesgo en los algoritmos puede tener una grave repercusión en la vida de las personas.

Por ejemplo, en un proceso de selección de personal se emplea un algoritmo de IA que presenta sesgos por los cuales desfavorece a personas de un determinado género. Los desarrolladores de este programa serán los responsables por los daños causados por ese sistema. La responsabilidad civil podrá incluir una compensación por daños a las personas perjudicadas, y la responsabilidad penal podría implicar sanciones administrativas si se demuestra negligencia o conocimiento previo de los sesgos por parte de los desarrolladores.

Los desarrolladores tienen la responsabilidad moral de diseñar sistemas de IA que respeten la dignidad humana y no discriminen a las personas de manera injusta. La negligencia en la creación de sistemas discriminatorios perpetúa desigualdades sociales presentes en la sociedad que vulneran el principio kantiano de tratar a todas las personas como fines en sí mismos ya que los trataría como medio para fines comerciales sin tener en cuenta su dignidad y derechos.

- **4.2 Responsabilidad de los usuarios de IA**

Los usuarios deberán ser responsables de los daños causados por un uso indebido de la IA como puede ser la suplantación de identidad, violaciones de privacidad o fraudes. Por ejemplo, si una persona emplea un sistema de IA para evadir una sanción legal, será responsable por un uso indebido de la misma. Será responsable penalmente por emplear la IA con el fin de cometer un delito.

Si un usuario emplea la IA con el fin de infringir normas o vulnerar los derechos de otras personas, está vulnerando el principio kantiano de tratar a los demás como fines y no como medios para lograr sus propios intereses. Kant, decía que los individuos eran moralmente responsables de sus acciones y que debían actuar conforme a principios éticos que respeten la dignidad humana. Como consecuencia de esta autonomía moral si el usuario decide libremente hacer un mal uso de la IA deberá responder rindiendo cuentas por ello.

- **4.3 Posibles sanciones civiles y penales**

En primer lugar, las sanciones penales que se establezcan deberán tener por finalidad castigar conductas que provoquen un daño grave y directo a la seguridad pública o los derechos de las personas por un uso irresponsable de la IA. Estas sanciones podrán ser multas o incluso pena de prisión en función de la gravedad del delito. Se buscará establecer unas penas comunes a todos los países para garantizar equidad, igualdad y justicia, al mismo tiempo que evitar cualquier tipo de inseguridad jurídica.

En cuanto a las sanciones civiles, deberán tener como fin compensar a quienes hayan sido víctimas de daños causados por el uso y desarrollo de la IA. Estas compensaciones pueden incluir indemnizaciones, como compensaciones económicas en el caso de daños personales o sociales, multas, en casos de mala praxis o negligencia, e incluso inhabilitaciones de sistemas de IA o prohibiciones temporales o permanentes para los casos más graves.

CAPÍTULO IX: CONCLUSIONES

1. REFLEXIONES FINALES SOBRE LOS PROBLEMAS ÉTICOS Y JURÍDICOS

El análisis de la IA desde la perspectiva de la Teoría Tridimensional del derecho ha permitido identificar los principales problemas éticos y jurídicos derivados de su desarrollo y uso. En la dimensión de la materia, se ha demostrado la creciente presencia de la IA en sectores fundamentales como la justicia, la sanidad o la seguridad, planteando beneficios, pero al mismo tiempo importantes riesgos. En la dimensión de la forma, se han examinado las normas vigentes y se ha constatado su imposibilidad de hacer frente a los retos éticos y jurídicos causados por la IA, tanto por la falta de mecanismos de control efectivos como por la ausencia de armonización internacional. En relación con la finalidad, se ha demostrado la urgencia de garantizar que la IA se desarrolle y emplee de acuerdo a principios éticos universales que garanticen el respeto a la dignidad y autonomía de las personas, evitando su instrumentalización.

Hoy en día, los principales desafíos identificados con respecto a la IA comprenden la falta de transparencia en la toma de decisiones, los sesgos algoritmos, la responsabilidad en caso de daños causados por programas de IA y la falta de protección de los datos personales de los individuos. Estos problemas revelan la necesidad de instituir un marco normativo internacional que permita mitigar estos riesgos y asegurar un uso ético y justo de la IA. Asimismo, se ha comprobado que las normas jurídicas vigentes tienen un alto grado de equívoco en su interpretación, lo que exige dotarlas de mayor coherencia y precisión.

La solución que se propone en este trabajo es la creación de un Tratado Internacional sobre IA, que suscriban las principales potencias mundiales en materia de IA, como China, Estados Unidos, La Unión Europea y Rusia. En ese tratado se fijarían principios éticos universales para el uso y desarrollo de la IA, que tengan por base la ética kantiana, englobando la responsabilidad, transparencia y respeto a los derechos humanos. Del mismo modo, se plantea la creación de órganos independientes responsables de la supervisión y auditoría de los algoritmos de IA con la finalidad de evitar discriminaciones y asegurar la justicia y equidad en la toma de decisiones de sistemas automatizados. Por último, se establecería un régimen común de responsabilidad civil y

penal por los daños causados por el uso de sistemas de IA, con el fin de brindar seguridad jurídica a los ciudadanos.

En conclusión, el desarrollo y uso de la IA genera problemas que deben abordarse desde una perspectiva jurídica y ética. Por medio de la Teoría Tridimensional del Derecho, se ha examinado la materia, la forma y la finalidad de la IA, concluyendo que es indispensable una normativa internacional basada en la ética kantiana para asegurar un uso responsable y justo de esta tecnología. Los agentes jurídicos como cuarto elemento de la Teoría Tridimensional del Derecho deben estructurar y ordenar la regulación de la IA. Sólo se podrá alcanzar un equilibrio entre la protección de los derechos fundamentales y la innovación tecnológica por medio de un esfuerzo conjunto entre desarrolladores, gobiernos, juristas y usuarios.

2. FUTURAS LÍNEAS DE INVESTIGACIÓN Y DESARROLLO EN IA

La IA es un campo que está constantemente evolucionando lo que hace necesaria una investigación continua sobre sus impactos jurídicos y éticos. Esta investigación deberá centrarse en:

- Análisis de la relación entre derechos humanos e IA: estudiar de qué forma puede la IA ser usada para promover y proteger los derechos fundamentales.
- Implementación de sistemas de control ético en IA: diseño de herramientas que aseguren que la IA respeta los principios éticos en todos los contextos.
- Evaluación de los impactos socioeconómicos de IA: análisis del impacto de la IA y la automatización en la educación, el empleo y la distribución de la riqueza.
- Desarrollo de IA auditable y explicable: crear algoritmos que concedan a los usuarios la posibilidad de comprender la forma en qué se toman las decisiones automatizadas.
- Exploración de nuevos modelos de gobernanza para la IA: investigación de estructuras de regulación y políticas que garanticen un control democrático de la tecnología.

En definitiva, la regulación de la IA debe basarse en una constante innovación que respete siempre los valores jurídicos y éticos fundamentales y que parta de la colaboración entre

sociedad civil, empresas y gobiernos para garantizar que esta se utilice de forma transparente, justa y respetuosa de la dignidad y autonomía humana.

BIBLIOGRAFÍA

1) LEGISLACIÓN

Reglamento (UE) 2021/0106 del Parlamento Europeo y del Consejo, de 21 de abril de 2021, sobre la inteligencia artificial (Ley de IA) (DOUE L 122, 18 de mayo de 2021).

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (DOUE L 119, 4 de mayo de 2016).

Ley de Protección de la Información Personal (PIPL), aprobada el 20 de agosto de 2021 por la Asamblea Popular Nacional de China (disponible en https://www.npc.gov.cn

Cyberspace Administration of China, *Regulaciones de Gestión de Servicios de Recomendación Algorítmica de Información en Internet*, 2022 (disponible en https://www.cac.gov.cn

2) JURISPRUDENCIA

Supreme Court of the United States, Nos. 24-656 y 24-657, *TikTok Inc., et al., Petitioners 24–656 v. Merrick B. Garland*, 17 de enero de 2025.

3) OBRAS DOCTRINALES

Kant, I. (1785). *Fundamentación de la metafísica de las costumbres* (A. González, Trans.). Editorial Losada, Buenos Aires, 2002, p. 421, 429, 433, 438

Kant, I. (1781). *Crítica de la razón pura* (A. González, Trans.). Editorial Losada, Buenos Aires, 2006, pp. 455-456, 463.

Kant, I. (1788). *Crítica de la razón práctica* (A. González, Trans.). Editorial Losada, Buenos Aires, 2002, p. A33.

Kelsen, H., *Teoría pura del derecho* (2ª ed.), Ediciones Jurídicas Euroamérica, Buenos Aires, 2003

Lazcoz Moratinos, G., "Sistemas de Inteligencia Artificial en la asistencia sanitaria: cómo garantizar la supervisión humana desde la normativa de protección de datos", *Centro de Investigación Biomédica en Red (CIBERER - ISCIII) Instituto de Investigación Sanitaria Fundación Jiménez Díaz (IIS-FJD)*

Reale, M. (1997). *Teoría Tridimensional del Derecho. Una visión integral del derecho.* Trad. A. Mateos, Editorial Tecnos, Madrid. Memoria para optar al grado de doctor, Universidad Complutense de Madrid, España.

Sandel, M. J., *Justice: What's the Right Thing to Do?*, Farrar, Straus and Giroux, Nueva York, 2009.

Strickland, E., "IBM Watson, Heal Thyself: How AI Fails in Health Care", *IEEE Spectrum*, 2019 (disponible en https://spectrum.ieee.org; última consulta 31/03/2025).

Vaticano, *Antiqua et Nova*, 2025 (disponible en https://www.vatican.va/roman_curia/congregations/cfaith/documents/rc_ddf_doc_20250128_antiqua-et-nova_sp.html

4) RECURSOS DE INTERNET Y ARTÍCULOS DE PRENSA

Montero Carretero, L, "*El fin de la era del todo vale para la IA"*, ABC Empresa, domingo 28 de mayo de 2023, pp.3

European Commission, *High-Level Expert Group on Artificial Intelligence: Definition of Artificial Intelligence*, disponible en https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf.

OCDE, *Recomendación sobre la inteligencia artificial*, 2019 (disponible en https://www.oecd.org

Pérez Bes, F., "El G20 busca un equilibrio entre regulación e innovación para la IA", *Demócrata*, 30 de noviembre de 2024, disponible en https://www.democrata.es/analisis-y-opinion/el-g20-busca-un-equilibrio-entre-regulacion-e-innovacion-para-la-ia/

UNESCO, *Recommendation on the Ethics of Artificial Intelligence*, 2021 (disponible en https://unesdoc.unesco.org

Nature Medicine, "Evaluation of a deep learning algorithm for detection of skin cancer in a clinical setting", *Nature Medicine*, vol. 26, n. 9, 2020, pp. 1344-1350 (disponible en https://www.nature.com/articles/s41591-020-1004-1

Cyberspace Administration of China, *Regulación de Algoritmos de IA*, 2022

European Union, *AI and Human Rights*, curso digital, disponible en https://help.elearning.ext.coe.int/user/view.php?id=210138&course=7331

Laboratoria, "10 casos donde la inteligencia artificial jugó en contra de la diversidad", *Laboratoria*, disponible en https://www.laboratoria.la/post/10-casos-donde-la-inteligencia-artificial-jugo-en-contra-de-la-diversidad

Smiley, A., "Uber's Self-Driving Car Didn't Know Pedestrians Could Jaywalk", *The Atlantic*, 2022 (disponible en https://www.theatlantic.com; última consulta 31/03/2025).

Salud y Fármacos, "La regulación de la inteligencia artificial en la asistencia sanitaria: aspectos éticos y legales", *Salud y Fármacos*, noviembre 2018 (disponible en https://www.saludyfarmacos.org/lang/es/boletin-farmacos/boletines/nov201804/20_la/; última consulta 31/03/2025).

UNESCO, *Recomendación sobre la ética de la inteligencia artificial*, 2021 (disponible en https://unesdoc.unesco.org/

Angwin, J., Larson, J., Mattu, S., & Kirchner, L., “Machine Bias”, *ProPublica*, 2016 (disponible en https://www.propublica.org