UNIVERSIDAD MIGUEL HERNÁNDEZ
FACULTAD DE CIENCIAS SOCIALES Y HUMANAS

TRABAJO FIN DE GRADO
EN SEGURIDAD PÚBLICA Y PRIVADA

Algoritmo, Ética y Riesgos de la Inteligencia Artificial: Implicaciones para los Derechos Humanos

Algorithm, Ethics, and Risks of Artificial Intelligence: Implications for Human Rights

UNIVERSITAS
Miguel Hernández

Curso académico 2024-25

- ✓ **AUTOR:** Sergio Martí Donet
- ✓ **TUTOR:** Rafael Cuesta Ávila

RESUMEN

Esta investigación analiza la intersección entre la Inteligencia Artificial (IA) y los Derechos Humanos, con un enfoque particular en el uso de algoritmos dentro de la Administración de Justicia, el cómo tecnologías como la vigilancia masiva, la predicción del delito y la gestión de riesgos afectan los derechos fundamentales, haciendo hincapié en la vulneración de derechos como la intimidad y la libertad de expresión, así como los dilemas éticos y los riesgos inherentes al uso de algoritmos, así mismo, se discuten las brechas regulatorias existentes y la urgencia de desarrollar marcos normativos que aseguren un uso responsable y ético de la IA, resaltando la importancia de proteger los derechos de los ciudadanos en la era digital mediante regulaciones que equilibren innovación tecnológica y respeto por los Derechos Humanos.

Palabras clave: Algoritmos, Inteligencia Artificial, Derechos Humanos, Ética Tecnológica

Abstract

This research analyses the intersection between Artificial Intelligence (IA) and Human Rights, with a particular focus on the use of algorithms in the administration of justice. It explores how technologies such as mass surveillance, crime prediction, and risk management affect fundamental rights, emphasizing the violation of rights such as privacy and freedom of expression, as well as the ethical dilemmas and risks inherent in the use of algorithms. Additionally, the study discusses existing regulatory gaps and the urgent need to develop normative frameworks that ensure the responsible and ethical use of IA, highlighting the importance of protecting citizens' rights in the digital age through regulations that balance technological innovation with respect for human rights.

Keywords: Algorithms, Artificial Intelligence, Human Rights, Technological Ethics.

Tabla de contenidos

1. Introducción.

Vivimos en una era marcada por avances tecnológicos sin precedentes, una era digital donde la inteligencia artificial (IA) ha dejado de ser un concepto reservado para la ciencia ficción para convertirse en una realidad cotidiana. Desde las recomendaciones personalizadas que recibimos plataformas como *Netflix*, *Youtube*, *Spotify* o incluso *Google*, hasta los más complejos algoritmos que ayudan a resolver problemas globales, la IA se ha instaurado en nuestras vidas, transformando no solo cómo interactuamos con el mundo, sino también cómo entendemos conceptos tan fundamentales como la justicia, la privacidad o la ética.

Sin embargo, este progreso plantea preguntas inquietantes: ¿qué ocurre cuando un algoritmo, diseñado para tomar decisiones *objetivas*, perpetúa prejuicios sociales? ¿Cómo podemos garantizar que los derechos humanos (DDHH), piedra angular de toda sociedad democrática, sean respetados en un entorno repleto de máquinas? La IA, aunque prometedora, no está exenta de riesgos, especialmente cuando su desarrollo y aplicación se enfrentan a marcos regulatorios aún en construcción.

Este trabajo se sitúa en la intersección entre tecnología, ética y DDHH. Si antes la IA era un concepto futurista, hoy se ha convertido en una parte integral de la vida cotidiana que ha transformado radicalmente el mundo laboral, académico, la interacción entre las personas e incluso, en cierta manera, el pensamiento. También ha pasado a generalizarse en todos los campos de la vida social: desde la vida cotidiana y doméstica, el campo de la medicina, la educación, las finanzas hasta el de la seguridad pública y, es que la IA ha marcado un antes y un después como el invento tecnológico del siglo XXI.

Hoy en día, la amplia variedad de aplicaciones de la IA hace que muchas veces no seamos conscientes de su presencia, puesto que su entrada y oleada al mundo ha sido casi inmediata y prácticamente imperceptible, sin embargo, esta misma razón, incluida su capacidad transformadora, por supuesto, plantea una

serie de retos y cuestiones que superan con creces la simple mejora en la eficiencia de nuestros sistemas y dispositivos y es que, como toda invención, siempre viene con novedad y un campo inexplorado en el cual la humanidad debe adaptarse y acoplar a su forma de vida con nuevas herramientas que le resulten provechosas, en lugar de contrarias a su bienestar.

La incorporación de IA en casi todos los ámbitos de la vida humana plantea innumerables posibilidades, pero no menos preocupaciones, ya que su funcionamiento se basa en la aplicación de una serie de algoritmos que analizan grandes cantidades de datos, hacen predicciones y toman decisiones de forma autónoma. Una inteligencia que toma sus propias elecciones como si se le diera criterio y pensamiento propio. Es precisamente esta funcionalidad la que sugiere ciertas cuestiones éticas de nuestro presente, pues específicamente en determinados procesos críticos como la justicia, la vigilancia o la gestión de riesgos, es el punto de partida donde convendría establecer con precisión los límites éticos que han de regir tanto su uso como su desarrollo.

No es casual que la IA no sea un ente "neutral" en absoluto: sus decisiones y predicciones dependen de los datos con los que ha sido entrenada y de los objetivos para los que ha sido diseñada, y por ello cualquier sesgo en los datos originales se reflejará más tarde en el comportamiento del algoritmo y, a menudo, se traducirá en consecuencias desiguales y perjudiciales para ciertos grupos que históricamente han sido relegados. Estas cuestiones revisten una especial relevancia en el ámbito de la justicia, dado que todos los procesos legales descansan en la aplicación rigurosa de los principios rectores del Estado de Derecho, entre ellos los de imparcialidad, igualdad ante la ley o la seguridad jurídica. Cuando los algoritmos operan sobre datos históricos cargados de prejuicios sociales, el resultado puede ser una perpetuación de esos mismos prejuicios, lo que plantea una disyuntiva ética que no se puede ignorar.

Desde el punto de vista de la seguridad pública, la IA es una herramienta especulativa que puede hacer prevalecer la política y la administración preventivas

del delito: como señala Zuboff (2020) en su libro *La era del capitalismo de vigilancia*, la IA es una "caja de Pandora" donde su uso corrupto y la falta de regulación conduce a la opresión, incluyendo la vigilancia masiva, de acuerdo con la cual los ciudadanos están constantemente bajo observación. Este concepto va en contra del principio de la sociedad democrática que garantiza la privacidad y la libertad. ¿Dónde se encuentra el límite entre la necesidad de seguridad y el respeto a los derechos fundamentales?

Según otros autores como Gil, López (2018), y otros, agregan que la falta de marco regulador empeora el asunto, dado que las grandes corporaciones y los estados tienden a inmiscuirse de forma desmesurada en la vida privada del individuo, vulnerando derechos fundamentales cuando no existen límites legales claros ni mecanismos de control y supervisión efecticos. Y es que su rápida inmersión ha hecho notar la falta de regulación, ya que la tecnología evoluciona van más rápidamente que la propia ley.

En el marco de un Estado democrático moderno estos derechos son pilares fundamentales para la convivencia libre y justa. Por lo tanto, antes de abordar la regulación de la tecnología en sí misma, es necesario cuestionar previamente las políticas que la rigen y los compromisos éticos asociados.

Este trabajo busca analizar más a fondo cómo la IA está afectando a diferentes regiones y barrios, a las administraciones públicas, y particularmente al sistema de justicia, así como a los derechos de los ciudadanos. Se pone énfasis en la necesidad de abordar estos dilemas desde una perspectiva crítica, considerando no solo los beneficios que la tecnología puede ofrecer, sino también los riesgos y posibles abusos que podrían surgir.

Por este motivo y como último aspecto del trabajo, se espera aplicar dicho análisis al plano práctico, a través de un cuestionario en un marco social y laboral pertinente al tema de estudio, todo ello con el fin de obtener opiniones genuinas sobre la percepción de la sociedad sobre el uso de la IA. Con esta aproximación, se espera abordar así no solo una perspectiva meramente académica, sino

también un enfoque íntimo y humano acerca del tema, teniendo en cuenta el hecho de que la IA ya no es solo una herramienta, sino un actor social en sí misma. Por ello, a través del referido cuestionario, se espera recibir visiones de primera mano sobre el panorama real de la IA en la sociedad, así como el rol de la seguridad en ella, y, más específicamente, cómo los ciudadanos comunes perciben, viven y valoran esta experiencia. Con esta aproximación se tratará de enriquecer esta con visiones auténticas que pudieran complementar las fuentes académicas al respecto, teniendo en cuenta que es una realidad tangible que moldea nuestra percepción acerca del sistema de seguridad y justicia.

En el ámbito de la Administración de Justicia y la protección de derechos fundamentales, su impacto se vuelve mayor ya que es cada vez más evidente y por lo tanto requiere un análisis riguroso, este trabajo busca explorar estas implicaciones en el contexto específico revisando consecuencias particulares en los DDHH, la vigilancia masiva, la predicción del delito y la gestión de riesgos, aspectos en los cuales los algoritmos de IA han comenzado a influir notablemente. En cada uno de estos casos no solo se encuentra en juego la eficiencia o precisión tecnológica, sino más bien el equilibrio entre seguridad y DDHH, un dilema que se presenta de manera constante en la era digital.

La importancia de este análisis radica precisamente en el nivel de poder de los algoritmos para moldear decisiones que afectan directamente la vida de los individuos, sobre todo en el contexto de la justicia y seguridad pública, donde los avances en IA permiten una capacidad de procesamiento y predicción sin precedentes, pero a su vez plantean también serios riesgos de vigilancia excesiva y discriminación, como señalan autores como Castells: la sociedad enfrenta una era de vigilancia, donde al acceso masivo a datos y su procesamiento automatizado pueden comprometer derechos básico de la población. Es precisamente esta dinámica la que requiere del análisis y atención de los especialistas del derecho, seguridad y prevención, dado que el mal uso de la IA podría transformar estas instituciones en herramientas de control, en lugar de ser garantes de libertad y justicia.

A la vista de todos estos desafíos, la necesidad de un marco de regulación sólido es innegable. De acuerdo con la Agencia de los Derechos Fundamentales de la Unión Europea (FRA), en un futuro inmediato y en ausencia de una regulación adecuada, los riesgos éticos y legales asociados con la IA solo van a incrementarse, corriendo el peligro de desbordar las capacidades de los sistemas actualmente existentes de protección de derechos. Un marco legal que establezca límites claros y fomente el uso responsable de la IA es un requisito previo para evitar que el desarrollo tecnológico conlleve regresiones sociales. En este sentido, este trabajo se propone no solo como un intento de comprender cómo opera la IA en el ámbito judicial; su objetivo esencial es insistir en la necesidad de proteger los DDHH en un tiempo donde la tecnología avanza más rápido que las leyes destinadas a regularla.

2. Marco teórico.

2.1. Definición de la inteligencia artificial (IA).

La IA se concibe como un campo de la informática que intenta crear sistemas que pueden realizar tareas propias sin necesidad del ser humano: análisis de datos, el reconocimiento de patrones y decisiones automatizadas. En su sentido más básico, la inteligencia trata de hacer que las máquinas tengan capacidad para aprender y adaptarse a las circunstancias experimentales, sin que se les den instrucciones explícitas. (Alonso Betanzos & Bolón Canedo, 2020)

2.2. Breve historia desde los orígenes a la actualidad de la IA.

Según tecnólogos, expertos y académicos, la IA está a punto de transformar el mundo que conocemos. Aunque corta, la historia de esta tecnología tiene un recorrido que parte de la primera mitad del siglo XX, con sus inicios de la mano de Alan Turing, para llegar hasta nuestros días, con el uso aplicado a escala cotidiana en nuestros dispositivos digitales. La expansión de la IA afecta a distintas dimensiones que van desde la antropológica hasta la ética, pasando por los cambios sociales, políticos, económicos o legales, entre tantos otros.

Para Brad Smith, presidente de Microsoft, la IA es la tecnología más importante que se ha creado desde la invención de la imprenta. Según el reconocido profesor de economía de Harvard, Dani Rodrik, galardonado en 2020 con el premio Princesa de Asturias, “si no reflexionamos a fondo, la IA tendrá efectos indeseables” (El País, 01.10.2023), de modo que está en nuestra mano moldear esta tecnología para evitar males mayores. El físico teórico británico Stephen Hawking expuso que “el éxito en la creación de la IA sería el evento más grande en la historia de la humanidad, pero desafortunadamente también podría ser el último, a menos que aprendamos a evitar sus riesgos”. Investigadores de la plantilla de OpenAI, fabricante del ChatGPT, alertaron que este poderoso descubrimiento tecnológico podría convertirse en una posible amenaza para la humanidad. De esta manera, estamos igual de fascinados y asustados con la revolución de la IA.

La IA no funciona únicamente en base a unas instrucciones predefinidas, sino que va actualizando su funcionamiento a medida que se integran nuevos datos, y precisamente esta capacidad de autoaprendizaje es la que la hace todavía más preocupante. Por consiguiente, es de esperar que la IA se convierta en una tecnología en continua mejora en cuanto a precisión y, según este argumento, si se procesan datos nuevos es probable que los sistemas de IA sean más eficientes con el pasar del tiempo. A su vez, esta capacidad de adaptación y aprendizaje puede convertirse también en un obstáculo a la hora de comprender y explicar procesos de decisión algorítmica, que es lo que Castells (2019) llama *opacidad algorítmica*, puesto que muchas decisiones resultantes de los sistemas de IA son difíciles de entender incluso para los propios programadores o diseñadores.

2.3. Tipos y lógicas de la IA.

Según la forma de procesar la información, la IA se divide en varias categorías, la frontera entre las cuales es muy difusa:

- **IA Clásica o Simbólica**: Surgida en los años 50 en Estados Unidos, utiliza sistemas de símbolos para resolver problemas.
- **Redes Neuronales Artificiales (RNA):** Inspiradas en el funcionamiento del cerebro, procesan información mediante interacciones entre unidades conectadas, empleando técnicas como aprendizaje profundo o automático. Este modelo tiene gran repercusión en psicología, neurociencia computacional y filosofía del lenguaje.
- **Autómatas Celulares**: Siguiendo reglas simples, simulan procesos biológicos como el desarrollo de organismos pluricelulares, donde cada unidad de procesamiento depende a su vez de las unidades vecinas. Este tipo de inteligencia ha aportado grandes avances en biología.
- **IA Evolutiva:** esta no solo se basa en los cálculos matemáticos (modelo conexionista) o en la lógica (modelo simbólico), sino que emula el modelo de la evolución biológica: para resolver un problema determinado, se parte de un programa inicial el cual es ineficiente, este programa se va modificando al azar y conforme a los principios de selección natural de Darwin dando lugar a versiones mejoradas del programa original, hasta que llegar a un programa que ya es capaz de resolver el problema inicial. Este programa se modifica por sí solo sin intervención humana, por lo que es imposible controlar todos sus pasos y detalles.

Con el tiempo la IA se ha transformado en una tecnología que tiene la capacidad de aprender, adaptarse y que es altamente eficaz para el análisis de datos; puede optimizar tareas repetitivas y procesos complejos los cuales requieren grandes cantidades de tiempo. Todo lo anterior la hace una herramienta práctica, principalmente en el campo de la minería de datos, el procesamiento de grandes volúmenes de información, aplicándola diferentes ámbitos como la medicina, la educación, la ingeniería o la seguridad. Esto se refleja en una integración de distintos algoritmos y procesos de aprendizaje automático, las cuales se basan en grandes volúmenes de datos, y en base a ellos se hacen predicciones con una gran precisión. (Alonso Betanzos & Bolón Canedo, 2020)

2.4. La dieta de datos que alimenta a la IA: big data y algoritmos

El ***big data*** se refiere al manejo y análisis de grandes volúmenes de datos que, debido a su tamaño, velocidad de generación y diversidad, no pueden ser procesados con herramientas convencionales. Su importancia radica en la capacidad de extraer información valiosa para la toma de decisiones, optimización de procesos y desarrollo de la IA.

La IA funciona gracias a los **algoritmos**, que son como su núcleo; una serie de instrucciones concretas, ordenadas, acotadas y finitas que sirven para resolver un problema o alcanzar un objetivo porque estos le permiten analizar datos anteriores, encontrar patrones y, con base en eso, tomar decisiones o hacer predicciones basadas en lo que ya ha aprendido. Y partiendo de este concepto la IA ha ido evolucionando a un ritmo exponencial en las últimas décadas.

Su influencia en la vida cotidiana y, especialmente, en áreas tan sensibles como la seguridad pública, la predicción de la criminalidad y la Administración de Justicia es cada vez más evidente y es donde se torna preocupante, puesto que los avances tecnológicos en cuanto a la IA están avanzando a pasos agigantados. Lo preocupante del modelo es que, mediante el uso del *big data,* esta acumulación masiva de información privada, puede ser procesada y utilizada de manera que puede violar los derechos de las personas sin que estas ni siquiera se den siquiera cuenta de ello. Si la eficiencia va de la mano de importantes dilemas éticos y legales, autores como Castells (2019) y Zuboff (2020) ya advierten que, aunque la IA puede convertirse en una poderosa herramienta, sin regulación suficiente aquélla resultará en prácticas de monitoreo y control masivo que amenacen la privacidad o derecho fundamental de los ciudadanos.

La inteligencia artificial se nutre de una ingente cantidad de datos que posibilitan su funcionamiento y evolución. Detrás de cada algoritmo que optimiza procesos y mejora la eficiencia de sistemas, subyace una realidad ineludible: la recolección masiva de información de los usuarios, en muchas ocasiones sin su conocimiento expreso y con fines que no siempre son transparentes. Este uso de

la información no solo incide en el ámbito privado de las personas, sino que también afecta ámbitos sensibles como la seguridad pública, la predicción del delito y la administración de justicia, generando interrogantes éticos y jurídicos de gran calado.

Estos sistemas procesan y analizan enormes cantidades de información en un tiempo récord, facilitando la detección de patrones y tendencias que luego se traducen en predicciones y toma de decisiones automatizadas. Por ejemplo, en el ámbito judicial esta tecnología promete optimizar procesos y hacerlos más eficientes. Sin embargo, la acumulación y el procesamiento masivo de datos privados pueden derivar en usos que vulneren derechos fundamentales sin que los individuos sean plenamente conscientes de ello.

La eficacia de la IA, cuando no va acompañada de una regulación adecuada, plantea dilemas complejos. Investigadores como Castells (2019) y Zuboff (2020) han advertido que, sin los controles apropiados, la IA podría derivar en sistemas de monitoreo y control masivo que amenacen el derecho a la privacidad y otras libertades individuales. El avance de los sistemas de vigilancia basados en IA ha incrementado su presencia en la vida cotidiana, permitiendo un rastreo constante y minucioso de los ciudadanos. Esto plantea una cuestión fundamental: ¿hasta qué punto estamos dispuestos a ceder nuestra privacidad en aras de la seguridad y la eficiencia?

Los algoritmos de IA han revolucionado los procesos de vigilancia, permitiendo identificar patrones de comportamiento, predecir amenazas potenciales y mejorar la respuesta de las fuerzas de seguridad. En teoría, esto podría traducirse en un refuerzo de la seguridad pública, pero, como señala Sánchez Almeida (2020), también conlleva importantes riesgos. La vigilancia masiva puede afectar derechos esenciales como la privacidad y la libertad de movimiento, especialmente cuando su aplicación se focaliza en determinados grupos poblacionales, favoreciendo la discriminación y debilitando la cohesión social.

En esta línea, Castells (2019) sostiene que la vigilancia impulsada por IA no se limita a la recopilación de datos, sino que permite un control social mucho más

profundo, capaz de influir en la autonomía de los ciudadanos. Esta situación resulta especialmente alarmante en países donde la privacidad y la libertad de expresión son pilares fundamentales de la sociedad democrática. A medida que los sistemas de vigilancia inteligentes almacenan y procesan datos sensibles, obligatoriamente hay que cuestionarse no solo quién accede a dicha información, sino también con qué propósito y bajo qué criterios o normas.

La repercusión de estas tecnologías ha reavivado el debate sobre el derecho a la privacidad en la era digital. La Agencia de Derechos Fundamentales de la Unión Europea ha alertado sobre la urgencia de establecer salvaguardias que protejan los datos personales y prevengan su uso indebido. La falta de transparencia en la recolección y el análisis de datos podrían derivar en serias vulneraciones de derechos. Sin una regulación clara y efectiva, la vigilancia masiva podría dar lugar a nuevas formas de control social que operen en la sombra, escapando a la vista y control de la ciudadanía.

En el contexto de la seguridad pública, se nos vende la idea de que estas herramientas pueden mejorar la prevención del delito y fortalecer la protección ciudadana. No obstante, la eficiencia tecnológica debe ir siempre acompañada de una supervisión ética y legal que garantice el respeto a los derechos fundamentales. Solo mediante un equilibrio adecuado entre innovación y regulación se podrá aprovechar el potencial de la IA sin comprometer los principios democráticos ni la dignidad de las personas.

3. Riesgos algorítmicos derivados de la IA.

A pesar de que el desarrollo y la evolución de la IA está impulsando avances que transforman nuestra vida cotidiana, que van desde la optimización de procesos hasta la automatización en la toma de decisiones esta evolución tecnológica también conlleva riesgos importantes, sobre todo cuando los algoritmos que la sustentan operan con sesgos, carecen de transparencia o se implementan sin un marco regulador adecuado. Por ello, la seguridad en el uso de la IA debe estar en sintonía con los principios de justicia, equidad y respeto por los derechos fundamentales, evitando así consecuencias negativas para la sociedad.

Uno de los principales obstáculos de la inteligencia artificial es la falta de explicabilidad de muchos de sus sistemas. En numerosas ocasiones, los algoritmos operan como una "caja negra", donde las decisiones resultan de procesos internos tan complejos que ni siquiera sus propios desarrolladores pueden explicarlos con precisión. Esta opacidad genera una cuestión clave: si una decisión tomada por la IA afecta negativamente a una persona o a un grupo, ¿quién es el responsable? Sin un marco claro de rendición de cuentas, pueden perpetuarse injusticias sin mecanismos eficaces para corregirlas.

Otro riesgo que se observa es la amplificación de sesgos preexistentes en los datos con los que se entrenan estos sistemas. Si la información utilizada refleja desigualdades estructurales o estereotipos discriminatorios, los algoritmos no solo los replican, sino que pueden intensificarlos. Esto es especialmente preocupante en ámbitos como la contratación laboral, la concesión de créditos o la seguridad pública, donde una decisión sesgada puede tener consecuencias directas en la vida de las personas. En lugar de actuar como herramientas imparciales, los sistemas de IA pueden reforzar dinámicas de exclusión si no se diseñan con mecanismos efectivos para mitigar estos sesgos.

Asimismo, el creciente uso de la inteligencia artificial en la toma de decisiones críticas plantea un desafío adicional. A medida que delegamos más responsabilidades en estos sistemas, existe el riesgo de que se reduzca la supervisión humana o de que se confíe excesivamente en su supuesta objetividad. En sectores como la justicia, la salud o la seguridad, donde las decisiones pueden afectar directamente los derechos y el bienestar de las personas, es fundamental que la IA funcione como un apoyo a la evaluación humana, y no como un sustituto de ella.

Otro aspecto que no puede pasarse por alto es el uso de la inteligencia artificial en la vigilancia y el control social. La recopilación masiva de datos personales, combinada con tecnologías como el reconocimiento facial o los sistemas predictivos de comportamiento, podría derivar en un modelo de monitoreo que vulnere la privacidad y la libertad individual. Si no existen regulaciones claras, estas herramientas pueden ser utilizadas con fines que

trascienden su propósito inicial, generando un escenario en el que la vigilancia constante se normalice y los derechos fundamentales se vean comprometidos.

Es imprescindible fomentar que estos sistemas garanticen la transparencia en su funcionamiento, asegurar la supervisión humana en decisiones críticas y promover un uso responsable que priorice la dignidad y los derechos de las personas.

La seguridad de la IA debe ponerse al servicio de la humanidad y del bien común para impedir que se violen derechos fundamentales de los ciudadanos que introduzcan consecuencias indeseables

3.1. Falta de veracidad: generación de bulos, fake news y postverdad.

En los últimos años la desinformación ha ido en aumento y se está convirtiendo en un grave problema ya que pone en riesgo la calidad de los medios de comunicación tradicionales y por ende de los sistemas democráticos. Con la llegada de la IA la creación y difusión de las fake news se ha vuelto más sencilla y sofisticada, lo que facilita la manipulación de la opinión pública a gran escala. Ejemplos recientes, como el referéndum del Brexit o las recientes elecciones presidenciales en Estados Unidos, han demostrado cómo estas herramientas pueden influir en la percepción ciudadana e incluso en el resultado de procesos electorales. Cada día la creación y distribución de las *fake news* y las suplantaciones de identidad y de voz mediante aplicaciones de IA cada vez más frecuente y más accesible al público en general con el consiguiente riesgo de un uso malintencionado.

El problema no radica únicamente en la facilidad con la que se pueden generar bulos, sino también en la capacidad de las plataformas digitales para dirigir estos contenidos a audiencias específicas, reforzando prejuicios y creencias preexistentes. En este escenario, la postverdad se afianza como un modelo en el que las emociones y opiniones personales adquieren más peso que los hechos verificables, debilitando el debate público y la toma de decisiones informadas.

Si esta tendencia sigue sin mecanismos de control y sin estrategias para fomentar una información veraz y contrastada, las democracias podrían

transformarse en sistemas donde el poder de decisión recaiga en manos las grandes tecnológicas o grupos con intereses particulares. En consecuencia, se corre el riesgo de que la ciudadanía pierda protagonismo en el debate político, abriendo la puerta a modelos de gobierno menos transparentes y más alejados de los principios democráticos.

Como hemos visto en el terreno de la postverdad la opinión y la manipulación pesan más que los hechos reales y verificables, lo que supone una merma del pensamiento crítico y de los valores democráticos de la sociedad.

3.2. Falta de diversidad: estandarización de la información generada.

La recopilación y procesamiento de datos, así como el uso de IAG (inteligencia artificial generativa) reflejan una predominancia de los grupos con mayor poder y representación en la sociedad, como los hombres blancos y asiáticos. Esta tendencia se debe a que los datos utilizados para entrenar los modelos de IA suelen ser seleccionados por individuos pertenecientes a estos grupos, lo que puede resultar en una representación sesgada de la realidad. Por ejemplo, en el caso de sistemas de detección de rostros mediante IA, si un algoritmo se entrena mayoritariamente con un perfil caucásico podría verse afectada la relación entre la exactitud de la identificación y las variables de evaluación, siendo más difícil el reconocimiento de rostros de tipo asiático o africano.

Por consiguiente, la información generada por la IA tiende a replicar y amplificar los sesgos presentes en los datos de entrenamiento, perpetuando así una visión homogénea que favorece a los grupos mayoritarios y marginaliza a las minorías. Este fenómeno puede dar lugar a la imposición de un canon que no refleja la diversidad de perspectivas y experiencias presentes en la sociedad, lo que limita la capacidad de la IA para abordar de manera equitativa las necesidades y realidades de todos los grupos sociales.

Para mitigar estos sesgos y promover una representación más inclusiva, es esencial que los equipos encargados del desarrollo y la implementación de sistemas de IA estén compuestos por individuos de diversos orígenes, géneros,

etnias y perspectivas. La inclusión de una variedad de voces y experiencias en el proceso de diseño y entrenamiento de la IA contribuye a la creación de modelos más justos y representativos, capaces de reflejar la complejidad y pluralidad de la sociedad en su conjunto.

En resumen, la falta de diversidad tanto en la recopilación de datos como en los equipos de desarrollo de IA puede contribuir a la estandarización de la información generada, y favorecer a ciertos grupos a la vez que margina a otros más minoritarios.

Al igual que en la sociedad actual la promoción de la inclusión y la diversidad es esencial para garantizar que estas tecnologías sirvan de manera equitativa

Imagen: elpais.com

3.3. Falta de transparencia: desconocimiento de las fuentes de origen.

El sistema de la IA genera un resultado a partir de información que extrae o copia de materiales que ya están protegidos, incurriendo en una posible violación de la ley, ya que de esta forma estaría conculcando los derechos de autor (*copy right*) y la protección de datos, los cual hace patente una falta de regulación que garantice la seguridad de los autores y actores.

Otro de sus principales problemas es la falta de claridad sobre el origen de la información que emplea para generar sus respuestas. La ausencia de un mecanismo que garantice la trazabilidad de estos contenidos no solo plantea interrogantes legales, sino que también compromete la autenticidad y fiabilidad de los resultados que la IA ofrece.

Este desconocimiento sobre las fuentes originales dificulta la identificación de posibles sesgos, errores o incluso la reproducción indebida de material ajeno sin el consentimiento de sus creadores. Como consecuencia, se ha abierto un debate sobre la necesidad de establecer una normativa que regule la transparencia y trazabilidad en el uso de datos por parte de la IA. Garantizar que los autores y propietarios de los contenidos reciban el reconocimiento adecuado a la vez que los usuarios puedan confiar en la validez de la información generada debería ser una de las prioridades en el desarrollo responsable de la IA.

3.4. Falta de equidad: sesgos contra colectivos vulnerables.

Lo que resulta evidente es que, en general, los algoritmos pueden parecer neutros, pero todos ellos están impregnados de los prejuicios que sociedad ha depositado en los datos históricos. Ciertamente, no debiera extrañarnos en absoluto que para estos sistemas de IA resultará más fácil utilizar datos que coincidan con la realidad, y la nuestra, por desgracia, no siempre es la más justa.

El sesgo en el proceso de toma de decisiones por parte del sistema legal, por ejemplo, es que al hacerlo el algoritmo hereda (innumerables) características que pueden establecerse a través de los datos históricos. Todo esto nos advierte de que su utilización bien nos conduce a juicios de moralidad e injusticias. (Gil, 2021)

Sí los datos recopilados de antecedentes penales muestran una tendencia a detener a mujeres que son de minorías étnicas y/o se encuentran en situaciones económicas más bajas, el sistema lo detecta y de ese modo establece las pautas para llegar a tratar a esas mujeres como altamente riesgosas en base a tales características. En este ejemplo, el algoritmo no está discriminando

deliberadamente, tan sólo está repitiendo una característica que ya tenía de antemano. El problema es aún mayor en los casos donde se predicen futuras reincidencias, como nos advierte la Agencia de los Derechos Fundamentales de la Unión Europea (2021), si simplemente se entrenan algoritmos, hay que tener un proceso de supervisión para ayudarlos a evitar que sean discriminatorios o perjudiciales.

De acuerdo a Gil (2021), con su 'carga de sesgo', el uso de algoritmos en el sistema judicial puede representar un problema ético si éstos no son supervisados por parte de humanos. Este tipo de sesgos, que influye en las decisiones sobre las personas, puede llegar a la conclusión de que determinadas conductas sean evaluadas como más peligrosas y repitiendo así el ciclo y realimentándose la desviación. Por ello toda tecnología ha de ser desarrollada bajo principios éticos y con la suficiente supervisión humana, a fin de evitar que puedan darse situaciones de injusticia o perjudiciales para los ciudadanos.

Un estudio de la Unesco publicado a principios de febrero de 2024, demuestra como la IA muestra prejuicios raciales y homofobia contra colectivos vulnerables como el femenino. Por ejemplo la IA confirma y reafirma los estereotipos contra las mujeres, asociadas a roles domésticos y vinculadas a palabras como hogar, familia y niños, mientras que los papeles masculinos se asocian a términos como negocios, ejecutivo, salario y carrera.

Cabe exigir la urgente necesidad de corregir sesgos algorítmicos para decidir sobre si los bancos conceden préstamos, seguros o selección de contratación empleos pues los datos de entrenamiento sobrevaloran a los hombres, blancos y heterosexuales. El problema para una de las mayores expertas en ética y tecnología, Margaret Michell (El País, 27.11.23) es que "las personas a las que más puede perjudicar la IA no deciden sobre su regulación".

3.5. Falta de privacidad: la intimidad deja de ser un derecho individual

El temor a una inteligencia artificial masiva e invasiva, cuya intervención no permita resistencia alguna, está propiciando el auge de un "capitalismo de la vigilancia". Aunque en los países occidentales este fenómeno ya se manifiesta mediante una opacidad informativa por parte del Mercado y el Estado, su desarrollo más extremo se observa en naciones como China, donde se ha implementado un sistema de créditos sociales basado en el comportamiento. En este contexto, el cibercontrol y la tecno-vigilancia se han normalizado, especialmente en una población que ha vivido bajo regímenes autocráticos, desde la era imperial hasta el Partido Comunista Chino (PCCh). En este escenario, la inteligencia artificial podría ampliar y profundizar la vigilancia masiva, amenazando así la privacidad individual.

El hecho de que la IA sea extraordinariamente eficaz en el procesamiento de extensos volúmenes de información hace que los sistemas que la incorporan sean idóneos para perseguir diferentes objetivos en el tratamiento de la justicia y seguridad pública.

Por otra parte, esta capacidad para procesar datos a gran escala también plantea dudas en el ámbito de la ética, puesto que como se ha mencionado, su uso masivo de datos puede conducir a una pérdida significativa de la privacidad personal, recopilando datos para sus propias predicciones y optimizando su propia actuación dependiendo de datos relacionados con la personas. Este elemento de la IA, en el sentido de la eficiencia antes referida, aunque útil para obtener mayores rendimientos, también plantea el dilema en relación con el sacrificio de la privacidad de los ciudadanos en aras de una supuesta mejora de la seguridad. (Zuboff 2020)

Como hemos visto la implementación de algoritmos inteligentes en Justicia no solo afecta a la forma en que esta institución lleva a cabo sus procesos internos sino que tiene un efecto notable sobre la percepción de los propios ciudadanos en relación con lo que supuestamente deben esperar de la Justicia, pues la expansión del *capitalismo del monitoreo* ha propiciado que los propios ciudadanos

sean cada vez más conscientes del uso de sus datos personales en ámbitos que anteriormente parecían estar reservados únicamente a la privacidad. Por este motivo, esta forma de percibir la vigilancia puede deteriorar la representación que deben tener los ciudadanos de la Justicia y la supuesta imparcialidad que debe existir en las instituciones judiciales, sobre todo en el caso de que los ciudadanos se perciban como objeto de un uso indiscriminado de sus datos, Zuboff, 2020)

En definitiva aunque su uso presenta una serie de beneficios en términos de eficacia y optimización de recursos también plantea algunos problemas y cuestiones importantes en cuanto a la protección de la privacidad. La lógica algorítmica puede beneficiar a aumentar la eficiencia en la toma de decisiones pero debe ser usada con precaución para evitar violaciones a los derechos fundamentales.

4. La aplicación de la IA sobre el campo de la seguridad.

Entre las aplicaciones más destacadas se encuentra la predicción del delito, que permite a las administraciones públicas y a las Fuerzas y Cuerpos de Seguridad (FCS), utilizando datos históricos y patrones delictivos, identificar áreas con mayor incidencia criminal y, en algunos casos, prever futuros comportamientos delictivos.

La IA también permite a FCS poder distribuir los recursos de manera más eficiente y realizar una labor más focalizada y precisa. Sin embargo, esta práctica también ha sido tachada negativamente por su propia potencialidad para perpetuar sesgos de tipo racial, étnico y socioeconómico, puesto que los datos históricos sobre la criminalidad están inevitablemente marcados por factores sociales y económicos y es por ello que ciertos grupos o comunidades pasan a estar desproporcionadamente afectados por las medidas de vigilancia predictiva. Todo esto conllevando a un problema ético importante, infringiendo en la equidad y justicia en el uso de la IA en el ámbito de la seguridad pública. (Castro, 2018)

En cuanto a la seguridad pública y control del tráfico, los sistemas inteligentes son altamente efectivos y los sistemas videovigilancia inteligente ya permiten supervisión en tiempo real y resolución automática de comportamientos sospechosos o hechos antirreglamentarios. Un ejemplo actual es la próxima incorporación de 25 cámaras con inteligencia artificial en la Policía Local de Valencia para monitorizar en tiempo real la ocupación de los estacionamientos de carga y descarga (*Las Provincias. 2024*). Si bien se puede prever un aumento de la seguridad, a su vez también y al mismo tiempo esto abre la puerta a una videovigilancia masiva que podría infringir derechos fundamentales (*Agencia de los Derechos Fundamentales de la Unión Europea, 2021*)

Por ejemplo, si las FCS disponen de un programa inteligente para la identificación de personas sospechosas basado en el *big data* obtenido a partir de archivos policiales previos, las FCS pueden saber rápidamente cuales son los *hot spots* o puntos calientes, las zonas más conflictivas e incluso el posible aspecto físico de los delincuentes. Este tipo de sistemas no atribuyen ningún delito o hecho incívico antes de que se haya cometido y contribuyen a una mejor gestión policial y en definitiva a una mejor calidad de vida del ciudadano. Pero al analizar con detalle estos sistemas podemos observar que debido a los diferentes sesgos raciales, socioeconómicos, de nacionalidad etc. se producen alteraciones que pueden modificar la relación de causalidad de estos hechos. Es decir, el algoritmo interpreta los datos para determinados colectivos antes de que el delito suceda, pero de esta manera se condena a determinados colectivos a ser el foco de la policía y determina donde se concentrarán los cacheos e identificaciones. Además, puesto que el modelo se retroalimenta y aprende de sus errores, estos finalmente se acentúan todavía más.

Estos sistemas de gestión policial se presentan bajo el amparo de la ciencia como sistemas más justos, imparciales e implacables, pero al final esta *justicia artificial* desemboca en mayores desigualdades e injusticias sociales, y aunque la intención de las FCS es legítima el resultado final tiene el efecto contrario.

El uso de inteligencia artificial en aplicaciones militares, como los programas israelíes Lavender (Lavanda) y Pegasus, representa un claro ejemplo de cómo se prioriza la seguridad de unos frente a la vulnerabilidad y la inseguridad de otros. El primero de estos programas, diseñado para identificar objetivos terroristas de grupos como Hamás o el autodenominado Estado Islámico (ISIS), ha tenido como consecuencia la muerte de miles de civiles palestinos sin vínculo alguno con dichas organizaciones, quienes han sido tratados como meros datos estadísticos en el sistema. Este tipo de intervenciones, basadas en algoritmos de inteligencia artificial, obvian la distinción entre combatientes y civiles, considerando a las víctimas como daños colaterales no significativos.

El uso de misiles, drones y armas operadas mediante IA y completamente autónomas es una de las aplicaciones de la IA que está generando mayor inquietud. Entre los mayores expertos a nivel mundial en esta materia existe un consenso casi unánime sobre los riesgos que plantea esta tecnología.

Si bien actualmente estos sistemas operan bajo supervisión humana, su eventual autonomía podría implicar consecuencias fatales en caso de errores, poniendo en riesgo la vida de civiles y provocando graves violaciones a los derechos humanos. En este sentido, existe un consenso creciente entre los expertos internacionales sobre los peligros inherentes a estas tecnologías.

La comunidad global de especialistas en IA ha emitido un manifiesto a través del *Future of Life Institute*, alertando sobre los riesgos éticos, humanitarios y de seguridad asociados al uso de armamento autónomo. Este manifiesto hace un llamado urgente a la regulación y limitación de estos sistemas, con el fin de evitar consecuencias irreversibles para la humanidad.

4.1. Sobre la predicción del delito: ¿hacia Minority Report?

Entre las aplicaciones más destacadas se encuentra la predicción del delito, que permite a las Fuerzas y Cuerpos de Seguridad (FCS) y a las administraciones públicas identificar áreas con alta incidencia criminal utilizando datos históricos y patrones delictivos. En algunos casos, incluso es posible prever futuros comportamientos delictivos. Gracias a esta tecnología, las FCS pueden distribuir

sus recursos de manera más eficiente y focalizada, mejorando la gestión policial y la calidad de vida del ciudadano.

Sin embargo, la predicción del delito presenta serias preocupaciones éticas. El uso de datos históricos sobre criminalidad, que a menudo reflejan factores sociales y económicos, puede perpetuar sesgos raciales, étnicos y socioeconómicos. Como resultado, ciertos grupos o comunidades pueden ser desproporcionadamente afectados por las medidas de vigilancia predictiva, lo que genera un problema de equidad y justicia en el uso de la IA en el ámbito de la seguridad pública (Castro, 2018).

La capacidad predictiva de la IA plantea preguntas fundamentales sobre la intervención tecnológica en los aspectos más íntimos de la vida social. ¿Es realmente posible anticipar el crimen con precisión y justicia? ¿En qué medida debemos confiar en los algoritmos para modelar la seguridad, y cómo se garantizan, al mismo tiempo, los derechos de los ciudadanos? Mientras que la IA se presenta como una herramienta milagrosa para la vigilancia eficiente a gran escala, su implementación requiere una profunda reflexión sobre el equilibrio entre el progreso tecnológico y la protección de los valores humanos, ya que existe el riesgo de caer en un estado de vigilancia constante que podría socavar los derechos fundamentales de libertad individual e incluso la propia democracia.

Por otro lado, el uso de algoritmos en la justicia y la seguridad plantea un desafío relacionado con la responsabilidad algorítmica, que aún se encuentra en desarrollo. Según López (2022), muchas veces es difícil identificar quién es el responsable de una decisión tomada por un sistema de IA, especialmente cuando se trata de algoritmos de aprendizaje profundo. Esta falta de claridad sobre la rendición de cuentas es problemática, sobre todo si las decisiones tomadas por la IA afectan gravemente a la vida de los ciudadanos y no existe alguien que pueda rendir cuentas por un error o injusticia cometida.

Un ejemplo claro de esto es el uso de algoritmos predictivos basados en big data, utilizados por las FCS para identificar personas sospechosas y detectar puntos calientes, es decir, zonas con mayor probabilidad de incidentes delictivos. Estos sistemas no atribuyen delitos antes de que ocurran, pero permiten una

mejor gestión policial y un mayor control. Sin embargo, la realidad es que, al analizar estos sistemas con detenimiento, se observa que pueden alterar la relación de causalidad, afectando especialmente a colectivos marginados por sesgos raciales, socioeconómicos y de nacionalidad. Los algoritmos, al aprender de sus errores y retroalimentarse, tienden a reforzar estos sesgos, condenando a ciertos colectivos a ser objeto de un mayor escrutinio policial. Esto resulta en un círculo vicioso donde los errores cometidos por el sistema son amplificados y perpetuados.

Estos sistemas, aunque presentados bajo el amparo de la ciencia como justos, imparciales e infalibles, pueden en realidad acentuar las desigualdades sociales. La intención de las FCS de utilizar herramientas tecnológicas para mejorar la seguridad es legítima, pero el resultado final podría ser todo lo contrario: una mayor injusticia y desigualdad.

La reflexión ética y la regulación adecuada serán clave para evitar que avancemos hacia un futuro de predicción del crimen similar al descrito en *Minority Report*, donde la tecnología, en lugar de proteger, pueda terminar vulnerando los derechos de los ciudadanos.

4.2. Sobre la detección de áreas urbanas de alto riesgo delictivo: puntos calientes.

Por ejemplo, si las FCS disponen de un programa inteligente para la identificación de personas sospechosas basado en el *big data* obtenido a partir de archivos policiales previos, las FCS pueden saber rápidamente cuales son los *hot spots* o puntos calientes, las zonas más conflictivas e incluso el posible aspecto físico de los delincuentes. Este tipo de sistemas no atribuyen ningún delito o hecho incívico antes de que se haya cometido y contribuyen a una mejor gestión policial y en definitiva a una mejor calidad de vida del ciudadano. Pero al analizar con detalle estos sistemas podemos observar que debido a los diferentes sesgos raciales, socioeconómicos, de nacionalidad etc. se producen alteraciones que pueden modificar la relación de causalidad de estos hechos. Es decir, el algoritmo interpreta los datos para determinados colectivos antes de que el delito suceda,

pero de esta manera se condena a determinados colectivos a ser el foco de la policía y determina donde se concentrarán los cacheos e identificaciones. Además, puesto que el modelo se retroalimenta y aprende de sus errores, estos finalmente se acentúan todavía más.

Estos sistemas de gestión policial se presentan bajo el amparo de la ciencia como sistemas más justos, imparciales e implacables, pero al final esta *justicia artificial* desemboca en mayores desigualdades e injusticias sociales, y aunque la intención de las FCS es legítima el resultado final tiene el efecto contrario.

4.3. Sobre la evaluación de la valoración policial del riesgo de maltrato (VPR).

La Valoración Policial del Riesgo (VPR) es un procedimiento que permite a los agentes de policía determinar el nivel de peligro que una persona o situación puede representar para la seguridad pública. Se aplica principalmente en casos de violencia de género y doméstica. Su finalidad es identificar posibles factores de riesgo, establecer las medidas preventivas necesarias y tomar decisiones informadas que garanticen la correcta protección de la víctima, asegurando una respuesta proporcional y adecuada a cada caso.

La aplicación de algoritmos de inteligencia artificial en el ámbito policial también ha llevado a una mejora en la eficacia en este sector, y ahora se están empezando a implementar sistemas de IA en procesos de VPR y VPER. Esta incorporación de la IA en casos de violencia de género ha supuesto un avance en la protección de las víctimas, sin embargo aún es necesario abordar diversos aspectos, tanto legales como éticos, algunos de los cuales ya hemos mencionado previamente.

La filosofía de estos sistemas es analizar datos como antecedentes penales, perfiles sociodemográficos y conducta sospechosa, y prever con ello el riesgo de delincuencia y/o reincidencia. De esta forma se pretende, con un menor gasto en recursos judiciales, obligar a los organismos de justicia a que traten con

mayor *eficacia* estos casos, y es aquí donde este uso de algoritmos trae consigo preguntas complejas relacionadas con la indiferencia y la fiabilidad. (Gil, 2022)

En España el sistema VioGén está integrando algoritmos de IA para analizar grandes volúmenes de datos y predecir con mayor precisión la probabilidad de reincidencia de agresiones. Con ello se pretende una evaluación más rápida y precisa de los riesgos, facilitando la asignación de recursos y medidas de protección adecuadas. Por ejemplo, la Secretaría de Estado de Seguridad (SES) del Ministerio del Interior ha incorporado la analítica avanzada y la inteligencia artificial para reforzar la eficiencia del proceso de VPR en los casos de violencia de género, con el objetivo de mejorar la predicción de posibles agresiones reincidentes y proporcionar una protección más personalizada a las víctimas.

Además, la IA ha permitido la creación de aplicaciones y *chatbots* que proporcionan apoyo y orientación a las víctimas de violencia de género, como AinhoAid en Valencia, diseñada para ayudar a las mujeres que han sido víctimas de violencia de género en el proceso de denuncia, ya que muchas de ellas no se atreven a dar este paso.

Pero no podemos olvidar la importancia de que en este proceso asistido se garantice la protección de los derechos de las víctimas y la transparencia en los procesos de toma de decisiones.

4.4. Sobre la legitimidad democrática de la presunción de inocencia.

Para tratar de eliminar los sesgos en la inteligencia artificial, es necesario establecer un razonamiento filosófico tanto para los creadores de IA como para el legislador, ya que una educación en este sentido hará que se tomen conciencia de los posibles riesgos. Este camino no es sencillo y requiere de una voluntad política que impulse tanto la incorporación de normativa como de los mecanismos que garanticen su cumplimiento.

Los derechos humanos (DDHH) son fundamentales para detectar y corregir cualquier tipo de desviación en el desarrollo de la IA. Por ello, sería recomendable revisar la Declaración Universal de los Derechos Humanos (1948) para incorporar nuevos derechos digitales, como el derecho a la privacidad, la transparencia, la seguridad digital o el derecho al olvido, entre otros. No deben quedar como un mero reconocimiento teórico; su defensa debe ser real y efectiva. Los DDHH deben jugar un papel central en toda posible regulación sobre IA, pues esta afecta a aspectos tan relevantes como la privacidad, la libertad, la dignidad y la justicia. Aunque la IA puede contribuir a mejorar la calidad de vida, su utilización debe siempre ceñirse a los derechos fundamentales. No es una mera consideración ética; es un compromiso con los valores democráticos que sustentan nuestra sociedad.

Uno de los derechos fundamentales que debe protegerse en la regulación de la IA es el derecho a la privacidad, un pilar esencial de la democracia y constantemente amenazado por la IA. Los algoritmos tienen la capacidad de recoger y analizar datos a gran escala, lo que presenta riesgos para la privacidad de la ciudadanía. Por tanto, cualquier posible regulación de la IA debe incluir límites sobre cómo, qué y quién puede acceder a los datos y bajo qué condiciones. Los ciudadanos deben tener control sobre sus propios datos y el derecho de elegir qué revelar o cómo serán utilizados (Sánchez Almeida, 2020).

Otro derecho clave es la igualdad y la no discriminación. Como se ha visto, los algoritmos pueden reproducir y amplificar sesgos presentes en los datos con los que fueron entrenados, lo que puede llevar a decisiones discriminatorias que afecten a minorías o grupos vulnerables.

Castells (2019) explica que, si la IA no se supervisa adecuadamente, podría generar una "nueva forma de control social" en la que ciertos colectivos sean tratados con mayor severidad. Por ello, es crucial que las regulaciones incluyan medidas para detectar y corregir estos sesgos, asegurando que todos los ciudadanos sean tratados con justicia y equidad.

La presunción de inocencia, uno de los principios más fundamentales en el ámbito judicial, también se ve amenazada por la IA. Los sistemas de IA que

intervienen en procesos judiciales deben ser suficientemente transparentes y explicables, para que los afectados puedan comprender las razones de una decisión y, si es necesario, impugnarlas. La Agencia de los Derechos Fundamentales de la Unión Europea (2021) ha insistido en que el uso de la IA en el sistema judicial no debe comprometer la presunción de inocencia ni el derecho a una tutela judicial efectiva. La Justicia no debe ser solo eficiente, sino también accesible y comprensible.

Además, los DDHH deben ser el núcleo de cualquier legislación sobre IA, siendo el punto de partida para el desarrollo de toda tecnología, incluso las que avanzan rápidamente. Estos derechos son innegociables y deben prevalecer sobre la eficiencia, la conveniencia o la economía. Al colocarlos en el centro de la regulación, podemos construir una IA que respete la dignidad y la libertad de todas las personas, asegurando que esta herramienta sirva para mejorar nuestras vidas sin comprometer los valores esenciales de nuestra sociedad.

Por supuesto, la IA puede transformar el funcionamiento de la vigilancia y la justicia, pero la cuestión es si dicha transformación será positiva o negativa. Es aquí donde entran en juego diferentes cuestiones éticas, que deben ser revisadas y examinadas detenidamente.

Autores como Atawa analizan cómo la IA no solo puede realizar tareas cotidianas, sino que también es capaz de realizar razonamientos no lógicos que pueden influir en derechos fundamentales. Lejos de ser una simple mejora tecnológica, la IA puede penetrar en los procesos humanos desde los más simples hasta los más complejos, y ahora tiene la capacidad de recoger y analizar información privada sin el consentimiento de los ciudadanos, lo que vulnera el derecho a la privacidad. Esta tecnología también tiene el potencial de transformar la vigilancia y la administración de la justicia. Según la Agencia de los Derechos Fundamentales de la Unión Europea (2021), existen sistemas de IA capaces de procesar datos a una velocidad y precisión que superan la capacidad humana, lo que permite un análisis exhaustivo en tiempo real. Sin embargo, este poder de procesamiento viene asociado con riesgos. La IA, al identificar patrones basados en datos históricos, puede llevar a situaciones de "profecía autocumplida", en las

que ciertos grupos estigmatizados por categorías delictivas se sometan a un nivel de vigilancia desproporcionado (Gil, 2021).

La definición de IA en el ámbito de la seguridad y la justicia es compleja. No se trata de una simple mejora de la eficiencia, sino de una tecnología capaz de influir en la existencia y libertad de las personas. Su capacidad de recoger, procesar y analizar información privada contradice el derecho a la privacidad. Además, su uso para predecir delitos pone en duda la presunción de inocencia, un principio fundamental en el ámbito judicial.

La IA también está vinculada al concepto de capitalismo de vigilancia, como lo explica Zuboff, quien señala que la recopilación masiva de datos y la predicción de comportamientos son factores que ponen en peligro tanto la democracia como la privacidad del individuo. Si no se regula, el uso de la IA para la vigilancia y la administración de la justicia podría desembocar en una vigilancia masiva que amenace la libertad de expresión y la privacidad.

Por otro lado, Castells (2019) argumenta que la IA forma parte de un proceso más amplio de digitalización y control social. Es un mecanismo de poder que refuerza la capacidad de los Estados y las corporaciones para influir e intervenir en los comportamientos sociales mediante la vigilancia constante y la toma de decisiones no siempre transparentes. Este contexto hace evidente la necesidad de regular la IA para proteger los derechos fundamentales.

De Castro (2018) sugiere que la IA está ampliando el concepto de justicia en una "sociedad de control", donde la tecnología no solo previene la delincuencia, sino que puede anticipar conductas potencialmente delictivas. Esto plantea una cuestión clave: si la IA se define en términos de seguridad y justicia, es una tecnología con un potencial transformador, pero también con un alto riesgo de revocar derechos fundamentales. A medida que los sistemas de IA evolucionan, la sociedad y el ámbito jurídico deben adaptarse, estableciendo principios éticos que orienten su desarrollo y aplicación responsable.

5. El desarrollo de la IA sobre el terreno de la justicia artificial.

5.1. Sobre los programas carcelarios de concesión de libertad condicional.

El impacto de la IA en la Administración de Justicia es notable i abarca varios aspectos clave como la eficiencia procesal y el apoyo en la toma de decisiones, optimizando la carga de trabajo en los juzgados. Durante todo el proceso judicial, desde las primeras diligencias hasta la sentencia final, los sistemas de IA pueden asistir a jueces y fiscales en el proceso de toma de decisiones, identificar documentos legales, analizar riegos y en definitiva apoyarlos en su labor. Un ejemplo es el proyecto RisCanvi, diseñado para dar respuesta a la evaluación y gestión del riesgo de toda la población penitenciaria en Cataluña y ayudar a los jueces a decidir sobre la libertad condicional, el cual presenta tanto ventajas como riesgos en su implementación. *El País (2021),*

Otro aspecto fundamental que tiene gran impacto el de la transparencia. Y es que las entidades que implementan IA han de ser capaces de explicar cómo funcionan sus sistemas y cómo se toman las decisiones y, además los datos utilizados han ser accesibles para todas las partes involucradas en el proceso. La falta de transparencia en los algoritmos puede socavar el derecho de defensa y el derecho a la tutela judicial efectiva.

Existe el temor de que un uso excesivo de IA acabe deshumanizando el sistema judicial y desplazando el papel central del juez. Por ello, a pesar de los avances digitales se subraya la importancia de la supervisión humana en las decisiones judiciales. Los sistemas de IA deben utilizarse únicamente como herramientas de asistencia a los jueces pero nunca como sustitutos de estos.

Otra de las áreas con mayor impacto ha sido la evaluación de riesgo de reincidencia y la que ha pasado a ser un elemento importante a la hora de tomar decisiones, tales como la libertad condicional, el tipo y duración de las penas o el nivel de riesgo de víctimas de violencia de género, donde los algoritmos pueden

procesar datos históricos sobre antecedentes penales y características sociodemográficas de las personas procesadas, identificando patrones que acaban por predecir la probabilidad de reincidencia, no obstante, este tipo de determinación está sujeto a sesgos existentes ya que se nutre de datos históricos, convirtiéndose en un mecanismo que puede conducir a decisiones discriminatorias. Castells (2019) alerta de que la implementación de los algoritmos en la Administración de Justicia puede acabar por perpetuar prejuicios existentes -si no están bien diseñados y supervisados- con el riesgo de incriminar de forma estigmatizada a las comunidades más vulnerables.

Un aspecto positivo es la automatización en el análisis de grandes volúmenes de información jurídica, donde la IA está aportando una notable eficiencia. A través del procesamiento automatizado de documentos y la identificación de precedentes legales, los sistemas de IA pueden agilizar la labor de abogados y jueces, reduciendo los tiempos de revisión y facilitando el acceso a datos relevantes. Sin embargo, De Castro (2018) señala que la dependencia excesiva en estos sistemas automatizados también puede crear una brecha en la comprensión humana del caso, especialmente si los operadores de justicia no tienen un conocimiento detallado del funcionamiento de los algoritmos que utilizan. Esto podría derivar en una *opacidad tecnológica* donde las decisiones judiciales se basen en herramientas cuyo funcionamiento no es del todo claro para quienes las aplican.

5.2. Sobre la efectividad del algoritmo COMPAS en las prisiones de EEUU.

Para comprender cómo esas cuestiones propias de la ética se incorporan en el día a día, se hace oportuno fijar la atención en algunos casos donde los propios algoritmos son protagonistas en el ámbito judicial o en el campo de la seguridad.

Según un artículo de *ProPublica* (2016), el análisis del algoritmo COMPAS revela importantes preocupaciones sobre su precisión y sesgo en la predicción de la reincidencia. El sistema COMPAS (*Correctional Offender Management Profiling*

for Alternative Santcions), que en español puede traducirse como Administración de Perfiles de Criminales para Sanciones Alternativas del Sistema de Prisiones de EEUU, un sistema que respaldo a la decisión judicial que se ha convertido en uno de los algoritmos de evaluación de riesgos más utilizados en la predicción del delito en este país. Pese a ello, el sistema COMPAS y su uso en el ámbito judicial ha sido objeto de duras críticas por sus presuntos sesgos raciales, dado que teniendo en cuenta algunos estudios independientes. COMPAS tiende a clasificar a las personas negras como de mayor riesgo frente a otro tipo de personas, incluso en casos y circunstancias similares. Este tipo de casos revelan que no basta con implementar sistemas de IA en el sistema judicial, sino que es necesario realiza auditorias que evalúen sus efectos sobre las decisiones finales.

Además, las investigaciones apuntaron a que COMPAS podría asignar puntuaciones desproporcionadamente altas determinadas minorías étnicas, supone un debilidad importante sistema penal, en un momento en que las tensiones raciales y el trato desigual por parte de la policía en EEUU están tan presentes.

La investigación de *ProPublica* (2016) analizó la puntuación de riesgo de 7000 personas detenidas en Florida durante dos años, y los resultados fueron sorprendentes:

A menudo se predijo que los acusados de raza negra tenían un riesgo mayor de reincidencia de lo que en realidad tenían. El análisis descubrió que los acusados de raza negra que no reincidieron en un período de dos años tenían casi el doble de probabilidades de ser clasificados erróneamente como de alto riesgo en comparación con sus contrapartes blancas (45 por ciento frente a 23 por ciento).

Se predijo que los acusados blancos representaban un menor riesgo del que en realidad representaban. El análisis descubrió que los acusados blancos que reincidieron en los dos años siguientes fueron etiquetados erróneamente

como de bajo riesgo casi el doble de veces que los reincidentes negros (48 por ciento frente a 28 por ciento).

El análisis también mostró que incluso cuando se controlan los delitos anteriores, la reincidencia futura, la edad y el género, los acusados negros tenían un 45 por ciento más de probabilidades de que se les asignaran puntuaciones de riesgo más altas que los acusados blancos.

Los acusados de raza negra también tenían el doble de probabilidades que los acusados de raza blanca de ser clasificados erróneamente como de mayor riesgo de reincidencia violenta. Y los reincidentes violentos blancos tenían un 63 por ciento más de probabilidades de haber sido clasificados erróneamente como de bajo riesgo de reincidencia violenta, en comparación con los reincidentes violentos negros.

El análisis de reincidencia violenta también mostró que incluso cuando se controlan los delitos anteriores, la reincidencia futura, la edad y el género, los acusados negros tenían un 77 por ciento más de probabilidades de que se les asignaran puntuaciones de riesgo más altas que los acusados blancos.

Una de las preguntas que surge es cómo este algoritmo determina el sesgo racial, si no tiene en cuenta el origen étnico o el color de la piel. La respuesta parece estar en que el algoritmo incluye preguntas que sí pueden reflejar indirectamente estos factores étnicos, y que pueden estar más relacionadas con la experiencia de las minorías que han tenido mayor intervención policial.

Los hallazgos de ProPublica generaron un intenso debate en los Estados Unidos sobre el uso de estas calificaciones de riesgo en el sistema judicial. Aunque los resultados de este sistema fueron criticados, los diseñadores de COMPAS no aceptaron las conclusiones de la investigación de ProPublica, defendiendo que su algoritmo no incurría en sesgos raciales.

5.3. Sobre la fiabilidad de las predicciones algorítmicas

La fiabilidad de las predicciones realizadas mediante los algoritmos de IA es un fundamental para evaluar su efectividad, pero esta depende de varios factores que deben ser considerados.

En primer lugar, la calidad de los datos utilizados para entrenar los modelos algorítmicos: si los datos son incompletos, sesgados o no representan adecuadamente a la población o la situación en cuestión, los resultados obtenidos pueden ser erróneos o incluso perjudiciales. Esto es particularmente relevante en ámbitos como la salud, donde modelos entrenados con datos homogéneos pueden no ofrecer soluciones aplicables a toda la diversidad de la población.

Además la explicabilidad de un sistema, como ya se ha dicho, resulta imprescindible para saber cómo y por qué un modelo llega a determinadas conclusiones y la falta de esta puede generar incertidumbre y desconfianza entre los usuarios.

Otro aspecto importante es la necesidad de evaluar y validar de forma continua los modelos de IA. Dado a que las condiciones y los contextos pueden cambiar con el tiempo, es necesario realizar pruebas rigurosas y ajustes periódicos para mantener la precisión y relevancia de las predicciones. Una evaluación adecuada, respaldada por métodos estadísticos sólidos, garantiza que los modelos permanezcan fiables a lo largo del tiempo.

Aunque la inteligencia artificial es una herramienta poderosa para la predicción y la toma de decisiones, su fiabilidad depende de una cuidadosa gestión de los datos, de la transparencia de sus sistemas y de una constante evaluación.

5.4. Sobre la deshumanización en la aplicación de la Justicia: el efecto VAR.

El avance de la inteligencia artificial en el ámbito jurídico plantea un dilema crucial: la tensión entre la objetividad algorítmica y la necesaria interpretación humana en la administración de justicia.

Una analogía esclarecedora es el denominado "efecto VAR", inspirado en la tecnología utilizada en el arbitraje deportivo. En el fútbol, el VAR asiste a los árbitros en la toma de decisiones mediante la revisión de imágenes y el análisis de jugadas con criterios reglamentarios estrictos. Sin embargo, su aplicación ha generado una gran controversia al sustituir la percepción y el criterio del árbitro humano por un sistema exclusivamente técnico, lo que en ocasiones ha desembocado en decisiones que, aunque objetivamente correctas, resultan contrarias al espíritu del juego.

Este fenómeno tiene un claro paralelismo con la aplicación de la IA en la justicia. La promesa de un sistema más eficiente, rápido y libre de sesgos humanos ha llevado al desarrollo de sistemas automatizados para la evaluación de riesgos, la predicción de reincidencia criminal y la toma de decisiones judiciales. No obstante, esta automatización del juicio plantea otra problemática: la deshumanización del proceso. Un algoritmo, por avanzado que sea, no posee la capacidad de comprender el contexto social, la complejidad emocional o las circunstancias individuales de cada caso. Su lógica binaria no admite matices ni excepciones, elementos indispensables para una justicia verdaderamente equitativa.

Además, la confianza ciega en estas herramientas tecnológicas puede generar una falsa percepción de infalibilidad, relegando la intervención humana a un segundo plano y transformando el ejercicio judicial en una mera validación de decisiones algorítmicas. Esto no solo erosiona el principio de imparcialidad, sino que también amplifica los riesgos de sesgos sistemáticos, dado que los modelos de IA se entrenan con datos históricos, y nunca con actuales o futuros, pueden perpetuar desigualdades preexistentes.

Si bien la tecnología puede y debe ser un aliado en la aplicación de la justicia, su implementación no puede desligarse de una supervisión ética y humana. Al igual que en el fútbol, donde el VAR debe ser una herramienta

complementaria y no un sustituto del árbitro, en la justicia, la IA debe actuar como un apoyo al criterio jurídico, nunca como un juez autónomo.

La verdadera justicia no ha de buscar la aplicación literal de la norma, sino que también ha de tener en cuenta la equidad, la empatía y la capacidad de interpretar cada caso concreto. Por tanto reservar la humanidad en la justicia es, al fin y al cabo, el mayor reto que tiene la era de la IA.

6. Narrativas de la IA: integrados, apocalípticos y alternativos.

Aunque es prácticamente imposible prever con certeza las consecuencias de la IA en las próximas décadas, podemos comenzar a imaginar cómo podría evolucionar desde el presente y reflexionar sobre su futuro. Siguiendo las ideas del semiólogo Umberto Eco, podemos dividir las posturas éticas respecto a los algoritmos en tres grandes grupos: los utópicos o 'integrados' los 'apocalípticos' o distópicos, y los 'alternativos'.

6.1. Escenario utópico: la IA como utopía salvadora.

Desde una perspectiva tecno-optimista o utópica, los 'integrados' defienden que la inteligencia artificial marca el inicio de una era de prosperidad ilimitada, actuando como un acelerador del progreso humano. Según esta visión, su desarrollo representa la oportunidad de contar con una herramienta sumamente poderosa para analizar grandes volúmenes de datos, lo que posibilita avanzar en direcciones prometedoras, como el descubrimiento de nuevos tratamientos que podrían curar enfermedades como el cáncer o enfrentar fenómenos como el cambio climático. Además, se abre la puerta a una mayor participación ciudadana y a la liberación de tareas monótonas, permitiéndonos disfrutar de más tiempo libre o encontrar un equilibrio ideal entre trabajo y ocio, fomentando la sociabilidad. En su momento, durante la Revolución Industrial, una parte de la población también dudó de los beneficios de la nueva economía, las vacunas o la electrificación urbana, sin prever que serían ideas revolucionarias que

transformarían el mundo para bien. Desde esta perspectiva, se busca acelerar dicho proceso de transformación.

Con esta visión se llega al concepto de *cyborg*: un ente en estado de simbiosis entre el humano y la máquina que promete una mejora sin precedentes del hombre como especie.

6.2. Escenarios apocalípticos: la IA como distopía catastrofista.

El cine, por su parte, ha presentado dos visiones contrapuestas sobre las máquinas inteligentes. En una de ellas, de carácter distópico, se plantea una IA fuerte capaz de desencadenar una singularidad tecnológica peligrosa que podría resultar en el sometimiento y control de la humanidad, o incluso en su exterminio. En contraste, la visión utópica sugiere que las máquinas son controlables, y que la humanidad podrá desconectarlas si fuera necesario.

Desde una postura tecno-pesimista, los 'apocalípticos' advierten sobre las terribles consecuencias del desarrollo de la IA avanzada, proyectando un futuro distópico que amenaza múltiples aspectos de la vida humana, como los ámbitos antropológico, social, político, económico, legal y ético. Entre los temores que expresan se encuentra el potencial desempleo masivo, que podría afectar al 60% de los trabajadores en las economías avanzadas, incluyendo, por primera vez, la eliminación de empleos altamente cualificados. En un nivel más extremo, la IA es vista como una amenaza para la propia existencia humana, no solo por la posibilidad de una esclavización tecnológica, sino también por el riesgo de extinción de nuestra especie en un escenario post-humano donde las máquinas tomen decisiones autónomas. Este punto crítico se conoce como la 'singularidad tecnológica', momento en el que la creación supera al creador, permitiendo que la máquina desarrolle conciencia propia en forma de superinteligencia, tal como se ve en películas como *Terminator* o *Matrix*. Desde esta perspectiva, la solución ideal sería frenar de golpe el avance, apagando los motores hasta que sepamos cómo manejar de forma segura una herramienta que, de no controlarse, podría causar un daño considerable.

Aunque la mayoría de los expertos se muestran escépticos respecto a la posibilidad de que se materialice este escenario distópico en el corto plazo, reconocen con cautela que los problemas éticos derivados del desarrollo de la IA ya son una realidad palpable y exigen atención urgente.

6.3. Escenario alternativo: la IA como proyecto regulado.

Los 'alternativos', especialmente influyentes en la Unión Europea, proponen una tercera vía que busca equilibrar ambos extremos. Abogan por un uso de la inteligencia artificial abierto, sin abusos, con un enfoque ético, democrático, sostenible y que respete los derechos de los ciudadanos. Además, defienden la preservación de la veracidad informativa y la protección de los puestos de trabajo, reconociendo la importancia de dar pasos firmes y responsables desde el principio para evitar los problemas derivados de la implementación de esta innovadora tecnología generativa. Desde esta perspectiva, sugieren frenar el ritmo de los avances mediante una moratoria reflexiva, con el fin de evitar daños a las personas. A nivel europeo, se están tomando diversas iniciativas jurídicas, como las impulsadas por el Tribunal Europeo de Derechos Humanos y la Agencia de los Derechos Fundamentales de la Unión Europea (2020), la publicación del Libro Blanco sobre la IA (2020), y la entrada en vigor del Reglamento General de Protección de Datos (RGPD, 2021), promovido por la Comisión Europea. A nivel nacional, se han implementado medidas complementarias, como la Estrategia Nacional de IA (ENIA, 2020) y la Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales, entre otras.

En la esfera internacional, durante el mes de febrero de 2025 ha tenido lugar en París la Cumbre de Acción sobre Inteligencia Artificial. Este ha sido un encuentro internacional de gran relevancia en el que se han reunido líderes políticos, empresariales, académicos y representantes de la sociedad civil para abordar el futuro de la inteligencia artificial desde perspectivas éticas, regulatorias y tecnológicas.

Organizada por el gobierno francés en colaboración con la Comisión Europea y organismos internacionales, la cumbre se ha basado en cinco ejes

temáticos: la regulación y gobernanza de la IA, el impulso a la financiación y desarrollo tecnológico, la seguridad y gestión de riesgos, el impacto social y laboral, y la ética y protección de los derechos humanos.

Durante el evento se han presentado varias propuestas clave, como la instauración de un marco normativo internacional que incluye la creación de una agencia europea de supervisión de la IA, y se han anunciado inversiones de hasta 200.000 millones de euros destinadas a fortalecer la competitividad y la innovación en el sector. Asimismo, se ha hecho hincapié en la importancia de establecer mecanismos de verificación y auditación para los modelos de IA de alto riesgo, promover la formación en esta área y garantizar la transparencia de los algoritmos para evitar sesgos discriminatorios.

El cierre de la cumbre se ha escenificado con la firma de la "Declaración de París sobre Inteligencia Artificial", documento que refleja el compromiso colectivo por un desarrollo seguro, inclusivo y transparente de la IA. Esta declaración muestra una clara divergencia en las estrategias regulatorias internacionales al no ser suscrita por Estados Unidos ni Reino Unido.

Este acontecimiento marca un hito en cuanto la regulación mundial sobre la inteligencia artificial, y que sienta las bases para futuras negociaciones y regulaciones.

7. Sondeo de percepciones subjetivas sobre los efectos de la IA.

Para completar la parte teórica y comprender mejor la percepción social sobre la IA se han realizado una serie de entrevistas en profundidad (incluidas en el anexo) a varios profesionales de diferentes sectores, especialmente del sector secundario y terciario. Con ello no se busca una muestra representativa, sino más bien aproximativa de la percepción actual del ciudadano de a pie sobre la IA. El objetivo principal es captar una variedad de puntos de vista que muestren cómo distintas personas entienden y se relacionan con la IA.

7.1. Elaboración y diseño del sondeo.

El estudio se ha estructurado una serie de trece preguntas divididas en dos tipos: estructuradas, con respuestas escaladas (Escala de Lickert) para un análisis de tipo cuantitativo, y unas preguntas con respuestas abiertas, para valorar cualitativamente las opiniones e inquietudes de los entrevistados.

La elección de los participantes para este pequeño sondeo ha sido realizada con el objetivo de obtener una muestra diversa y equilibrada de opiniones sobre la Inteligencia Artificial tratando de evitar cualquier tipo de sesgo. Con este fin, se ha seleccionado a personas de diferentes edades, géneros, niveles educativos y sectores profesionales, para reflejar una variedad de perspectivas que puedan enriquecer el análisis. La muestra no tiene la intención de ser representativa, sino que busca ofrecer una aproximación cualitativa sobre las percepciones de la sociedad frente a la IA.

Los participantes incluyen entre otros un ingeniero informático; jefe del departamento de informática de un gran ayuntamiento, un ingeniero industrial; director técnico, una trabajadora social, administrativas del servicio público de sanidad o agentes de la Policía Local de la Comunidad Valenciana.

7.2. Resultados del sondeo.

A partir de las entrevistas, se han identificado una serie de percepciones sobre la IA, tanto en el ámbito laboral como en la vida cotidiana. Para organizar y entender mejor estos hallazgos, se ha realizado un análisis DAFO (Debilidades, Amenazas, Fortalezas y Oportunidades), la cual permite examinar de manera integral los factores que influyen en la implementación de la inteligencia artificial.

Este análisis tiene como objetivo resaltar los aspectos positivos y negativos identificados en la percepción de los entrevistados, con el fin de proponer líneas de acción para aprovechar las oportunidades y mitigar las amenazas.

A continuación, se presentan las principales categorías del análisis DAFO, basadas en los resultados obtenidos a través de las entrevistas realizadas.

En cuanto a las fortalezas los entrevistados señalan lo que perciben positivamente sobre la inteligencia artificial y su impacto en distintos sectores. Entre las principales fortalezas identificadas se encuentran:

- **Eficiencia y reducción de costes**: Muchos de los entrevistados coinciden en que la IA puede ser clave para mejorar la eficiencia en los procesos administrativos, la toma de decisiones basadas en datos y la gestión de la información. En particular, se destaca la capacidad de la IA para automatizar tareas repetitivas y agilizar trámites.
- **Mejora de la competitividad empresarial**: Los profesionales del sector empresarial destacan cómo la IA facilita la mejora de la competitividad a través de la automatización y la optimización de recursos. Esto se percibe como una ventaja fundamental para reducir costes y mejorar el rendimiento.
- **Facilitación de la toma de decisiones**: La IA es vista como una herramienta potente para tomar decisiones informadas gracias a la gran cantidad de datos que puede analizar y procesar, lo que permite a los profesionales tomar decisiones basadas en hechos y patrones detectados.

Las debilidades reflejan los aspectos negativos que los entrevistados perciben sobre la adopción de la inteligencia artificial, los cuales podrían frenar su implementación o generar resistencia en ciertos sectores. Entre las principales debilidades destacadas se encuentran:

- **Desinformación y falta de conocimiento**: La mayoría de los entrevistados expresan que, aunque reconocen el potencial de la IA, hay una falta de comprensión sobre qué es realmente y cómo va a influir a largo plazo en la sociedad. Esto refleja una necesidad urgente de más información sobre este tema, la cual debe ser clara y accesible.
- **Resistencia al cambio en ciertos sectores**: En algunos sectores, como el psicosocial y de seguridad pública, se perciben limitaciones en la implementación de IA debido a la importancia del contacto humano y

la necesidad de tomar decisiones con sensibilidad, lo que hace que la tecnología sea vista con recelo.

Las oportunidades se refieren a los aspectos externos que podrían aprovecharse para maximizar el uso de la IA en la sociedad. Entre las principales oportunidades identificadas se encuentran:

- **Educación y formación**: Existe una gran oportunidad para diseñar campañas educativas que informen a la población sobre los beneficios y riesgos de la IA. Los entrevistados sugieren que estas campañas deben ser accesibles, con contenido claro que permita una comprensión general.
- **Expansión a otros sectores**: La IA tiene un gran potencial para expandirse en sectores como la salud, la educación y la seguridad pública. A medida que se vaya desarrollando la tecnología, estos sectores podrán aprovechar sus ventajas para mejorar la eficiencia y la calidad de los servicios.

Las amenazas son factores externos que podrían suponer un riesgo para el avance o la correcta implementación de la IA. Los entrevistados señalan varias amenazas importantes que deben ser gestionadas adecuadamente:

- **Ciberdelincuencia y mal uso**: Una de las preocupaciones más comunes es el uso malintencionado de la IA, especialmente en lo que respecta a la ciberdelincuencia, suplantación de identidades y estafas. Los entrevistados destacan la necesidad de regulaciones estrictas para mitigar estos riesgos.
- **Desigualdad en el acceso a la tecnología**: La brecha digital puede ser una amenaza importante, ya que el acceso desigual a la IA puede incrementar la desigualdad social, dejando a algunas poblaciones en desventaja frente a otros sectores que cuentan con mayores recursos.
- **Impacto social y laboral**: Muchos entrevistados temen que la IA pueda llevar a la deshumanización de ciertos trabajos, así como a la pérdida de empleo, especialmente entre los sectores más vulnerables, como aquellos en puestos precarios.

7.3. Conclusión del análisis DAFO

El análisis DAFO realizado a partir de las entrevistas proporciona una visión clara sobre las percepciones y actitudes hacia la inteligencia artificial. Si bien las fortalezas y oportunidades destacan el potencial de la IA para mejorar la eficiencia y la competitividad en diversos sectores, las debilidades y amenazas revelan cuales son las preocupaciones de la población sobre su implementación, especialmente en relación con la falta de regulación y los riesgos sociales asociados. Es fundamental que se tomen medidas para abordar estas cuestiones, fomentando una educación digital adecuada, un marco regulatorio robusto y garantizando un acceso equitativo a la tecnología.

Fuente: elaboración propia.

8. A modo de conclusión.

A lo largo de este trabajo hemos analizado las consecuencias de la inteligencia artificial en nuestra sociedad, un campo que, si bien aporta notables avances, plantea profundos interrogantes éticos y humanos. La automatización de procesos y la creciente influencia de la IA en la toma de decisiones nos enfrentan a un desafío sin precedentes: equilibrar los beneficios que nos ofrece con la protección de los derechos fundamentales.

Es evidente que la IA tiene la capacidad de mejorar nuestras vidas mejorando la eficiencia, la innovación y la optimización de recursos. Sin embargo, hemos constatado que su implementación sin una regulación adecuada puede profundizar las desigualdades sociales y vulnerar principios democráticos esenciales. Los sistemas automatizados pueden, de forma inadvertida, reproducir sesgos preexistentes, afectando de manera desigual a determinados colectivos. Este riesgo requiere una reflexión ética profunda que nos lleve a revisar los marcos normativos que guían el desarrollo y la implementación de estas tecnologías.

En primer lugar, cabe replantearse el concepto de justicia dentro de la era digital: la IA al automatizar procesos de toma de decisiones que antes estaban exclusivamente reservados al ser humano, nos confronta con la necesidad de repensar la justicia en sus cimientos. La creciente utilización de algoritmos e IA en áreas tan críticas como la justicia o la seguridad pública, entre otras, introduce un riesgo inherente: la discriminación algorítmica. Si estos sistemas no son suficientemente justos y libres de sesgos, corremos el peligro de perpetuar y amplificar desigualdades sociales existentes, incluso se corre el riesgo de generar nuevas formas de discriminación automatizada, erosionando los principios fundamentales de toda sociedad democrática.

En este sentido, podemos apelar a una ética de la IA creativa a la vez que adaptativa. La complejidad que caracteriza a la inteligencia artificial demanda la aplicación de una ética que trascienda la simple aplicación de reglas predefinidas.

Esta ética no puede ser estática ni reduccionista; ha de ser creativa y en constante evolución. Ante las situaciones inéditas y los nuevos dilemas morales que plantea la IA, es imprescindible desarrollar un marco ético que sea capaz de adaptarse a la rápida transformación tecnológica. Este marco debe tener como pilares la promoción real de la inclusión, la diversidad y la equidad, asegurando que los beneficios de la IA se distribuyan de manera justa y que nadie quede excluido.

En este sentido habría que plantearse la integración de la ética mencionada desde su nacimiento. Hemos de abandonar la concepción errónea de la ética como un mero añadido final al proceso de desarrollo tecnológico. La ética de la IA debe ser un principio rector el cual se aplique tanto al diseño desde su fase inicial como a proceso de implementación y uso final. Esto implica una transformación profunda en la mentalidad de los desarrolladores de IA. Desde el primer momento, deberían ser conscientes de las implicaciones éticas de su trabajo, teniendo en cuenta cómo sus creaciones afectarán a la sociedad. La responsabilidad ética no debe ser una etapa más al final del proceso, sino los cimientos sobre los que se construye todo sistema de IA.

En última instancia, y a la vista de la rápida evolución de la IA, sería conveniente la creación de nueva generación de especialistas en ética de la IA. Para responder a las nuevas cuestiones éticas que vayan surgiendo medida que avanza tecnológicamente la inteligencia artificial. Es importante contar con profesionales capacitados que además de comprender su funcionamiento técnico, también posean conocimientos éticos, filosóficos y sociológicos, además de un profundo respeto hacia los valores humanos. Solo así se podrá establecer una colaboración efectiva entre ingenieros, científicos y empresas tecnológicas. Su papel será fundamental para el desarrollo y la implementación de una IA que sea eficiente y socialmente responsable.

En definitiva, la inteligencia artificial representa la mayor revolución tecnológica de nuestra era, con un potencial transformador que abarca todos los aspectos de la sociedad. No obstante, su progreso no puede desligarse de una

reflexión constante sobre sus implicaciones éticas y morales. Solo desde un punto de vista responsable, transparente y humanista podremos garantizar que el avance de la inteligencia artificial verdaderamente contribuya al bienestar colectivo sin menoscabar los derechos y valores fundamentales de la humanidad.

9. Web-bibliografía.

- Agencia de los Derechos Fundamentales de la Unión Europea. (2021). *Construir correctamente el futuro: La IA y los derechos fundamentales*. https://doi.org/10.2811/818206
- Alonso Betanzos, A., & Bolón Canedo, V. (2020). *IA, algoritmos y derecho: Una introducción*. UOC. https://openaccess.uoc.edu/bitstream/10609/150223/4/InteligenciaArtificialAlgoritmosYDerechoUnIAntroduccion.pdf
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, mayo 23). How we analyzed the COMPAS recidivism algorithm. *ProPublica*. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm
- Castells, M. (2019). *Vigilancia y control social en la era digital: Ética y derecho en la sociedad del riesgo*. Anagrama.
- De Castro, J. C. (2018). *Tecnología y derecho: Nuevas formas de control social*. Comares.
- El País. (2021, julio 11). *Riscanvi: luces y sombras del algoritmo que ayuda al juez en Cataluña a decidir si mereces la condicional*. El País. https://elpais.com/tecnologia/2021-07-11/riscanvi-luces-y-sombras-del-algoritmo-que-ayuda-al-juez-en-cataluna-a-decidir-si-mereces-la-condicional.html
- Europa Press. (2017, noviembre 4). *El proyecto de inteligencia artificial que ayuda ya a esclarecer la autoría de los incendios*. Europa Press. https://www.europapress.es/galicia/agro-00246/noticia-proyecto-inteligencia-artificial-ayuda-ya-esclarecer-autoria-incendios-20171104110452.html
- Gil, E. (2021). *Algoritmos y derecho: El impacto de las tecnologías en la Administración de Justicia*. Dykinson.
- Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. https://www.boe.es/eli/es/lo/2018/12/05/3/con
- Ley Orgánica 7/2021, de 26 de mayo, de protección de datos personales tratados para fines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales. https://www.boe.es/eli/es/lo/2021/05/26/7/con
- López, F. (2022). *Derecho y tecnología: La regulación de la IA en el ámbito jurídico*. Tirant lo Blanch.
- Miranda Bonilla, H. (2021). Algoritmos y derechos humanos. *Revista De La Facultad De Derecho De México, 71*(280-2), 705–732. https://doi.org/10.22201/fder.24488933e.2021.280-2.79666
- Pérez, L. (Director). (2024). *Justicia artificial* [Película]. CineTech Studios.
- Propuesta de Directiva sobre la Responsabilidad Civil de la IA. (2022, septiembre). Estado: En proceso de discusión y desarrollo.

- Reglamentos sobre la IA (AI Act). https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence
- Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas sobre la IA. http://data.europa.eu/eli/reg/2024/1689/oj
- Real Decreto 817/2023, de 8 de noviembre, que establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo. https://www.boe.es/eli/es/rd/2023/11/08/817
- Sánchez Almeida, C. (2020). *Privacidad y protección de datos en la era digital*. Reus.
- Zuboff, S. (2020). *La era de la vigilancia: El capitalismo de datos y la crisis de la democracia*. Taurus.
- Las Provincias. (2024, diciembre 9). *El Ayuntamiento impulsa nuevas cámaras para controlar las zonas de carga y descarga*. Las Provincias. https://www.lasprovincias.es/valencia-ciudad/ayuntamiento-impulsa-nuevas-camaras-controlar-zonas-carga-20241209110549-nt.html
- Estrategia Nacional de IA (ENIA). https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/ENIA2B.pdf
- Estrategia Europea de IA. (2018). https://digital-strategy.ec.europa.eu/en/policies/european-strategy-artificial-intelligence
- Libro Blanco sobre IA. (2020). https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf

10. anexo

EDAD	GÉNERO	ESTUDIOS	PROFESIÓN	RESIDENCIA	PROCEDENCIA

La plantilla del cuestionario se estructura en trece preguntas, distribuidas en tres baterías o apartados:

4 preguntas 'fáciles o sugerentes' a modo de enganche para ganar el interés o la confianza del encuestado.	1	¿Consideras que los ciudadanos están suficientemente informados sobre cómo la IA puede afectar al conjunto de sus vidas en general?. ¿Conoces la legislación de la UE en materia de regulación IA?. ¿Qué opinión te merece?.
	2	Qué papel cree que juega la IA hoy en la vida cotidiana de tu sector profesional. Exprésalo en términos de porcentaje aproximado (%). 0-20% 20-40% 40-60% 60-80% 80-100% *(Discriminar en sectores primario, secundario, terciario)*
	3	¿Estimas que el uso de la IA puede aumentar la eficiencia y la competitividad en tu sector profesional? Nada (0), Poco (1), Algo (2), Bastante (3), Mucho (4), Todo (5) Por qué.
	4	¿Supones que los algoritmos de la IA generativa pueden reducir el grado de subjetividad y falibilidad en la toma de decisiones para introducir una mayor objetividad y fiabilidad en los resultados? Nada (0), Poco (1), Algo (2), Bastante (3), Mucho (4), Todo (5) Por qué.

4 preguntas diana que vayan directamente al objetivo de la investigación.	5	¿Podría citar tres de los beneficios que introduce esta tecnología innovadora en su ámbito profesional?.
	6	¿Presumes que a corto o medio plazo la IA podría llegar a sustituir con buen criterio las decisiones humanas que antes se tomaban de manera deliberativa?
	7	Y por el contrario, qué tres problemas detectas que podrían afectar la introducción de la IA en el ámbito de tu actividad profesional.
	8	A quién harías responsable en caso de que un algoritmo tomara una decisión que afectara negativamente a una persona o a un grupo humano. a. Al programador que lo diseñó. b. A la empresa que lo comercializa. c. A la falta de regulación de la Administración pública. d. A quiénes otros.

5 preguntas comprometidas que supongan una toma de posición del encuestado.	9	¿Cree que los desarrolladores de IA tienen en cuenta los aspectos éticos en el diseño y funcionamiento de los algoritmos que se aplican para la realización de cualquier actividad?
	10	¿Hasta qué punto crees que la aplicación indiscriminada de los algoritmos vulnera los derechos ciudadanos en ámbitos públicos tales como la sanidad, la seguridad o la administración?. Por qué.
	11	¿Concibes que el uso de la IA pueda llegar a la discriminación de ciertos grupos de personas al introducir sesgos de género, étnicos, de clase social u otros?. Por qué. ¿Podrías poner algún ejemplo?.
	12	¿Piensa que en el futuro la IA podría suponer una amenaza real para el ser humano? Nada (0), Poco (1), Algo (2), Bastante (3), Mucho (4), Todo (5)

		Por qué
	13	Qué medidas concretas implementarías a tu juicio para garantizar un uso responsable y ético de la IA