



Universidad Internacional de La Rioja
Facultad de Derecho

Máster Universitario en Investigación Criminal
**Sesgos de la Inteligencia Artificial en la
investigación criminal**

Trabajo fin de estudio presentado por:	Nicasio Armenteros Generoso
Tipo de trabajo:	Trabajo Fin de Master
Tema:	Inteligencia Artificial e investigación Criminal
Director/a:	Sergio Colado García
Fecha:	07-02-2023

Resumen

No hay duda que la inteligencia artificial supone todo un reto para la sociedad actual. Los sistemas de inteligencia artificial traen consigo grandes ventajas, pero también problemas, a veces inesperados, que comprometen en algunos campos de aplicación los derechos fundamentales de las personas. La existencia de errores y sesgos en los algoritmos, que son consustanciales a la inteligencia artificial, son parte importante del problema y además, no pueden evitarse pero si mitigarse. Es en la esfera judicial y policial, en su aspecto de investigación criminal, donde quizás, el peligro de la conculcación de los derechos de las personas sea más patente y preocupante. Es por ello necesario realizar un sereno análisis de la cuestión y no dejarse llevar por una visión fatalista, intentando ser conscientes de las grandes ventajas que puede aportar estas nuevas herramientas pero sin perder de vista sus posibles consecuencias si no se presta el cuidado necesario.

Palabras clave: Inteligencia artificial (IA), sesgo, algoritmo, policía predictiva.

Abstract

There is no doubt that artificial intelligence poses quite a challenge for today's society. AI systems bring with them significant advantages, but also problems, sometimes unexpected ones, which compromise fundamental rights of individuals in certain fields of application. The existence of errors and biases in algorithms, inherent to artificial intelligence, is a significant part of the problem and moreover, cannot be avoided but can be mitigated. It is perhaps in the judicial and law enforcement sphere, particularly in criminal investigation, where the danger of violating people's rights is most evident and concerning. It is therefore necessary to conduct a calm analysis of the issue and not be swayed by a fatalistic view, trying to be aware of the great advantages that these new tools can bring but without losing sight of their possible consequences if the necessary care is not taken.

Keywords: Artificial Intelligence (AI), bias, algorithm, predictive policing.

Índice de contenidos

1. Introducción	8
1.1. Justificación del tema elegido.....	9
1.2. Problema y finalidad del trabajo.....	10
1.3. Objetivos	10
2. Marco teórico y desarrollo	12
2.1. Inteligencia Artificial	12
2.1.1. Breve reseña histórica	14
2.1.1.1. Antes del siglo XX	14
2.1.1.2. Desarrollo de la IA en el siglo XX.....	14
2.1.1.3. Siglo XXI. El presente.....	15
2.1.2. Big Data.....	15
2.1.3. Definición de Inteligencia Artificial.....	17
2.1.4. Tipos de Inteligencia Artificial	19
2.1.5. Aprendizaje automático « <i>Machine learning</i> »	22
2.2. Algoritmos.....	24
2.2.1. Definición de Algoritmo.....	24
2.2.2. Algoritmo e Inteligencia Artificial	24
2.2.3. Algoritmos como cajas negras y posibles fuentes de sesgos y discriminación..	25
2.3. Sesgos.....	29
2.3.1. Definición de sesgo.....	30
2.3.2. Tipos de sesgos	31
2.3.3. Solución a los sesgos algorítmicos.....	36
2.4. Sesgos y Errores humanos, íntimamente ligados a la máquina	38
2.5. IA en el ámbito judicial y de la investigación criminal.....	40

2.5.1.	Aplicación a la investigación criminal y al ámbito judicial	41
2.5.2.	Policía predictiva.....	46
2.5.3.	Problemática de la aplicación de la IA al ámbito de la investigación criminal...48	
2.5.3.1.	Responsabilidad de los errores de la inteligencia artificial.....	52
2.5.3.2.	Protección de datos	52
2.6.	Implicaciones ético-morales	54
2.7.	Presente y futuro del marco legal.....	58
2.7.1.	Presente.....	58
2.7.2.	Futuro	60
3.	Resultados	63
3.1.	Primera hipótesis. El mito de la malignidad de la IA	63
3.2.	Segunda hipótesis. Falta de transparencia como como principal problema de la IA64	
4.	Conclusiones.....	65
4.1.	Primera hipótesis	65
4.2.	Segunda hipótesis	66
4.3.	Algunas consideraciones finales	66
4.4.	Dificultades encontradas	67
4.5.	Propuestas para líneas futuras de investigación	67
	Referencias bibliográficas.....	68
	Listado de abreviaturas	85
	Anexo A. Figuras 1 y 2.....	86

Índice de figuras

Figura 1: Inteligencia artificial.	86
Figura 2: Tipos de aprendizaje automático	86

Índice de tablas

Tabla 1. <i>Resumen de las distintas utilidades de la IA y herramientas</i>	46
Tabla 2. <i>Resumen principios éticos aplicables a la IA</i>	57

1. Introducción

Cuando hablamos de Inteligencia Artificial (en adelante IA) seguramente, de forma ineludible, se dibuja en nuestra mente la imagen de un robot de apariencia humanoide, muchas veces recreada a través de novelas o el cine, bien como la descrita en los relatos de «Yo, Robot» (ASIMOV, 1950), donde los robots pueden aprender y tomar decisiones en base a tres leyes de la robótica, bien como las memorables imágenes de «The Terminator» (CAMERÓN, 1984) donde una poderosa máquina de IA ha desarrollado conciencia propia y busca la aniquilación de la humanidad. Sin embargo, aunque todavía estamos lejos de esa realidad imaginada, no podemos obviar lo que representa, hoy en día, el potencial de la IA en diversas y dispares áreas del conocimiento humano, como herramienta de ayuda al hombre en el desarrollo de sus actividades. Efectivamente, desde la robótica hasta la publicidad, pasando por la seguridad o las ciencias sociales, la aplicación de la IA se constituye como una valiosísima herramienta al servicio del hombre, que como indica SORIA OLIVAS (2022) avanza con tal velocidad que hace muy complicado seguir y compilar todos los cambios que se producen, en este campo, en unos pocos años a esta parte.

Desde un punto de vista judicial, la aplicación de la IA plantea nuevos retos derivados de riesgos producidos por el desarrollo de dicha tecnología, tal y como indica NÚÑEZ ZORRILLA (2019). Y para COBO CANO y LLORET IGLESIAS (2023), los avances de la IA, en relación con la medicina, podrían llevarnos a pensar que no existen limitaciones para ella.

Una de las áreas de conocimiento donde la IA puede ayudar, lo representa la criminología. En concreto, puede ser útil en la prevención y predicción del delito, facilitando la tarea de interpretación de resultados y toma de decisiones a partir de las características de un delito, un delincuente o una víctima. Sin embargo, a pesar de su potencial y su prometedor futuro se plantean diversas dudas como: ¿Es ético o moralmente aceptable que participe una máquina en decisiones que pueda afectar a la vida y derechos fundamentales de una persona?; ¿Los algoritmos utilizados en la IA en esa toma de decisiones complejas, son seguros, o pueden presentar sesgos que se transmitan a los resultados?; ¿En qué grado esos sesgos son mejores o peores que los que cometerían un humano?; ¿Sería viable la toma de decisiones por IA sin participación humana?; ¿Qué problemas y consecuencias podría plantear?; ¿Tenemos que temer la IA o ese temor es un sentimiento irracional?

1.1. Justificación del tema elegido

En palabras de LÓPEZ DE MÁNTARAS y MESEGUER GONZÁLEZ (2017) sería más correcto hablar de distintas inteligencias en lugar de únicamente de inteligencia, por lo que la investigación en IA podría llevarnos al desarrollo de un nuevo tipo de inteligencia no presente en la naturaleza hasta el momento.

La propia importancia del fenómeno de la IA y su capacidad de influir profundamente sobre la vida de los hombres en nuestra sociedad, nos lleva a pensar que es de sumo interés estudiar este fenómeno desde el mayor número de perspectivas posibles. De este modo, la visión criminológica constituiría un punto de vista más, que nos permitiría conocer y analizar con mayor profundidad las consecuencias de la aplicación de la IA en uno de los campos del conocimiento humano, el campo judicial.

En el campo de la investigación criminal las posibilidades de estudio son enormes, desde investigar cómo la IA puede mejorar la precisión y eficacia del reconocimiento facial en la identificación de sospechosos y víctimas en escenas del crimen, hasta la utilización técnicas de aprendizaje automático para lograr perfilados criminales mucho más precisos, atendiendo a un gran volumen de datos recogidos, pasando por la utilización de algoritmos que permitan la optimización de los medios policiales, mejorando la capacidad predictiva de estos a través del cálculo de zonas en las sea más probable el hecho delictivo.

De todos los aspectos posibles a someter a estudio, la relación de la IA con los algoritmos, que permiten tomar decisiones, y concretamente sus sesgos, constituye uno de los focos principales fuente de polémica, como puede verse a través de artículos como el expuesto por VICENTE y MATUTE (2023), VALENZUELA (2021), MUÑOZ GUTIÉRREZ (2021), FALIERO (2021) o incluso se recoge en la Ley de Inteligencia Artificial europea que a fecha de elaboración de este trabajo está a punto de publicarse.

En el estado actual de la cuestión, donde la aplicación de la IA dentro de nuestra sociedad supone grandes retos y donde abundan gran cantidad de titulares alarmistas sobre los peligros de la IA . Sirva de ejemplo «Gobiernos mundiales alertan sobre las potenciales consecuencias “catastróficas” de la inteligencia artificial (IA)» de la agencia EFE (2023) o «La inteligencia artificial puede llevar a la extinción de la humanidad: la advertencia de los grandes expertos en IA» de BBC NEWS MUNDO (2023). Creemos que es fundamental intentar arrojar algo de

luz sobre dicho tema tratando de desmitificar el supuesto problema, allí donde sea necesario y llamar la atención sobre los puntos donde se plantee, realmente, el peligro anunciado.

1.2. Problema y finalidad del trabajo

A través de los diversos autores que se exponen, como O`NEIL (2016), se argumenta con frecuencia que la IA es un precursor de la discriminación, hacia el hombre. El miedo hacia los sesgos que se desprenden de algunas de las aplicaciones empleadas, hoy en día, sería el principal motivo de dicha conclusión. Sin embargo, es quizás necesaria una visión más calmada del problema, tratando de averiguar cuál es el verdadero origen de esos sesgos y como consecuencia, de la posible discriminación que acarrea.

Tema también de gran importancia es la ética de la aplicación de la IA al ámbito criminal y judicial. Sustituir las decisiones de una persona por decisiones de una máquina supone un reto a la hora de admitir como válidos los resultados que esta arroje. Tampoco, se puede olvidar el peligro potencial de las predicciones de acciones delictivas futuras, que podrían suponer y ya en muchas ocasiones lo supone un adelanto de la barrera punitiva de nuestra sociedad.

Con la elaboración de este texto trataremos de dar respuesta a varias preguntas que son claras en su planteamiento, pero su respuesta es más compleja de lo que en un principio pudiera parecer. Estas preguntas son las siguientes: ¿el daño que producen los sesgos en la IA, invalidan esta como herramienta?, ¿son evitables?, ¿es posible reducir dichos daños?, ¿son peores que los errores humanos cometidos en el mismo ámbito judicial o de investigación criminal?.

1.3. Objetivos

La hipótesis sobre la que trabajaremos plantea que los efectos malignos catastrofistas de la IA, son en general un mito que debe ser combatido, puesto que los beneficios que puede aportar serán, con mucho, superiores a los perjuicios. Aunque, sin perder de vista, los citados efectos adversos. Es decir, plantearíamos una visión menos alarmante respecto del uso la IA dentro del ámbito de la investigación criminal.

De este modo tendríamos como hipótesis principal de trabajo, que a pesar de la indiscutible existencia de sesgos, tanto en la construcción de los algoritmos como dentro del conjunto de datos utilizados para entrenar la IA, es posible la implementación de medidas de control que

mitiguen de forma efectiva sus influencia evitando o reduciendo en última instancia resultados como la discriminación en sus distintas facetas.

Como hipótesis secundarias planteamos que, a pesar de la existencia de sesgos, es quizás, la falta de transparencia y explicabilidad de los algoritmos utilizados en IA, la que puede contribuir de manera significativa a la opacidad de los procesos judiciales penales y a la investigación policial en el ámbito criminal. Planteando además la necesidad de superar retos éticos cruciales sobre la responsabilidad de estas herramientas.

En atención a las hipótesis planteadas los objetivos serían los siguientes:

1. Objetivo general:
 - a. Tratar de desmitificar la visión oscura sobre la IA y aportar más claridad al tema de los sesgos existentes de la IA, planteando la necesidad de una actitud positiva aunque vigilante.
2. Objetivos secundarios:
 - a. Evaluar los distintos tipos de sesgos existentes y que se conoce por sesgo algorítmico.
 - b. Explicar si existen medidas de control y reducción de sesgos algorítmicos.
 - c. Conocer en que campos es útil la IA dentro del ámbito de la investigación criminal.
 - d. Conocer los problemas que plantea y su impacto en la investigación criminal y los procesos judiciales penales.
 - e. Implicaciones éticas de los algoritmos de IA utilizados en el ámbito judicial y policial.
 - f. Elaborar un cuadro de directrices éticas y morales a las que se debería someter la IA dentro del campo de estudio para reducir sus inconvenientes.
 - g. Breve planteamiento de la legislación existente a día de hoy y la nueva por venir.

2. Marco teórico y desarrollo

Con el fin de conocer el objeto de estudio que se presenta en este trabajo, será necesario establecer cuáles son los elementos fundamentales que tendremos que conocer y delimitar, que conforman parte del tema elegido. Por ello, a continuación expondremos esos elementos necesarios que conforman nuestro marco de trabajo definiendo y explicando algunos de los conceptos fundamentales.

2.1. Inteligencia Artificial

Podríamos asegurar sin gran riesgo a errar que, a día de hoy, la IA es uno de los temas de actualidad, que supone un gran reto mundial al que se enfrenta la humanidad. Prueba de ello pueden ser las diversas noticias que genera el hecho en sí y que hablan sobre las posibles consecuencias del desarrollo de dicha tecnología. Mostrando una realidad al más puro estilo Beckeriano, es decir, señalando que la teoría de la sociedad del riesgo de BECK (2002) puede todavía utilizarse como referente a la actualidad del avance en IA.

Baste como demostración de la afirmación expuesta arriba, solo algunos ejemplos sobre artículos, que a continuación se exponen, sin ánimo de ser exhaustivos: KAI-FU (2017), The New York Times, «La verdadera amenaza de la inteligencia artificial»; ROSE (2023) de BBC NEWS MUNDO con su artículo «Los trabajadores que ya han sido sustituidos por la inteligencia artificial»; MARR (2023) en Forbes10years en su artículo «Estos son los 15 mayores riesgos de la IA»; GONZÁLEZ (2023) del ABC en «La inteligencia artificial se forma como activista del clima»; RATTINGER (2023) de Merca2.0 en «Las Consecuencias de la Inteligencia Artificial en el Trabajo de Marketing»; MURGAS (2023) de Forbes10years en su artículo «Pausar la Inteligencia artificial: ¡tu seguridad está en juego!»; o artículos como los de Agencia EFE (2023) sobre «El alto comisionado de la ONU advierte sobre los peligros de la IA para los derechos humanos»; o titulares de la Redacción de BBC NEWS MUNDO (2023) como «La inteligencia artificial puede llevar a la extinción de la humanidad: la advertencia de los grandes expertos en IA».

Desde nuestro punto de vista, para el desarrollo de este trabajo, lo interesante para dar respuesta a las preguntas planteadas es resolver, a su vez, otras nuevas preguntas sobre qué supone realmente la IA, qué es lo que verdaderamente se puede hacer o qué nos permite

hacer la IA y en definitiva, responder a qué se entiende por IA. Orientando dichas respuestas hacia el campo de la criminología y más concretamente hacia el de la investigación criminal.

Como pequeño adelanto a lo que expondremos a continuación y a colación de lo anterior, podemos citar a AGRAWAL et al. (2020), quienes nos dicen que la IA, lo que realmente permite realizar es la construcción de predicciones. Objetivo este, que coincide con ciencias como la criminología que en su pretensión de explicar el fenómeno criminológico pretende también la predicción de los hechos criminales, ampliando su campo práctico a la su utilización en el campo de la investigación criminal y la predicción de futuros comportamiento dañinos o peligrosos para la sociedad.

Dentro del desarrollo de la IA, lo importante, sería la alimentación de dicha inteligencia con datos. Estos datos, una vez filtrados o adecuadamente tratados, serían generadores de conocimiento, tal y como nos indican ZARZA y LÓPEZ MURPHY (2017). Por lo tanto, IA y tratamiento de datos irían íntimamente unidos.

Cuando la cantidad de datos a tratar es enorme, podemos empezar a hablar de *Big Data*, aunque esta, la cantidad o volumen, no sería la única característica de dicho concepto. Si tomamos la definición de ZARZA y LÓPEZ MURPHY (2017), *Big Data* sería la resolución de problemas de tratamiento de datos cuando estos se presentan en gran cantidad, variedad y con una gran velocidad de generación de los mismos. Por lo tanto, volumen de datos, variedad de los mismos y velocidad con la que se generan, serían las tres características fundamentales de este concepto, estas sería las tres «V» que ya introdujo Doug Laney en 2001. A estas características, posteriormente, diversos autores, irían añadiendo otras distintas.

En conclusión, con la idea expuesta, la IA, sobre todo en su versión de *Deep learning*, sería muy dependiente de ser alimentada con gran cantidad de datos y como consecuencia estaría íntimamente ligada al *Big data*. De hecho es interesante la observación que hacen ZARZA y LÓPEZ MURPHY (2017, p. 151) cuando afirman que «la inteligencia artificial, como la conocemos hoy, es un fruto primariamente del *big data*». En el mismo sentido otros autores como VALLS PRIETO (2017), también resaltan esta idea de relacionar datos con la IA como elemento clave. El tratamiento de estos datos será, como expondremos más adelante, la fuente principal de los problemas de la aplicación de la IA al mundo de la Criminología.

2.1.1. Breve reseña histórica

A la hora de conocer cualquier fenómeno que pretendamos estudiar es interesante presentar su evolución histórica, al objeto de poder profundizar en los distintos aspectos del mismo, enmarcándolo en su contexto temporal. Por ello se expone a continuación una breve reseña histórica de la evolución de la IA a lo largo de la historia de la humanidad.

2.1.1.1. Antes del siglo XX

AMADOR HIDALGO (1996) hace referencia a una prehistoria de la inteligencia artificial que abarcaría desde tiempos inmemoriales hasta la mitad del siglo XX. Hace referencia, por ejemplo, a artilugios o autómatas elaborados que intentaban simular a seres vivos. Indica también las máquinas creadas para facilitar el trabajo de razonamiento del hombre, con la creación de máquinas de cálculo construidas por culturas como la China o Japonesa. Sin embargo, para el autor el punto de referencia sería el siglo XIX con Charles Babbage creador de las máquinas diferencial y analítica. Por ello se le considera el padre de la computación. Otros pioneros que cita serían Herman que crearía la primera máquina de computación de medios mecánicos o la primera sumadora mecánica de Borroughs en 1892.

Por otro lado RUSSELL, NORVIG y CORCHADO RODRÍGUEZ (2008) citan que, ya en el año 250 a.C., Ktesibios construiría la primera máquina autocontrolada, que consistía en un reloj por agua. Y ABELIUK y GUTIÉRREZ (2021) indican que la fue Ada Lovelace, en 1842, quién programaría el primer algoritmo con el objetivo de ser ejecutado por una máquina.

2.1.1.2. Desarrollo de la IA en el siglo XX

Durante el desarrollo de la IA existen distintos hitos históricos que no pueden desdeñarse uno de esos hitos sería el desarrollo de la prueba de Turing, creada por Alan Turing en 1950 y que citan RUSSELL et al. (2008) en su libro. Esta prueba vendría a establecer que capacidades debería cumplir una máquina para establecer si podría o no considerarse inteligente. De este modo dichas capacidades deberían de permitir a la máquina un procesamiento del lenguaje natural, una capacidad de representar el conocimiento, disponer de habilidades de razonamiento que permitiera responder preguntas y sacar conclusiones utilizando la información guardada y por último capacidad que le permitiera un aprendizaje compatible con la adaptación a nuevas circunstancias y obtención de nuevos patrones.

Autores como ABELIUK y GUTIÉRREZ (2021) y como RUSSELL et al. (2008), señalan que dentro del siglo XX, los primeros que trabajaron en IA fueron Warren McCulloch y Walter Pitts en 1943. Estos utilizaron los conocimientos de ese momento sobre neuronas, lógica proposicional de Russell y Whitehead y la teoría sobre computación elaborada por Turing, para proponer un modelo de neuronas artificiales. A partir de esto surgiría la regla de Hebb que daría paso a al aprendizaje de Hebb. Y será en 1951 cuando se construya el primer computador.

Hasta la década de los 70 la IA viviría una edad dorada de continuos avances. Sin embargo, a partir de los primeros años de 1970, hasta la década de los 90, llegó un periodo que es conocido como invierno de la IA, principalmente debido al menor avance en esta disciplina y el golpe de realidad percibido sobre sus limitaciones. A partir de esos año vuelve a resurgir un periodo floreciente de la IA con desarrollo de programas como «*Deep blue*» de IBM que vencería a Kasparov en una partida de ajedrez. (PINO, RICO-CARRILLO y HERNÁNDEZ-ARANGO, 2022).

2.1.1.3. Siglo XXI. El presente.

El despertar surgido a partir de la última década del siglo XX se continuará durante este siglo XXI, con continuos avances y desarrollos. Así, según exponen ABELIUK y GUTIÉRREZ (2021, p. 17) en: 2002 surgiría el robot «*Roomba*», un éxito comercial; en 2014 se desarrollan las redes generativas adversarias que permite a la máquina generar imágenes, video o voz; en 2014 se comercializa «*Alexa*», como asistente virtual; 2015 la IA al alcance de todos con la aparición de código abierto en librerías; 2017 «*AlphaGo*» una IA que vence en el juego del Go al campeón mundial.

A los anteriores hitos podríamos añadir otros aportados por PINO, RICO-CARRILLO y HERNÁNDEZ-ARANGO (2022), como el año 2000 cuando surge el concepto de aprendizaje profundo (*Deep Learning*) o 2007 con el inicio del proyecto de IBM Watson.

2.1.2. Big Data

En relación con la IA existirían una serie de conceptos que estarían íntimamente ligados con la misma. Este es el caso del *Big Data*. Hoy en día este término no parece estar tan en boca de todos, a pesar de que su importancia es vital para el desarrollo de la IA. Autores como

DIEBOLD (2012) consideran el *Big Data* como una disciplina emergente distinta de las disciplinas tradicionales como la estadística o ciencias de la computación.

Como bien señalan HERNÁNDEZ-LEAL, DUQUE-MÉNDEZ y MORENO-CADAVID (2017), el concepto *Big Data* no solo va dirigido a un enorme volumen de datos, sino también a su variedad y la velocidad con la que se procesan y se acceden a dichos datos. Según los mismos autores, del *Big Data* se desarrolla la ciencia de los datos como el conjunto de técnicas que permiten trabajar con los mismos a través de su manipulación y su tratamiento. Este procesado de datos implicaría cuatro fases que serían: la propia generación de los datos, la adquisición o captura de los mismos, su almacenamiento y finalmente, su análisis.

CASAS ROMA, NIN GUERRERO y JULBE LÓPEZ (2019) indican que el término *Big Data*, que apareció en el siglo XXI, podría traducirse por datos masivos. Estos autores destacan que el verdadero valor de los datos masivos no es la disposición de una gran cantidad de datos sino en su examen e interpretación de los mismos. Para este proceso se utilizarían distintas técnicas como el empleo de algoritmos, que podrían aplicarse de forma secuencial o en paralelo durante el análisis de los datos.

TABARES y HERNÁNDEZ (2014) también resaltaron que el verdadero núcleo del *Big Data* estaría en la capacidad de extraer la información adecuada de los datos y no en su mero almacenamiento. Para estos autores las características de los datos vendrían dadas por el volumen, la velocidad, la variedad y también por el valor de los datos que estaría relacionado a su vez con la veracidad, validez y la volatilidad de los mismos.

Y MEDINA LA PLATA (2023), precisa que el término tiene su origen en la función que pretende cumplir, que sería la manipulación de una gran cantidad de datos. Señalando, además, que no existiría un consenso para establecer que se entiende por gran cantidad de datos a partir de los cuales podríamos hablar de *Big Data*. Estos datos podrían presentarse de forma estructurada, o bien no estructurada, de forma que permitieran la toma de decisiones independientemente del tamaño del conjunto. Los algoritmos permitirían la recogida y lectura de esa gran cantidad de datos.

En la Resolución P8_TA (2017) 0076 del PARLAMENTO EUROPEO (2017, p. 2) sobre las implicaciones de los macrodatos en los derechos fundamentales, se habla de macrodatos como sinónimo de *Big Data* como «la recopilación, análisis y acumulación constante de

grandes cantidades de datos, incluidos datos personales, procedentes de diferentes fuentes y objeto de un tratamiento automatizado mediante algoritmos informáticos y avanzadas técnicas de tratamiento de datos». Esta resolución relaciona además, en su considerando B, la utilización de los macrodatos con la utilización de dispositivos de IA, al objeto de predecir determinados sucesos.

BERLANGA (2016) por último, diferenciaría el *Big Data* del *Data Analytics*. El primero se daría cuando se utiliza sobre los datos técnicas inductivas del aprendizaje automático. El segundo se referiría al caso de tratamiento de datos desde la estadística clásica cuantitativa y cualitativa. De este modo, el *Big Data* quedaría relacionado con la IA.

2.1.3. Definición de Inteligencia Artificial

Inteligencia Artificial se compone de dos palabras, por un lado inteligencia que según la Real Academia de la lengua Española (en adelante, RAE) podría entenderse como capacidad de entender o comprender algo. O en su segunda acepción, como capacidad de resolver problemas. Y por otro lado se compone de la palabra artificial, que desde la RAE, nos indica que es un adjetivo que indica la cualidad de algo hecho por hombre o algo que no es natural o falso. Si atendemos a ambas palabras juntas, la definición que nos ofrece la RAE, en lo referente a inteligencia artificial, nos habla de «Disciplina científica que se ocupa de programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana, como el aprendizaje o el razonamiento lógico».

AMADOR HIDALGO (1996) destaca que no existe una definición clara de lo que es la IA ni un consenso al respecto. La dificultad estriba entre otros motivos en la dificultad de definir qué se entiende por inteligencia. Por lo tanto, según el autor existirían distintas definiciones de IA en función de distintos criterios a tener en cuenta. De este modo existiría una definición basada en el objetivo del estudio de los procesos cognitivos y otra se basaría en la creación de sistemas automáticos.

En la misma línea que el anterior autor citado, RUSSELL et al. (2008), dan distintas definiciones, de distintos autores, sobre IA en función de dos conceptos distintos, los procesos mentales o el razonamiento. Así por ejemplo, mencionan, en relación con el primer concepto citado la definición de KURZWELL (1990) que hablaría de la IA como un arte que buscaría crear máquinas con capacidad para realizar funciones que cuando son realizadas por humanos

implicaría inteligencia. Mientras que en relación con el segundo concepto citarían entre otros a WINSTON (1992) quien definiría la IA como estudio de distintos cálculos que harían posible la percepción, el razonamiento y la actuación de las máquinas.

MUÑOZ RODRÍGUEZ (2020) nos indica en su artículo que la primera definición de IA fue dada por John McCarthy en 1955 y que según el cual, la definía como una ciencia y una ingeniería utilizada para crear máquinas inteligentes, orientadas a los resultados.

Para ROUHIAINEN (2018) inteligencia artificial sería aquella capacidad de las máquinas por la que estas pueden tomar decisiones, de forma similar a los humanos, a partir de la utilización de distintos algoritmos que permitiría la utilización y tratamiento de datos a partir de los cuales se generaría el aprendizaje citado.

Según dicho autor los errores cometidos de con la utilización de dicha tecnología serían mucho menores que el caso de las decisiones tomadas por humanos. Sin embargo, no habría que descartar la existencia de posibles problemas relacionados con esa tecnología, frente a los que habría que estar pendientes, en aras de evitar complicaciones y problemas futuros.

Desde el punto de vista de LÓPEZ DE MÁNTARAS BADIA y MESEGUER (2017) el objetivo final de la IA sería lograr que la maquina tuviera una inteligencia similar a la humana. Sin embargo, esto plantea el problema sobre que se entendería por inteligencia. Sobre esta cuestión los autores indican que la inteligencia no sería algo exclusivo del género humano y que por lo tanto existirían distintos tipos de inteligencia por lo que en el caso de IA se podría estar hablando de un nuevo tipo de inteligencia. Y BODEN (2017), también habla del objeto de la IA como la búsqueda de que los ordenadores realicen las mismas cosas que los humanos.

Por último, CABALLERO y MARTÍN (2022), hablan de IA como un campo de la informática que buscaría replicar tareas complejas que realizan los humanos de forma habitual. Entre dichas tareas se encontrarían acciones de reconocimiento de voz, escritura, imágenes, o la toma de decisiones en función de los datos disponibles.

Asimismo, desde la propia Unión Europea (en adelante UE), se ha definido que se entiende por IA. En su comunicación COM (2018) 237 final, la COMISIÓN EUROPEA (2018) indica que la definición inteligencia artificial se aplicaría a aquellos sistemas que para alcanzar los fines programados, tendrían la capacidad para realizar un análisis de lo que le rodea y

seguidamente reaccionar con cierto grado de autonomía y cuyo comportamiento podría considerarse inteligente.

Por otro lado, desde el Centro Común de Investigación (JRC en sus siglas en inglés), de la Comisión Europea, en su informe técnico sobre la definición de la IA, de 2020, presenta una definición muy extensa y completa:

Los sistemas de inteligencia artificial (IA) son sistemas de software (y posiblemente también hardware) diseñados por humanos(2) que, dado un objetivo complejo, actúan en la dimensión física o digital al percibir su entorno a través de la adquisición de datos, interpretando los datos recopilados estructurados o no estructurados, razonando sobre el conocimiento o procesando la información derivada de estos datos y decidiendo la mejor acción o acciones a tomar para lograr el objetivo dado. Los sistemas de IA pueden usar reglas simbólicas o aprender un modelo numérico, y también pueden adaptar su comportamiento analizando cómo el entorno se ve afectado por sus acciones anteriores. (JRC, 2020, p. 9)

En consecuencia, a través de lo expuesto podemos afirmar que la relación entre IA y *Big Data* es estrecha y complementaria. La propia COMISIÓN EUROPEA (2020, p. 10), en su Libro Blanco sobre la inteligencia artificial indica, «Sin datos, el desarrollo de la IA y otras aplicaciones digitales resulta imposible.», apuntalando la importancia de esta relación.

También, en la Resolución P8_TA (2019) 0081, el PARLAMENTO EUROPEO (2019) señala que la IA es dependiente de la recogida de un gran volumen de datos. Esta relación se basaría en: la utilización del *Big Data* como fuente de datos para la IA; la utilización IA como herramienta para analizar *Big Data*, contribuyendo a la mejora de su rendimiento y a la optimización. Ambas herramientas presentaría desafíos en el campo de la ética al plantear cuestiones sobre privacidad o la influencia de los algoritmos sobre la vida íntima de las personas. Otro desafío se referiría a la protección de derechos fundamentales así como en la necesidad de lograr transparencia de los procesos.

2.1.4. Tipos de Inteligencia Artificial

Según RUSSELL et al. (2008) podríamos clasificar la IA en función del punto de vista o enfoque que queramos desarrollar, de este modo tendríamos cuatro tipos o formas de IA que conllevaría, a su vez, cuatro distintas formas de definir la IA. Todo ello en función de dos parámetros fundamentales, el primero sería uno referido al comportamiento humano y el

segundo estaría referido a la capacidad racional o capacidad de razonar. De este modo existirían, de acuerdo con los autores citados, los siguientes tipos de IA:

1. IA basada en sistemas que actúa como una persona. Emulan a los humanos
2. IA basada en sistemas que piensa una persona. Con capacidad para resolver problemas, aprender y tomar decisiones.
3. IA basada en sistemas que piensan racionalmente de la misma forma que lo haría una persona. Sistemas que piensan racionalmente.
4. IA basada en sistemas que actúan racionalmente como una persona. Entendiendo esta actuación como una intención de lograr el mejor resultado posible, haciendo que la máquina realice las ilaciones correctas para tal fin.

REBOLLO DELGADO (2023), expone en su libro la misma clasificación indicada arriba, aunque también indica otra clasificación que divide la IA en tres tipos:

1. IA específica o débil. Caracterizada por la utilización de un proceso inductivo y simbólico, imitando los procesos humanos.
2. IA general o fuerte. Que permite a la máquina resolver problemas de forma autónoma, utilizando un proceso deductivo.
3. Superinteligencia. Que sería superior a la inteligencia humana siendo totalmente independiente de la intervención del hombre.

Desde otro punto de vista, CORTINA ORTS (2019) hablaría de distintos enfoques de la IA que distribuiría la misma en: *Machine Learning*, *Machine reasoning*, la robótica y aquel que integra todas las señaladas. Proponiendo a su vez de una división de la IA similar a la que propone REBOLLO DELGADO. Así tendríamos:

1. Inteligencia superior o superinteligencia. Que sería aquella que supera a la del hombre.
2. Inteligencia general. Que sería la inteligencia propia del ser humano.
3. Inteligencia especial. Que sería la que se utilizaría para trabajos muy específicos o tareas concretas como por ejemplo jugar al ajedrez.

ROJAS (2015), a su vez, establece que la IA se divide en:

1. IA clásica. Que emplearía una lógica simbólico- deductiva. Se daría un conocimiento previo así como las soluciones y se daría las respuestas en función de probabilidades.

2. IA computacional. Basada en una lógica simbólico- inductiva. En este caso se darían datos que proporcionarían un aprendizaje en función de la experiencia adaptándose a la realidad.

Y mucho antes, SEARLE (1980, p.2) hablaba de la necesidad de clasificar la IA en dos tipos:

1. «*Strong AI*» como aquella inteligencia que estaría impregnada de intencionalidad.
2. «*Weak AI*» o bien «*Cautions AI*». Que sería aquella que es programada por el hombre para ser utilizada como herramienta que mejore la capacidad de esta para tareas concretas.

En relación con los tipos de IA, otros autores como LÓPEZ DE MÁNTARAS BADIA y MESEGUER GONZÁLEZ (2017), asimismo, exponen los mismos tipos de IA de SEARLE (1980)

Desde el punto de vista del enfoque que se puede dar a la IA, autores como ROUHIAINEN (2018) hablan de *Machine Learning* o Aprendizaje automático y de Aprendizaje profundo o *Deep learning*, como parte del aprendizaje automático con mayor potencia y capacidad de resolución de problemas complejos. En este caso, se utilizarían las llamadas redes neuronales que requeriría una gran cantidad de datos para realizar su trabajo. (Ver Figura 1, en Anexo A).

Según ROUHIAINEN (2018), el aprendizaje automático supondría uno de los campos de la IA que consistiría en la capacidad de aprendizaje de las máquinas sin estar programadas para ello.

Respecto al concepto de aprendizaje automático, PINEDA PERTUZ (2022), toma la misma definición, citando para ello ARTHUR SAMUEL (1959) cuando define aprendizaje automático como un área de estudio que permitiría a un computador aprender, pero sin necesidad de estar programado específicamente para ello y SURESH y GUTTAG (2021) indican que el aprendizaje automático es un tipo de inferencia estadística que aprende, a partir de datos existentes, una función que puede generalizarse a datos nuevos y no vistos previamente.

Dentro del aprendizaje automático existirían tres campos: el aprendizaje supervisado, el no supervisado y el de refuerzo. (ROUHIAINEN, 2018). La misma clasificación es expuesta por PINEDA PERTUZ (2022). (Ver Figura 2, en Anexo A).

Autores como SIMÓN CASTELLANO (2022) refiere que bajo las mismas palabras, inteligencia artificial, se incluirían gran número de técnicas avanzadas en procesamiento de datos y supercomputación. El autor cita entre estas técnicas, las siguientes: *Data mining*, *Big Data* o

procesamiento masivo de datos, *Blockchain*, *Machine Learning*, *Deep Learning*, procesamiento del lenguaje Natural, reconocimiento facial o redes neuronales.

Y por último podemos citar a LANCHO PEDRERA (2003), quien diferencia entre sistema experto e IA, considerando el sistema experto como una parte de la IA que se basaría en un sistema que da respuestas a problemas planteados en función de la experiencia humana trasladada a la máquina. Planteando el dilema ético que supone la aplicación de estos sistemas al mundo judicial, en tanto que se daría a una maquina capacidad de toma de decisiones judiciales en base a una simulación de la lógica humana.

Por lo tanto, a parte de las distintas clasificaciones de IA según su naturaleza o tipo, podríamos a la vez hablar de subconjuntos de la IA, que de forma resumida podría clasificarse de la siguiente forma:

1. *Machine learning* que a su vez incluiría:
 - a. *Deep learning*
2. Sistemas expertos
3. Visión artificial que a su vez incluiría
 - a. Reconocimiento facial
 - b. Reconocimiento de imagen
4. Procesamiento del lenguaje natural
5. Robótica
6. Redes neuronales

2.1.5. Aprendizaje automático «*Machine learning*»

Para MC QUILLIAN (2018) no existiría inteligencia en la IA simplemente se trataría de una minimización matemática que trata de dibujar una línea en un lienzo lleno de puntos que más se ajuste a dichos puntos. En consecuencia, lo que se llama aprendizaje no sería realmente tal.

En el mismo orden de cosas, para PASQUINELLI (2022), el aprendizaje automático pretende predecir un resultado a partir de la entrada de unos datos. Realmente la salida de datos se correspondería con una probabilidad de clasificación, por ejemplo 97% de probabilidad de ser perro y 3% gato. En definitiva, lo que realiza es una aproximación, exponiendo, al igual que MCQUILLIAN, que, en realidad, el aprendizaje automático no es aprendizaje si no una

aproximación o ajuste de los datos de entrada a un modelo obtenido previamente del entrenamiento.

En relación con los tipos de aprendizaje automático, para PASQUINELLI (2022) existirían tres formas: el entrenamiento o abstracción de patrones, la clasificación o reconocimiento de patrones y la predicción o generación de patrones.

A su vez nos encontraríamos con tres fases o etapa del procesos: los datos de entrenamiento de donde el sistema IA deberá obtener los patrones, aquí el conjunto de datos será distinto en atención a si estamos hablando de supervisado o no supervisado; el propio algoritmo de aprendizaje que es el que obtendrá los patrones de los datos de entrenamiento y con ello elabora una descripción estadística. Esto siempre presentará errores, deteniéndose el aprendizaje cuando alcanza una precisión que se considera aceptable; y por último, tendríamos la aplicación solo cuando se considera suficientemente entrenado.

Respecto del proceso, para SURESH y GUTTAG (2021), el aprendizaje automático se dividiría en seis fases que incluirían: la recopilación de datos identificando la población objetivo,; una fase de preparación de datos, mediante distintas técnicas en función del objetivo pretendido y de la propia naturaleza de los datos; una fase de desarrollo del modelo utilizando datos de entrenamiento pero no los de validación, el entrenamiento ser realizaría en función del objetivo y se elige el mejor modelo para la situación dada; fase de evaluación del modelo, donde se utilizan datos de prueba o incluso otros datos disponibles conocidos como datos de referencia para demostrar la robustez del modelo seleccionado; fase de posprocesamiento del modelo donde se ajustaría más el modelo; por último, la fase de implementación del modelo.

Y por último podemos citar a FERRANTE (2021), quien explica que el aprendizaje supervisado sería aquel en el que el entrenamiento de la máquina se guía mediante la alimentación de datos con su etiqueta. Es decir, a cada dato se le indica a que categoría pertenece y en base a esto se pretende que encuentre patrones que sean capaces de predecir datos de entrada futuros. En estos modelos la calidad de los datos puede ser crucial ya que a desde las características de esos datos la IA realizará la categorización correspondiente y por lo tanto la predicción futura. Así siguiendo un ejemplo similar al expuesto por el autor, si en los datos solo aparecen elementos A blancos y B negros el color podría suponer una característica de

categorización relevante para la máquina pero claramente implicaría errores futuros de predicción si se introduce elementos A de color negro.

2.2. Algoritmos

2.2.1. Definición de Algoritmo

La RAE establece, para la palabra algoritmo, dos acepciones, la primera lo define como un «Conjunto ordenado y finito de operaciones que permite hallar la solución de un problema» y la segunda habla de un método y forma de notación con distintas formas de cálculo.

Respecto de la IA, para STEINER (2012), algoritmo sería una lista de reglas o indicaciones que llevarían a partir de unos datos iniciales a una respuesta o resultado determinado. Y COTINO HUESO (2019), define algoritmo como una serie de órdenes o comandos que hacen que una computadora transforme una entrada en una determinada salida. Y en relación a sus características señala que podríamos encontrarnos con algoritmos diseñados de forma intencionada para discriminar y otros que discriminen de forma colateral, es decir, sin buscar dicho efecto.

Por otro lado, en conexión con el concepto de algoritmo, desde el punto de vista del mundo de la informática, MONASTERIO ASTOBIZA (2017) hablaría de un «*software*» que va a trabajar con una serie de instrucciones, que serían limitadas.

Es interesante la aportación de DANAHER (2016), respecto al enorme desarrollo de los algoritmos en nuestra sociedad, el cual habla de la existencia de una «algorocracia» refiriéndose así a un tipo de sistema de gobierno donde se basa la estructura y la organización en computadores programados con algoritmos. Lo que implica que los algoritmos de por sí no tienen que estar, necesariamente, implementados en computadores.

Por último, podemos citar a MERCADER UGUINA (2022, p. 18), quien define algoritmo como una «receta» resultado de una serie de directrices matemáticas que pretende llegar a un resultado.

2.2.2. Algoritmo e Inteligencia Artificial

Al igual que la IA está unida a la existencia de datos y como consecuencia al *Big Data*, también está ligada, de forma ineludible, a los algoritmos como parte fundamental de su constitución. De hecho los algoritmos, considerados de forma aislada, son consustanciales a muchos

aspectos de nuestra sociedad actual como muestra SIMÓN CASTELLANO (2022), quien destaca que en la vida diaria no podemos obviar los algoritmos, ya que son una realidad, que condicionan nuestras decisiones y que está presente en muchos de los aspectos de nuestra vida, como puede ser en el caso de la publicidad. Además, como este autor resalta, estos algoritmos no estarían libres de fallos notables.

La afirmación inicial de este apartado, es confirmada por autores como MERCADER UGUINA (2022), para quien algoritmo e IA van necesariamente unidos. Aunque, no necesariamente todo algoritmo implicaría IA, es decir, el algoritmo puede o no existir dentro de la IA.

La Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (en adelante UNESCO), elaboró su Recomendación sobre la ética de la inteligencia artificial, adoptada el 23 de noviembre de 2021 que se constituyó como la primera norma con ámbito mundial sobre la ética en la IA. En la misma, la UNESCO (2022, p.10), relaciona los algoritmos con la IA como parte constituyente de su definición, cuando expresa que «los sistemas de IA son tecnologías de procesamiento de información que integran modelos y algoritmos que producen una capacidad para aprender y realizar tareas cognitivas, dando lugar a resultados como la predicción y la adopción de decisiones»

En conclusión podemos afirmar a la vista de lo planteado que algoritmo e IA constituyen realidades íntimamente relacionadas.

2.2.3. Algoritmos como cajas negras y posibles fuentes de sesgos y discriminación.

El hecho de que los algoritmos se encuentren muy presentes en nuestra vida diaria como expresa SIMÓN CASTELLANO (2022) hace que estas herramientas adquieran una notable importancia a la hora de la aplicación de la IA, y sobre todo en su aspecto de capacidad explicativa es decir, en que una de sus características sea que pueda ser explicada la razón de sus resultados.

En este sentido, MONASTERIO ASTOBIZA (2017, p. 186), es muy claro al expresar que «Los sistemas algorítmicos son laberintos incomprensibles» y lo son incluso para aquellos que los crearon. También, MONASTERIO ASTOBIZA (2017, p. 188), expresa su preocupación cuando dice que «los algoritmos son cuna caja de Pandora». Y AGRAWAL et al. (2020) igualmente, hablan sobre que los resultados obtenidos de la utilización de la IA parecen obtenidos de una caja negra donde el proceso por el cual se arroja la predicción no es controlado por nadie.

Sobre este particular podemos encontrar noticias como la que escribe MAYBIN (2016), donde se muestra que la existencia de sesgos determinados algoritmos en las herramientas de decisión automática, que utilizan actualmente en algunos países, puede tener serias consecuencias sobre la vida de aquellas personas sobre las que se les aplica. Efectivamente, BELLOSO MARTÍN (2022), advierte que diversos estudios alertan sobre posibles daños a derechos fundamentales, debido a la aplicación de las nuevas tecnologías digitales, las cuales llevarían implícitas los propios sesgos y exclusiones que se dan en las acciones humanas. Por lo tanto no habría que dejarse engañar acerca de la neutralidad y objetividad de la tecnología. Este hecho se debería a la existencia del sesgo algorítmico. Dichos sesgos vendrían dados a partir de tres momentos diferentes: el momento de recogida de datos, el de preparación de los datos recogidos y finalmente el momento de procesar los mismos para la toma de decisiones.

Respecto del primer momento FERRANTE (2021), que los sesgos producidos como causa de los datos recogidos, en muchas ocasiones, serían debidos a que los datos no representan, adecuadamente, a la población que pretenden describir. Incluso el autor asegura que, los sistemas de IA, al ser creados por personas con su propia visión de la realidad y sus propias formas de interpretarla, pueden, de alguna forma, transmitir esas características al propio sistema que irremediablemente se vería afectado por un sesgo.

Y por otro lado VÉLEZ, GÓMEZ SANTAMARÍA y OSORIO SANABRIA (2022) recuerdan que los sesgos dentro de la IA pueden producirse por dos vías distintas. La primera vendría dada por los propios datos con los que se entrena el algoritmo, los cuales adolecerían ya de por sí de un sesgo o error durante el proceso de recolección o tratamiento. La segunda vía sería la que se provocaría por la construcción del propio algoritmo, en el cual se introducirían los propios prejuicios del programador.

También, VALENZUELA (2021) especifica que, aunque relacionados íntimamente, algoritmo e IA no son exactamente lo mismo. La IA vendría a ser un intento de simular la inteligencia humana y los algoritmos serían procedimientos construidos en base a unas instrucciones o rutinas que buscarían resolver un problema a partir de unos datos facilitados. Estos algoritmos podrían acarrear graves consecuencias en relación con la justicia de las decisiones, en el sentido que, usado para la toma de decisiones en un entorno humano, los problemas a resolver serían complejos y difíciles de reducir a únicamente una serie de instrucciones o

rutinas. Respecto de esto último, la autora destaca que, en el proceso de aprendizaje automático, una de las dos etapas estaría fuera de control de cómo se lleva a cabo el proceso, corriendo el peligro de caer en el sesgo.

SURESH y GUTTAG (2021) defienden que es imprescindible conocer en qué momento, de todo el proceso de generación de aprendizaje automático, se puede introducir el error. Como dichos autores destacan, no todos los problemas derivados del aprendizaje a automático provienen de los datos.

La mención de la IA como cajas negras es realizada, en el mismo sentido, por MERCADER UGUINA (2022) en su libro, donde expone brevemente la dificultad de comprender el proceso de toma de decisiones de la IA debido a la suma de diversas capas de algoritmos (redes neuronales) donde cada capa realiza su propio cálculo y su resultado es emitido a la siguiente capa, creándose una cadena de resultados intermedios obtenidos a partir de los distintos algoritmos aplicados, lo que hace que el seguimiento de dichos resultados o decisiones de la IA sea algo altamente complicado.

Desde otro punto de vista, LEFEVRE CERVINI (2022), encuentra que el principal problema de la aplicación de la IA al entorno de la justicia sería su falta de transparencia puesto que en caso de existir un mal diseño eso podría llevar a decisiones erróneas. La existencia de sesgos en el diseño de la IA podría causar desmedidas consecuencias a determinadas comunidades.

VÁZQUEZ BARRIO y SALAZAR GARCÍA (2023) hacen referencia a la importancia de tener en cuenta al programador a la hora de evitar o reducir los posibles sesgos que esa misma persona puede transmitir al algoritmo ya que estos sesgos son fuente de distintos tipos de discriminación y generadores de injusticias o visiones sesgadas de la realidad que perjudican la toma de decisiones. También BARONA VILAR (2021b) pone de manifiesto el peligro de la existencia de sesgos dentro del ámbito de la IA que pueden influenciar negativamente generando problemas de desigualdad, segregación, etiquetamiento y desprotección. Para esta autora la solución estribaría en la cooperación internacional. Para la autora, la inevitable existencia de sesgos pone en entredicho la utópica objetividad de la máquina en todos los procesos de decisión donde se emplean, ya que los algoritmos utilizados no serían neutros. Entre las consecuencias que pueden verse ya en la actualidad cita por ejemplo, algoritmos utilizados en EEUU que discriminaba entre raza negra o blanca para la atención sanitaria o

algoritmos de predicción de la criminalidad que presentan mayor riesgo para las personas de raza negra.

COTINO HUESO (2022) señalan que cuanto menor sea el control humano, ejercido sobre la IA, mayor será la necesidad de tener transparencia en la aplicación de la misma. Y PASQUINELLI (2022) se plantea el significado de fallo en la IA, comparando la IA con una lente que aporta sus propias distorsiones. Es decir, la IA actuaría como una lente con defectos que distorsiona la realidad que se ve a través de ella. Este autor resalta la existencia de las dos visiones de la IA una negativa y otra positiva y optimista, no posicionándose en ninguna y defendiendo una visión crítica.

Respecto de los algoritmos O'NEIL (2016, p. 8) habla de «armas de destrucción matemática» como herramientas que a través de los datos, con los que se alimentan, plantean decisiones que en muchos casos son perjudiciales y sesgadas. Para la autora el verdadero problema se encontraría en la elección de los datos que son tratados de una forma opaca y en la que siempre serían discriminados grupos de personas más desfavorecidos por la sociedad.

Desde otro ámbito para el FRA (2022) no hay duda de que la tecnología no es neutral, es decir, allí donde está presente el hombre con sus errores de sesgo en la toma de decisiones, este puede transferir dicho sesgo a la máquina. La respuesta a cómo se introduce estos errores en el proceso algorítmico es compleja. Esta agencia indica que los algoritmos de la IA afectan a personas de todo el mundo. No obstante, la cuestión central radica en qué ocurre si los derechos fundamentales de los ciudadanos europeos se ven comprometidos debido a los sesgos presentes en los algoritmos. Según las investigaciones de este organismo la IA puede conculcar derechos como el de privacidad o influir sobre la toma de decisiones discriminatorias, con serias consecuencias en la vida de los individuos.

Dos de los problemas principales que pueden acarrear la existencia de sesgos serían en un primer lugar la retroalimentación del sistema, es decir, la salida de un algoritmo se utiliza como entrada que actualiza el propio sistema, produciéndose una retroalimentación del propio sesgo, ya que un sesgo en los datos de inicio se va repitiendo de forma continua al ser utilizado en como entrada en el siguiente ciclo. Consecuencia de lo anterior podría llevar a una potenciación del sesgo empeorándolo. Esto es particularmente problemático cuando se aplica a la IA en el ámbito de la aplicación de la ley, como podría ser la policía predictiva.

El segundo caso sería la detección de discursos de odio en línea, basado en metodologías avanzadas de aprendizaje automático y procesamiento del lenguaje, que sin embargo, pueden llevar a resultados sesgados debido al propio idioma utilizado para la predicción. Arrojando resultados distintos dependiendo de si es inglés, italiano o alemán el idioma de referencia.

A pesar de todo, hay que tener en cuenta que, como MILLER (2018) defiende, parte de la solución a los sesgos producidos por los algoritmos se encuentra en los mismos algoritmos.

En resumen, respecto del sesgo o error, el verdadero problema estaría en que, debido a la IA, puede arrojar resultados no esperados con mayores o menores consecuencias en función de lo que estemos pretendiendo predecir. Evidentemente no tendrá las mismas consecuencias que el sesgo trate sobre la no identificación de imágenes entre gatitos o perros que si trata de evaluar la futura reincidencia penal de un individuo determinado. Sin embargo, como ya se expuso anteriormente, existen distintas herramientas y procedimientos para detectar, medir y mitigar estos errores.

2.3. Sesgos

CUATRECASAS MONFORTE (2023) atribuye, en su trabajo, a BILL GATES la frase «La primera regla de cualquier tecnología utilizada en una empresa es que la automatización aplicada a una operación eficiente magnificará la eficiencia. La segunda es que la automatización aplicada a una operación ineficiente magnificará la ineficiencia». Esta idea podríamos enlazarla con el concepto de que toda tecnología aplicada a la realidad no es infalible y presenta errores.

Cuando pretendemos aplicar la tecnología y como consecuencia de ello, la matemática, a través de los algoritmos al mundo real, nos encontramos que esta no es infalible y se ve sometida a la imperfección de la realidad, apareciendo entonces, errores y fallos del sistema utilizado, que conviene tener en cuenta o al menos cuantificar.

Estos errores y sesgos pueden verse acrecentados, si son ignorados o si no somos capaces de detectarlos. A este respecto COLINA (2023) llama la atención sobre la necesidad de considerar la capacidad de manipulación que puede existir a través de los sesgos de distinto origen, con el objetivo de evitar la manipulación algorítmica.

En relación con esa incapacidad para conocer qué es lo que está pasando, LÓPEZ DE MÁNTARAS y MESEGUER GONZÁLEZ (2017, p. 151) indican que uno de los principales

problemas actuales en el desarrollo de la IA sería la existencia de las llamadas «cajas negras» donde se producirían procesos de cálculo arrojando resultados de los que no se tendría una explicación del porqué la máquina la ha tomado. Y para los autores esa explicación sería «una característica irrenunciable».

CASACUBERTA (2021) también reseña que durante el desarrollo de los distintos algoritmos de aprendizaje automático, la asociación entre datos de entrada y de salida que hace la máquina se realiza sin intervención humana desconociendo la mayoría de las veces como, la IA, ha llegado a los resultados que arroja.

Desde nuestro punto de vista, no conocer qué, cómo y por qué, del resultado dentro del proceso de desarrollo del algorítmico podría llevar errores que cuestionarían la utilidad de los sistemas de IA. En este sentido FERRANTE (2021), defiende que el uso cada vez mayor de la IA podría acarrear a través de los sesgos implícitos en sus algoritmos una sociedad más injusta.

También, en la misma línea, se manifiestan ABELIUK y GUTIÉRREZ (2021), indicando que si bien, los procesos donde las reglas son concretas, son fáciles de comprender y explorar sobre cuál ha sido el camino tomado por la máquina, esto no es así en el caso del aprendizaje profundo donde el proceso es mucho más complejo y difícil de comprender.

Si nos posicionamos en el ámbito de la estadística tenemos que diferenciar entre sesgo y error, puesto que no son exactamente lo mismo, de este modo ODDI et al. (2018) intentan clarificar ambos conceptos, indicando que error sería algo inseparable de la estadística, mientras que sesgo sería una tendencia hacia ciertos valores que de no corregirse implicarían resultados erróneos. De lo que se podría deducir que el error estadístico nos anunciaría que probabilidad o margen de equivocación mientras que el sesgo de no tratarlo directamente nos arrojaría en brazos de conclusiones falsas. Por lo tanto, no se debe confundir sesgo con error estadístico a pesar que en muchas ocasiones se hable de ellos de forma indistinta.

2.3.1. Definición de sesgo

Para entender mejor que se entiende por sesgo podemos comenzar por la definición que nos presenta la RAE, que establece 8 acepciones para la palabra sesgo. De todas ellas, atendiendo al fin de nuestro trabajo, tomaremos la primera y la séptima. Su primera acepción hace referencia al algo torcido, cortado o situado oblicuamente, mientras que en su séptima

definición nos habla de un error sistemático, en estadística, en el que se puede incurrir al realizar muestreos o ensayos que tendería a dar una respuesta frente a otra.

Si atendemos a distintos autores, MANTEROLA y OTZEN (2015) explican que los sesgos, dentro de la investigación, serían aquellos errores que se comenten de forma sistemática y que no tiene naturaleza aleatoria o azarosa. Y por otro lado, CORTADA DE KOHAN y MACBETH (2006) exponen la existencia de errores presentes en el hombre a la hora de la toma de decisiones, que serían los sesgos cognitivos.

De las definiciones presentadas hasta aquí podemos discernir, por un lado, una definición que se refiere a la propia condición humana, es decir, a errores o formas de interpretar la realidad debido a la propia naturaleza humana y, por otro lado, una definición que afectaría al campo de la estadística fundamentalmente debido a los métodos de muestreo o errores de las medidas realizadas.

Por lo tanto, como definición básica de sesgo, dentro del campo de la investigación y la ciencia, podríamos hablar de un error producido no de forma fortuita. Estos serían de distinta naturaleza dependiendo de si estamos tratando una cuestión estadística, psicológica, o tecnológica. Dichos errores pueden acarrear problemas a la hora de la elaboración de cualquier estudio o la búsqueda de cualquier resultado. O como HERNÁNDEZ-AVILA, GARRIDO y SALAZAR-MARTÍNEZ (2000, p. 438) aseguran, «pueden comprometer la validez del estudio».

En consecuencia, una vez establecida la definición, vemos que existirían distintos tipos de sesgos. Por lo que antes de estudiar los efectos de los sesgos dentro de la IA, sería conveniente, desde nuestro punto de vista conocer algunos tipos de sesgos y explicarlos.

2.3.2. Tipos de sesgos

Si nos centramos en distintos tipos de sesgos, por un lado podríamos hablar de sesgos cognitivos, que según GARCÍA-CAMPOS, SARABIA-LÓPEZ y HERNÁNDEZ-CHÁVEZ (2022) constituirían errores o tendencias presentes en los procesos mentales humanos que surgen cuando estos pretenden captar, interpretan o enjuician el mundo que les rodea. Estos errores serían en la mayoría de los casos involuntarios y estarían relacionados con falta de datos o el exceso de rapidez en el proceso. Estos fenómenos condicionarían la forma de percibir nuestra realidad.

También COLINA (2023), habla del sesgo cognitivo como un error que se produce de forma sistemática cuando el ser humano realiza la acción de pensar o procesar un conjunto de datos determinados. Desde este punto de vista existirían distintos tipos de sesgos cognitivos, así el autor nos hablaría de: sesgo de confirmación, sesgo de anclaje, sesgo de disponibilidad, efecto halo, sesgo de falso consenso, sesgo de inversionista, atractivo de grupo, sesgo de apoyo de nuestras elecciones, sesgo de familiaridad, sesgo de pensamiento grupal o sesgo de punto ciego. Sin embargo, es importante resaltar que este tipo de sesgo no sería en principio el tipo de sesgo que se produciría en las decisiones realizadas a través de la IA. Aunque podrían encontrarse implícitas como veremos dentro del propio proceso pero por causa distintas al propio funcionamiento de la máquina.

De la segunda definición escogida podemos deducir que existiría un sesgo de carácter estadístico. Este sesgo estadístico, según ODDI et al. (2018) sería un error que es evitable ya que se correspondería con cierta tendencia a seleccionar determinados datos sobre otros. Los mismos autores hablarían de: sesgo de muestreo, sesgo de medición, sesgo de estimación y sesgo de predicción. A diferencia de los distintos tipos de errores (error estándar de estimación o el error de muestreo), que desde el punto de vista estadístico no son, de por sí un problema, en cuanto que son medibles y podrían estar controlados. Sí lo es, en cambio, el sesgo ya que podría llevar a la obtención de una solución equivocada del problema planteado.

Sin embargo, en el campo de la IA podríamos definir un nuevo tipo de sesgo que se correspondería con el algoritmo que se implementa en la máquina, que permitirá la obtención de los resultados buscados y que no debemos confundir con los sesgos anteriormente citados, este sesgo es el sesgo algorítmico.

Según la Agencia de los Derechos Fundamentales de la Unión Europea (en adelante FRA) (2022) es importante distinguir entre lo que sería sesgo o «*bias*» en terminología inglesa y discriminación, ya que una cosa no implica necesariamente la otra.

Según el FRA el término sesgo puede significar distintas cosas, como ya hemos expuesto. Puede referirse a una inclinación hacia el favorecimiento de determinados grupos humanos por características de protección como serían edad, sexo, origen étnico, etc. También puede hacer referencia a la propia diferenciación entre dos objetos o categorías. Podemos hablar de sesgo estadístico como la diferencia entre el valor real y el estimado de un determinado

parámetro. Y por último se plantea la posibilidad de ver sesgo como un desplazamiento desde el origen como parámetro estimado.

Por otro lado la discriminación estaría vinculada a los datos utilizados o incluso al resultado del propio algoritmo. Existiendo distintas formas de discriminación.

Desde la literatura de habla inglesa se habla de «*bias*». De este modo, en relación con la IA, ANDREW NG (2018) hablaría de dos tipos de errores que se producen durante el aprendizaje automático: el «*bias*» (sesgo) y la varianza. Ambos estarían relacionados el conjunto de datos que se utilizan durante el entrenamiento de los modelos de IA.

En su libro el autor define tres tipos de conjunto de datos:

1. Conjunto de entrenamientos, «*Training set*», que será sobre el que se ejecuta el algoritmos de aprendizaje.
2. Conjunto de desarrollo, «*Developing set*», que es el que se utiliza para ajustar parámetros del algoritmo. También se llama conjunto de validación.
3. Conjunto de prueba, «*Test set*», que es aquel que se utiliza para evaluar el rendimiento del propio algoritmo pero no para ajustar sus parámetros.

A grandes rasgos, el sesgo sería la tasa de error de un algoritmo en el conjunto de datos de entrenamiento cuando se tiene un conjunto de entrenamiento muy grande. Mientras que, la varianza, sería la diferencia entre error que se realiza en el conjunto de prueba y el de entrenamiento, es decir, cuánto peor va el algoritmo el conjunto de prueba en comparación con el conjunto de entrenamiento. El error total, en forma simplificada, sería la suma del sesgo y la varianza. Por lo tanto, podemos concluir que el origen de estos sesgos algorítmicos serían un mal entrenamiento en la máquina o una mala selección o tratamiento en el conjunto de datos.

CAÑADAS (2021) también habla de varianza y sesgo como dos términos que deben buscar cierto equilibrio para una correcta aplicación del aprendizaje dentro de la IA. En este sentido nos indica que el entrenamiento de un modelo determinado de IA implicaría construir una función determinada f' , a partir de los datos de entrada, siendo la varianza la diferencia entre las distintas funciones f' calculadas cuando variamos el conjunto de entrenamiento. En cambio el «*bias*» o sesgo sería el resultado de obtener varias f' muy parecidas o prácticamente sin

cambios. En el primer caso hablaríamos de una varianza alta o sobreajuste y en el segundo se hablaría de subajuste del modelo.

En sentido parecido PASQUINELLI (2022) también habla de sobreajuste del modelo cuando el modelo se ajusta demasiado a los datos de entrenamiento con lo cual pierde capacidad de predicción, y habla de infraajuste cuando el grupo de datos de entrenamiento es tan pobre que no es capaz de obtener patrones a partir de ellos debido a la pérdida de información. Este autor también señala la posibilidad de un tercer caso y es el de que el modelo aprenda patrones erróneos.

BAGNATO (2017) habla también en términos «underfitting» o subajuste cuando debido a la falta de datos suficientes de datos la IA falla en las predicciones, es decir, no es capaz de generalizar y «overfitting» o sobreajuste cuando el entrenamiento ha sido demasiado estricto por decirlo de alguna forma y el modelo se ajusta en exceso a los datos facilitados por lo que cualquier variación de los nuevos datos en comparación con el conjunto de entrenamiento no será detectada.

En consecuencia, por lo que se puede ver respecto de los errores en los algoritmos, los datos jugarían un papel fundamental en su generación. Efectivamente, como confirma PASQUINELLI (2022, p. 20), el punto fundamental de la IA serían los datos con los que se alimenta, fuente y origen principal de los sesgos, afirmando de forma contundente que «Los datos neutrales no existen».

Por otro lado y desde una perspectiva más global, si atendemos a todo el proceso de aprendizaje automático, respecto de los sesgos encontramos que autores como SURESH y GUTTAG (2021) sugieren los siguientes tipos de sesgos: sesgos históricos, debido, entre otros motivos, a estereotipos; sesgos de representación cuando la muestra no es representativa de la población; sesgos de medición debido a la aproximación sobre el dato real; sesgo de agregación cuando se utiliza un modelo único; sesgo de aprendizaje; sesgo de evaluación, cuando los datos de referencia no representan a la población de uso; y por último sesgo de implementación cuando existen discrepancia entre el problema que pretende resolver el modelo y la forma que finalmente se usa dicho modelo.

En el mismo sentido PASQUINELLI (2022) determina que cualquier conjunto de datos es únicamente una visión parcial del mundo que se pretende estudiar, obtenida a partir de un

muestreo estadístico por lo que es inevitable la presencia de errores. De este modo, según el autor, existirían tres formas o tipo de sesgo distintos: el sesgo del mundo, el debido a los datos y el sesgo producido por el algoritmo o sesgo de la propia máquina del modelo.

El sesgo debido al mundo hace referencia a que en el propio mundo está presente la desigualdad y los estereotipos. El sesgo de los datos se referiría a aquel se produciría durante el proceso de captación de los datos, así como su tratamiento (dar formato y etiquetamiento), ya que durante este proceso se ven afectados la precisión y resolución de los datos. Por último el sesgo algorítmico sería aquel se produce por errores de la propia computación, la aproximación y comprensión aplicada durante el proceso y que amplificarían los dos anteriores.

MEHRABI et al. (2021), por su parte, enumeran en su trabajo diferentes fuentes de sesgos que pueden afectar a las aplicaciones de IA. Las fuentes serían tres: aquellas que provienen de los datos y la que provienen de los algoritmos y la de los usuarios. En función de estas fuentes la tipología de los sesgos que presentan los autores es muy amplia, la cual, expondremos aquí muy brevemente debido a condicionantes de espacio, sin entrar a explicar cada uno.

Respecto a los sesgo proveniente de los datos tendríamos: sesgos de medición, donde influye como se mide y como se utiliza cada característica; sesgo de variable omitida que sucede cuando se omiten una o varias variables que son importantes; sesgo de representación, cuando las muestras no son representativas de la población; sesgo de agregación, cuando se extraen conclusiones falsas sobre individuos al observar la población; sesgo de muestreo, parecido al de representación y se da cuando el muestreo no es aleatorio; falacia de datos longitudinales; sesgo de enlace.

Respecto de sesgo de cuyo origen es el algoritmo los autores establecen los siguientes tipos: sesgo algorítmico cuando se agrega únicamente por el propio algoritmo; sesgo de Interacción del usuario, el que activaría el propio usuario sobre el algoritmo por su propia forma sesgada de clasificar o seleccionar la información; sesgo de popularidad, los elementos más populares tienden a exponerse más; sesgo emergente, como resultado de cambios en la población, valores culturales o conocimientos sociales generalmente algún tiempo después de completar el diseño; sesgo de evaluación durante la evaluación del modelo.

Por último los sesgos de usuario serían: sesgo histórico, sesgo ya existente en los propios datos; sesgo de población, que surge cuando se crea datos no representativos; sesgo de autoselección, subtipo de sesgo de selección o muestreo en el cual los sujetos de la investigación se seleccionan a sí mismos; sesgo social, cuando nuestro criterio se ve afectado por acciones de otras personas; sesgo de comportamiento en atención a distintos comportamientos según contexto; sesgo temporal el surgiría por el simple paso del tiempo entre distintas poblaciones y comportamientos; sesgo de producción de contenido.

2.3.3. Solución a los sesgos algorítmicos

La cuestión que surge ante el problema del sesgo es inmediata ¿existe solución para los sesgos que se producen durante el proceso de aprendizaje de la IA? Como respuesta a la pregunta SURESH y GUTTAG (2021) indican que para cada sesgo existiría un abanico de posibilidades de mitigar su influencia teniendo en cuenta que la solución estaría relacionada con el tipo de sesgo y la etapa en la que se produce.

Centrándonos en la varianza y el sesgo, BAGNATO (2017) establece la necesidad de llegar a un término medio en el aprendizaje del modelo de forma que no sea ni demasiado sobrentrenado (sobreajuste), ni demasiado subentrenado (subajuste). Entre los métodos que puede utilizarse para evitar lo primero el autor destaca:

1. La necesidad de tener clases de datos variadas y equilibradas con un adecuado balanceo.
2. Ajustar el tiempo y las iteraciones del entrenamiento, reducir las dimensiones si son demasiadas de forma que comprometa la cantidad de muestra
3. O tener una cantidad mínima de muestra para el entrenamiento.

En su libro ANDREW NG (2018), establece distintas formas de atacar o solucionar el problema del sesgo y la varianza producida en el entrenamiento de los sistemas de IA. De este modo el autor plantea distintas estrategias para su reducción, que a continuación planteamos de forma muy resumida:

1. Un sesgo evitable alto puede resolverse con el aumento de tamaño del modelo utilizado. Por ejemplo, se podrían agregar capas a la red neuronal utilizada.
2. Si lo que existe es una varianza alta se podrían aumentar el volumen de datos del conjunto de entrenamiento.

De aquí se podría deducir que aumentar el tamaño de la red neuronal y agregar más datos de entrenamiento, puede llevarnos a buenos resultados. Sin embargo, en la práctica, aumentar el tamaño del modelo llevará a problemas de computación haciendo el modelo más lento. Y tampoco se puede adquirir datos de forma interminable.

Además, tenemos el problema añadido de que intentar reducir el sesgo implica aumentar la varianza y viceversa. Es decir, aumentar el tamaño del modelo generalmente reduce el sesgo, pero también puede aumentar la varianza y aumentar el conjunto de entrenamiento puede provocar el aumento del sesgo.

Otra forma de actuar sobre ambos sería actuar sobre la arquitectura del modelo siendo entonces posible, aunque difícil que se reduzcan sesgo y la varianza simultáneamente.

Por lo tanto, los sesgos y varianza, pueden controlarse y optimizarse mediante distintos métodos, aunque nunca pueden eliminarse del todo. Es necesario, por lo tanto, llegar a una solución de compromiso donde ambos errores sean lo más pequeños posibles sin comprometer la capacidad predictiva del algoritmo.

La solución no es sencilla teniendo además en cuenta que en cuanto al sesgo este respondería a la formula siguiente:

$$\text{Sesgo} = \text{Tasa de error óptima ("sesgo inevitable")} + \text{Sesgo evitable}$$

Siendo también conocida, la tasa de error óptima como error de Bayes o tasa de Bayes.

Ante el sesgo evitable alto el autor propone (no sería útil aumentar más datos de entrenamiento):

1. Aumentar el tamaño del modelo, aumentando el número de neuronas o capas esto permitiría un mejor ajuste el conjunto de entrenamiento.
2. Modificar las características de entrada según las percepciones del análisis de errores ya que nuevas características pueden mejorar el sesgo y la varianza (si utilizamos regularización).
3. Reducir o eliminar la regularización. Esto reducirá el sesgo evitable, pero aumentará la varianza.
4. Modificar la arquitectura del modelo que puede afectar tanto al sesgo como a la varianza.

En caso de alta varianza, se propone:

1. Agregar más datos de entrenamiento, sería la forma más simple.
2. Agregar regularización, reduce la varianza pero aumenta el sesgo.
3. Agregar parada temprana del entrenamiento. Esta técnica reduce la varianza pero aumenta el sesgo.
4. Selección de características para disminuir el número o el tipo de características de entrada.
5. Disminuir el tamaño del modelo.
6. Modificar las características de entrada según las percepciones del análisis de errores
7. Modificar la arquitectura del modelo

En resumen, según el campo de conocimiento que estemos tratando, se puede hablar de distintos tipos de sesgos, entendidos como error o desviación que puede provocar conclusiones erróneas, si no se trata de reducirlo. De este modo, desde el campo de la estadística hablaríamos de sesgos estadísticos los cuales provendrían de una mala selección de los datos o su tratamiento, desde el campo de las ciencias como la psicología hablaríamos de sesgos cognitivos los cuales se implicarían errores en la toma de decisiones provocadas por mecanismo internos del funcionamiento de la psique humana y que lleva a realizar generalizaciones erróneas de distintas situaciones. Y desde el campo de la IA tendríamos el sesgo algorítmico, provocado como consecuencia de un error en el procedimiento de la propia IA bien sea, debido a sesgos heredados del conjunto de datos que se utilizan, debido al propio algoritmo utilizado, o bien, debido a proceso de entrenamiento de la máquina.

2.4. Sesgos y Errores humanos, íntimamente ligados a la máquina

No podemos olvidar que la IA comete errores, pero, esos errores o sesgos son indisolubles de las personas que están involucradas en todos los pasos de creación de modelos de IA; ya sea en la captura de datos, la elaboración del propio algoritmo, la selección de características a tener en cuenta, el entrenamiento del modelo o incluso su posterior uso. En consecuencia, en todo este proceso, el hombre sería la parte más importante en la generación de errores.

Efectivamente, según RUIZ-MORENO y TRUJILLO (2012), el estudio del error cometido por humanos es más complejo que el de las máquinas, constituyendo aquel una parte muy importante del riesgo total cuando se hace un análisis de probabilidad de sistemas industriales

o técnicos complejos. Ambos autores afirman que según diversos estudios la contribución del hombre al error de un sistema implicaría entre el 50% y el 80% de la responsabilidad. Sin embargo, no todo sería negativo, el ser humano tendría la habilidad para corregir y anticipar las desviaciones no previstas, mostrando una gran adaptación y capacidad de aprendizaje de forma que le permitiera elaborar soluciones a los problemas planteados, anticiparse a ellos y corregir los errores. Habilidades en las que se aventajaría a cualquier dispositivo no humano.

CARRILLO-GUTIÉRREZ et al. (2021) hablan de que la incidencia de defectos en la industria por causa humana está entre el 70% y el 90% de los casos y ALEXANDER (2017), indica que según datos de la NASA, el 78% de los incidentes relacionados con operaciones del transbordador espacial en tierra fueron resultado de error humano. Por otro lado, STANTON et al (2005) confirman las ideas anteriores, es decir, los errores humanos constituirían un porcentaje muy alto en los incidentes en sistemas complejos. Las cifras que presentan son muy variables, por ejemplo, el error humano sería la causa del 50% de los accidentes ferroviarios y hasta 75% en accidentes de tráfico, señalando también, que son causa de un alto número de muertes por error médico. Con estos resultados podríamos concluir que, en todo caso, es el hombre y no la máquina sobre el que habría que implementar la mayor cantidad de herramientas de vigilancia y control.

En relación con los errores humanos, si nos centramos en campo de la investigación judicial, autores como SOLETO (2016), reconocen que determinadas técnicas de investigación basadas en la memoria tienen una alta probabilidad de causar errores. Y GRAJALES ARTURO (2023) señala que la neurociencia ha venido a confirmar que el ser humano comete errores en base a su predisposición de enjuiciar la realidad en función de su cultura, educación u origen social y por lo tanto los sesgos cognitivos de cualquier juez influyen en su decisión.

También, en el ámbito judicial, DROR (2020) habla sobre los sesgos presentes en patólogos forenses y como estos pueden influenciar en sus decisiones cuando las mismas evidencias se les presentan en distintos contextos. Constatando que, a pesar de las pruebas, el proceso de cómo se evalúan tiene una influencia crítica en las conclusiones que se obtienen de ellas. El autor presenta hasta ocho fuentes de sesgos implícitos en el proceso de toma de decisiones.

Para VILLANUEVA (2021), no se puede negar la realidad de los sesgos en jueces y jurados y defiende la necesidad de encontrar mecanismo que los reduzcan.

Dada la importancia que tiene el error humano en la gran cantidad de accidentes o fallos producidos bajo su intervención, se han desarrollado un variado número de métodos que pueden utilizarse para la medición del mismo. Entre ellos se podría citar el que exponen, por ejemplo, RUIZ-MORENO y TRUJILLO (2012) la metodología de Técnica de Evaluación de Riesgos de Pérdida de Fallo Humano, por sus siglas en inglés THERP, como una de las técnicas más características para el cálculo del error humano. También, el expuesto por ALEXANDER (2017) «*Human error assessment and reduction technique*» (HEART), que implicaría la utilización de guías y tablas para evaluar la probabilidad de errores y sus posibles consecuencias. Y por último podemos citar a STANTON et al (2005) que hablan de distintas técnicas según su enfoque cuantitativo o cualitativo. Así, presentarían varios métodos de medición del error como pueden ser el «*Cognitive Reliability and Error Analysis Method*» (CREAM), «*Technique for the Retrospective and Predictive Analysis of Cognitive Errors*» (TRACER) o el «*The systematic human error reduction and prediction approach*» (SHERPA)

A este respecto MEHRABI et al. (2021), señalan que la utilización de los algoritmos de aprendizaje automático aporta mayores ventajas ya que, a diferencia de los humanos, las máquinas no se cansan ni se aburren y pueden trabajar con mayor número de factores que las personas. Y MILLER (2018) afirma que posiblemente el sesgo humano sea más peligroso que el debido a los algoritmos.

En definitiva, el control del error en la IA es en última instancia el control del error humano, sobre el cual existirían muchos estudios en diversos campos. No se trata de medir si un sistema comete más errores que otro, sino de entender que el error es parte indisoluble de la IA ya que, en el proceso de creación de IA, es necesario que el hombre intervenga en algún momento del mismo y por lo tanto, es inevitable la introducción de sesgos. Solo es posible su disminución en un rango que haga viable la utilización de la IA.

2.5. IA en el ámbito judicial y de la investigación criminal.

Coincidimos con JANUÁRIO (2023) que el desarrollo tecnológico y científico supone un gran desafío para los ámbitos jurídico y penal. La aplicación de la IA al campo de la criminología estaría íntimamente relacionada con su aplicación a la resolución de problemas penales, es decir, quedaría incluido dentro del proceso penal y como consecuencia, sería inseparable del ámbito judicial. Cualquiera de las implicaciones que existiera dentro de la criminología tendría

que ser valorada también desde el ámbito judicial, en cuanto a su afectación a los derechos que pudieran verse afectados.

2.5.1. Aplicación a la investigación criminal y al ámbito judicial

La Comisión europea para la eficiencia de la justicia (en adelante CEPEJ) (2018), expone, en su Carta Ética Europea sobre el uso de la inteligencia artificial en Sistemas judiciales y su entorno que, entre las utilidades para la IA en el ámbito de la justicia a nivel administrativo, estaría el tratamiento de archivos o documentos, lo cual permitiría mejorar el servicio de la justicia hacia el ciudadano. Otras de las posibles aplicaciones serían la de acceso y búsqueda de jurisprudencia o la capacidad de relacionar distintas fuentes legales. La IA sería, también, útil para mejorar la transparencia de la propia justicia. Uno de los aspectos más interesantes, desde nuestro punto de vista, sería la utilización de la tecnología para la realización del perfilado de los jueces que les permitiría ser más eficaces a la hora de emitir sus decisiones.

CASANOVAS y BOURCIER (2012), hablan, que dentro del ámbito judicial, existiría la posibilidad de utilizar la IA en tareas de asistencia a la toma de decisiones o incluso en la elaboración de argumentación jurídica. También, reseñan la posibilidad de lectura de las reglas del derecho.

Para BARONA VILAR (2019, p. 24) «Los datos masivos son en el Siglo XXI lo que probablemente fue la energía del vapor en el Siglo XVIII o la electricidad en el Siglo XIX» estos datos deben, sin embargo, ser tratados y analizados a través de distintas técnicas para poder sacar de ellos una verdadera utilidad. Las nuevas tecnologías tienen su cabida, dentro del ámbito de la investigación penal, en el desarrollo de nuevas funciones que permitirían mejorar el descubrimiento de posibles responsables de una acción penal, así como la predicción de riesgo de reincidencia o incluso la veracidad de las declaraciones emitidas por testigos o peritos. El reconocimiento de personas a partir de su rostro, lectura de matrículas de vehículos, geolocalización a través de distintos aparatos electrónicos y el rastreo dentro de la red de internet serían otras de las funcionalidades de la IA aplicada a la investigación de delitos.

Entre los posibles empleos de la IA, según ROUHIINEN (2018), estarían: reconocimiento de imágenes, clasificación, etiquetado, aplicación de algoritmos en el sector financiero, utilización para el tratamiento de datos en cuestiones médicas, predicción en distintos sectores de la vida humana y su utilización en cuestiones de seguridad. En toda esta nueva

tecnología, la importancia radicaría en los datos, los cuales a través de sus análisis permitiría el funcionamiento adecuado de las IA.

En relación con los sistemas de investigación de delitos CUATRECASAS MONFORTE (2023), establece de la necesidad de clasificar dichas herramientas en tres tipos: aquellas dedicadas a la predicción y evaluación de riesgos; aquellas cuyo objetivo sería la investigación de delitos; y por último, las herramientas de tramitación.

Otro autor, MONTESINOS GARCÍA (2021), habla principalmente del carácter preventivo de las aplicaciones de la IA, resaltando la nueva era de la prevención y seguridad. De este modo, cuando habla de las aplicaciones de la IA las centra en dicho campo predictivo. Considerando el lado policial y el lado judicial. Respecto del primero indica la existencia de cuatro métodos: los encaminados a predecir el crimen; los que buscan el cálculo de probabilidades de que una persona cometa el crimen; aquellos que predicen las características a las que es más probable que se ajuste un autor de un determinado delito; y por último, aquellos métodos que predicen el riesgo de ser víctima. A su vez, todos estos métodos estarían basados en la predicción de lugares y en la predicción de individuos. Los primeros buscan la elaboración de mapas donde exista mayor probabilidad de aparición de hechos delictivos. Lo que podría presentar problemas de estigmatización de determinadas zonas geográficas con consecuencias hacia sus moradores, estén o no relacionados con la delincuencia. Los métodos basados en personas analizan sus características para arrojar una predicción acerca de la posibilidad de delinquir.

Algunas herramientas de lo indicado hasta ahora serían: Precobs, XLAW, Eurocop PRED-CRIME (en desarrollo), PredPol o HunchLab para las herramientas basadas en lugares y el algoritmo de ROSSMO, sistema VioGen para los centrados en la predicción de los individuos.

En relación a los sistemas de apoyo a los jueces MONTESINOS GARCÍA (2021) cita aquellos que permiten calcular el riesgo de reincidencia como el sistema COMPAS, el HART, el VioGen, o el RisCanvi. Y VALLS PRIETO (2017), habla de los diversos problemas jurídicos que presentan las técnicas de prevención y persecución del crimen. Y respecto de la justicia, SIMÓN CASTELLANO (2021), destaca, entre otras, que una de las utilidades sería la aplicación de la IA a la resolución extrajudicial o el perfilado de jueces. Este autor basándose en GONZÁLEZ FUSTER (2020), divide en dos bloques la aplicación de la IA en la jurisdicción penal. Estos serían: el bloque predictivo y el de reconocimiento facial. El predictivo buscaría predecir delitos, localizando lugares, tiempos o incluso personas que estén en predisposición de cometerlos o de sufrirlos.

Por otro lado, GONZÁLEZ FUSTER (2020) considera principalmente cuatro bloques relacionados con la IA en el campo penal: el de policía predictiva, el reconocimiento facial, IA y justicia criminal, y IA y fronteras. El campo de policía predictiva a su vez se descompondría en cuatro ámbitos o métodos: aquellos que pretenden predecir delitos o pronosticar espacios o tiempos con mayor probabilidad de delito; los que pretenden identificar individuos con riesgos de cometer o reincidir en delitos; otro ámbito sería el que pretende predecir perfiles en función de delincuentes anteriores; y finalmente el que busca predecir víctimas a través de la identificación de individuos o grupos en riesgo de victimización.

El CCBE (*Council Of Bars And Law Societies Of Europe*) (2020), en su informe sobre consideraciones de los aspectos legales de la IA, distribuye el uso de la IA en el campo policial y jurídico penal, en dos sectores: el de prevención de delitos, realizando un uso de la IA como herramienta de predicción; y el segundo sector, como recopilación y análisis de las pruebas. El uso de herramientas predictivas de IA, con el objetivo de generar predicciones sobre nuevos actos delictivos y su distribución espaciotemporal, se utilizaría para distinguir y examinar patrones de sucesos de carácter pena. Según este informe los algoritmos utilizados en esta área utilizarían datos sobre demografía, áreas delictivas, ubicaciones probables de delincuentes, etc. Estos algoritmos estarían basados en teorías como la de la actividad rutinaria de Cohen y Felson formuladas en 1979. Sin embargo, la utilización de dichos algoritmos no son del todo aceptados, fomentando profundas críticas debido a la base de datos sobre los que realizan sus predicciones, ya que estas incluirían delitos que se distribuirían con mayor frecuencia en zonas pobladas por individuos de menores recursos y donde existiría mayor concentración de grupos étnicos minoritarios. Con estos datos el posible riesgo de sesgo estaría servido, potenciando la criminalización y el etiquetamiento de una parte de la sociedad, en contraposición de la otra parte, por el mero hecho del espacio físico donde viven.

Desde su punto de vista BARONA VILAR (2021a) resalta la gran aceptación de la tecnología de IA en la predicción policial dentro del crecimiento de la cultura de la sociedad de la seguridad o del riesgo de Ulrich Beck. Este fenómeno ha permitido la creación de herramientas de predicción que trabajan con gran cantidad de datos, a partir de los cuales se construye una predicción sobre la criminalidad.

La autora referencia herramientas de asistencia en el ámbito jurídico, que permite elaborar estrategias de defensa/acusación o predicen el comportamiento de la defensa, tales como

ECHO y ALIBI. En el campo de la mediación o negociación cita herramientas como MODRIA, MEDIARÉ o iCAN SYSTEM y como ejemplos de simuladores de abogados señala DONOTPAY, REEM Y SPLIT-UP. En el campo policial, habla de los programas COMPAS y HART como herramientas de predicción de riesgo futuro de comisión de delitos por parte de una persona. Y para predicción geográfica, menciona el CAS. VERIPOL y VioGen serían herramientas utilizadas en nuestro país, para la detección de denuncias falsas y valoración de riesgo por violencia de género. Por último, hace mención al campo del reconocimiento biométrico, donde se plantea la preocupación de caer en un Gran Hermano a nivel mundial. Concluyendo que, es fundamental poner límites al uso de la tecnología aplicada a la vigilancia preventiva, por las posibles consecuencias negativas que pudiera derivarse de la misma.

GÓMEZ COLOMER (2023) plantea, por otro lado, la existencia de otras aplicaciones de la IA en la predicción, como puede ser el estudio de sentencias dictadas por un juez o tribunal que ayudaría a los abogados en el planteamiento de sus estrategias. Otra posibilidad sería la de utilizar la máquina como juez con capacidad para juzgar determinados asuntos tanto a nivel civil como penal al objeto de reducir el número de asuntos que colapsan la justicia hoy en día. Respecto a esto último resalta que dicha idea ya se estaría aplicando en países como Estonia (desde el año 2000) o China (desde el año 2019). Los riesgos de estas tecnologías radicarían en un aumento de la capacidad inquisitorial pudiendo potenciar riesgos de discriminación ataque a derechos de privacidad o incluso de la dignidad o libertad de expresión.

Otros autores, como PERALTA GUTIÉRREZ y TORRES LÓPEZ (2022) exponen distintos instrumentos utilizados en el ámbito jurídico. De ellos destacamos: los relacionados con la gestión de la administración de justicia como Adriano, Amberlo, eMatt, CourtView o Jworks entre otros; herramientas de análisis de documentos como Ross Intelligence o Ravn i Manage IA; Chatbots DONOTPAY que permitirían la realización de reclamaciones; y aquellos que permitirían análisis y búsqueda de textos legales y jurisprudencia (justicia predictiva) como Judicata, Lexmachina-Legal Analytics, Ravel Law, Jurimetría, Legal Data, Vlex o Gavelytics.

A nivel institucional, dentro del ámbito europeo, podemos citar la Resolución P9_TA (2021) 0405 del PARLAMENTO EUROPEO (2021), donde destaca que la IA proporciona una mayor eficacia en la lucha de delitos de especial gravedad como terrorismo o el de blanqueo de capitales. Respecto al uso por parte de los cuerpos de policía señala varias utilidades, como son: reconocimiento facial o de otras características humanas (voz, otras señales biométricas);

reconocimiento automático de matrículas; tecnologías de lectura de labios; vigilancia auditiva; investigación y análisis autónomos de bases de datos; predicción como por ejemplo puntos calientes de delincuencia; detección del comportamiento; detección de fraudes financieros y la financiación del terrorismo; vigilancia de las redes sociales; captadores de identidad internacional de abonados móviles; otros sistemas de vigilancia como cámaras térmicas o detección de latidos del corazón.

Según la citada Resolución el empleo de la IA por las autoridades judiciales sería: apoyo de decisiones de prisión o para dictar sentencias; resolución de litigios en línea; gestión de jurisprudencia y acceso a la justicia.

El texto europeo afirma que los riesgos en el ámbito judicial y policial se ven acrecentados, ya que afectan a derechos fundamentales de los individuos como: la presunción de inocencia, de libertad y seguridad de la persona, los derechos a una tutela judicial efectiva y a un juicio justo, a la no discriminación, libertad de circulación, libertad de reunión y asociación o libertad de expresión e información.

En resumen tratando de recopilar la información expuesta hasta el momento, podríamos decir que los distintos campos de aplicación de la IA con utilidad directa en la investigación criminal serían:

1. Reconocimiento Facial ayudando en la identificación personas en imágenes y vídeos, facilitando la identificación de sospechosos o víctimas en investigaciones criminales.
2. Análisis de Textos mediante procesamiento de lenguaje natural con el fin de analizar textos de todo tipo para la obtención de información útil en la investigación.
3. Predicción del Crimen, a través de predicciones de algoritmos que arrojen áreas con mayor probabilidad de comisión de hechos delictivos.
4. Sistemas de Perfilación Criminal para la elaboración de perfiles criminales a partir de patrones y características de otros delincuentes y delitos cometidos.
5. Análisis de Redes Sociales y detección del fraude a través de la búsqueda de patrones y relaciones y detección de conexiones o transacciones determinadas o sospechosas.
6. Simulación y Reconstrucción de Escenas del Crimen a partir de la introducción de los datos recogidos en las mismas.
7. Automatización de Tareas Repetitivas evitando a los investigadores tareas arduas y fatigosas que les distraigan de su función fundamental.

Respecto de las distintas utilidades y herramientas expuestas, estas se podrían sintetizar, sin ánimo de ser exhaustivos en la siguiente tabla:

Tabla 1. Resumen de las distintas utilidades de la IA y herramientas

UTILIDADES		
Dedicadas a predicción	Dedicadas a investigación	Dedicadas a la tramitación
Predicción de riesgo de reincidencia.	Elaboración de posibles sospechosos perfilado criminal.	Control de tiempos y plazos.
Predicción de zonas calientes del delito.	Perfilado geográfico.	Control de medias cautelares, competencia,
Predicción de potenciales delincuentes.	Reconocimiento facial	Elaboración de defensa o hipótesis de acusación.
Predicción de potenciales víctimas	Reconocimiento de voz	Mediación y Resolución de conflictos.
Predicción de denuncias falsas.	Reconocimiento de escritura, análisis de textos, etc.	Multas.
Justicia predictiva.	Verosimilitud del testimonio.	Análisis de documentos. Ross
Dactiloscopia y ADN		
Rastreo de redes		
HERRAMIENTAS		
COMPAS, HART, VioGen, RisCanvi, ROSSMO, PredPol, Adriano, eMatt, CourtView, Jworks, CAS, Precobs, XLAW, Eurocop PRED-CRIME, HunchLab, ECHO, ALIBI MODRIA, MEDIARÉ, REEM iCAN SYSTEM, DONOTPAY, SPLIT-UP, Judicata, Lexmachina-Legal Analytistics, Ravel Law, Jurimetría, Legal Data, Vlex Gavelystics, Intelligence, Ravn i Manage IA , VERIPOL		

Fuente: elaboración propia a partir de datos expuestos en marco teórico

2.5.2. Policía predictiva

PERRY et al. (2013, p. 13) definen como policía predictiva «la aplicación de técnicas analíticas, especialmente técnicas cuantitativas, para identificar objetivos probables para la intervención policial y prevenir delitos o resolver crímenes pasados mediante predicciones estadísticas.»

esta definición es tomada por diverso cuando se refieren a la policía predictiva. Estos autores además, desmitifican la policía predictiva como el remedio supremo de la lucha contra el delito. A su vez llaman la atención sobre las trampas en las que se puede caer como la de pasar por alto derechos de los ciudadanos.

MIRÓ LLINARES (2020 a), reconoce la inexistencia de tecnologías que sean neutras o asépticas. La IA puede ser buena o mala, dependiendo de los fines que persiga. Incluso, si el fin fuera bueno, su aplicación al campo judicial y policial podría conllevar consecuencias que no deseadas. En su estudio habla, por un lado de la justicia predictiva penal y por otro de policía predictiva. Justicia predictiva sería aquella que emplea algoritmos para realizar una estimación de riesgos futuros, a partir de introducir valores sobre las características personales y ambientales de los sujetos. Para la policía predictiva utiliza la definición de PERRY et al. (2013) e indica que el objetivo de la misma estaría en la predicción de delitos y la identificación de objetivos, ya sean personas o espacios geográficos, sobre los que dirigir al acción policial.

Según el autor existirían dos tipos de técnicas de predicción policial: aquellas que se basan en los entornos geográficos y otras basadas en las características de los delincuentes. Cada técnica tendría consecuencias distintas, siendo las segundas las que más polémica habrían generado por sus implicaciones éticas y posibilidades de sesgos discriminatorios. También existirían otra serie de técnicas predictivas basadas en análisis de imágenes como el reconocimiento facial o de movimientos o lectura de placas.

CINELLI y MANRIQUE (2019) destaca que la utilidad de los sistemas de análisis predictivo en el ámbito policial todavía no es muy clara teniendo en cuenta la relación coste-beneficio. Entre los aspectos positivos se encontrarían la posibilidad de optimización de recursos policiales, así como la identificación de patrones que permitirían la anticipación ante hechos delictivos (acción preventiva frente a reactiva). Respecto a los aspectos negativos, se podría hablar de aquellos relacionados con los derechos fundamentales afectados, tales como la presunción de inocencia, privacidad o incluso discriminación. Entre las herramientas de este ámbito señala: Analyst's Notebook, KeyCrime, Predpol, Crime Anticipation System, Precobs y Skala.

GONZÁLEZ-ÁLVAREZ, HERMOSO y CAMACHO-COLLADOS (2020) entienden el concepto de policía predictiva como el uso de gran cantidad de datos que serían tratados mediante algoritmos, al objeto de poder realizar predicciones que permitieran una gestión más eficaz de los recursos policiales. Se pasaría por lo tanto, de una acción reactiva a una preventiva por

parte de las policías. Estas herramientas funcionarían buscando patrones dentro de una gran cantidad de datos. La complejidad de estos sistemas hace necesaria la colaboración con otros estamentos diferentes de los policiales. La aplicación de estos sistemas plantea problemas como la falta de transparencia y la posibilidad de sesgo si los datos que se administran aportan ese error. En España los principales campos de utilización de estas técnicas son: predicción de crímenes, delimitación de distritos policiales y optimización de zonas de patrullas, análisis y vinculación de crímenes en serie, estimación del riesgo en violencia de género, vulnerabilidades de las redes de delincuencia organizada, imágenes de cámaras policiales, detección de denuncias falsas.

NÓHPAL RODRÍGUEZ (2021) en la misma línea define la policía predictiva como un modelo policial donde se emplearían distintas técnicas para analizar la información recibida al objeto de predecir un delito. Incide, también, sobre la idea del paso de una policía de reacción a otra de prevención. Resalta la necesidad de que el equipo de policía se encuentre formado por un amplio elenco de profesionales (criminólogos, criminalistas, victimólogos, psicólogos, etc.). Entre las herramientas actuales que ya se utilizan en este campo el autor cita las herramientas Dagnet, Rigel, CrimeStat, y Predator.

Por último indicar que, en relación a las herramientas utilizadas en la policía predictiva autores como LUM e ISAAC (2016) señalaban que los datos policiales con los que se alimentan las mismas eran el principal problema al ser datos sesgados desde su origen.

2.5.3. Problemática de la aplicación de la IA al ámbito de la investigación criminal.

Desde la criminología se han desarrollado distintas teorías sobre el delito que han ido teniendo su aplicación en la ciencia policial en diversas formas y distintos grados de asimilación. Estas teorías, a su vez, han sido implementadas en algoritmos que han permitido el desarrollo de distintos instrumentos, dando paso a la introducción de la IA en el ámbito criminológico aplicado a la investigación, resolución de casos y explicación de fenómenos criminales. Todo ello no solo a nivel policial sino también a nivel judicial, como por ejemplo, el análisis de la previsión de la violencia o de la reincidencia para unos determinados individuos que ya han delinquido. Esta irrupción, como veremos a continuación, plantea distintos problemas y cuestiones sobre su uso y riesgos.

Para MIRÓ LLINARES (2020a), es difícil mantenerse en el término medio entre una visión negativa o catastrofista y una visión optimista o utópica de la aplicación de técnicas de policía predictiva. La respuesta estaría en ver que la tecnología no supone una herramienta neutral. Reconocer las limitaciones de la IA podría ser un paso para mejorar la misma orientándola hacia objetivos más realistas. En este sentido un enfoque multidisciplinar e integral del desarrollo de estas tecnologías se hace necesario.

Según CUATRECASAS MONFORTE (2023, p. 63) el potencial de magnificación de la IA podría conllevar, en relación con la ineficiencia del hombre, a «perpetuarla, lo cual resulta extremadamente peligroso, especialmente en el ámbito de la justicia». Aunque, la IA es todavía, muy poco aplicada en el entorno de la justicia, podría aportar grandes beneficios en la mejora de la calidad, minimizando los riesgos con una adecuada regulación. Esta regulación debería de cumplir una serie de principios mínimos, al objeto de garantizar su eficacia.

En su libro CASTELLANO y MAGRO SERVET (2021) destacan como problemas y riesgos principales de la IA aplicada a la justicia: la opacidad de los códigos, el riesgo de discriminación o la falta de legislación adecuada entre otros.

La falta de transparencia en los algoritmos es resaltada por SIMÓN CASTELLANO (2021). A este problema, se le añadiría que la legislación de aplicación en la actualidad, a nivel europeo, es compleja y se encuentra muy disgregada en distintas normas, que no contribuyen a clarificar la situación y permiten la aplicación de distintos criterios heterogéneos en los países de la Unión. Para el autor son destacable los retos que se plantean en el ámbito de la protección de los derechos fundamentales, ya que las decisiones basadas en las predicciones producidas por la IA, pueden implicar afectación de los mismos por el mero hecho de ser decisiones preventivas. Derechos que podrían verse afectados serían el derecho a la igualdad, el de defensa o a un juicio justo. Como ejemplo de esos peligros a los que nos referimos el autor expone los casos de los programas COMPAS aplicado en EE.UU y el de HART aplicado en Reino Unido, los cuales inciden en la discriminación de determinados colectivos. En sus conclusiones expone que el uso de IA en el ámbito judicial aporta tantos o más riesgos que beneficios. Aunque reconoce que la aplicación de esta tecnología al campo referido todavía es muy limitada y aplicada para casos concretos.

BARONA VILAR (2019) subraya la preocupación y resistencia existente a los nuevos cambios que se pueden prever con la revolución tecnológica y se pregunta si en el ámbito de la justicia

puede convertirse en una solución o, en cambio, será un problema. Destaca el peligro que pueden suponer errores en la predicción realizada por los sistemas algorítmicos indicando el caso real referido al *US Sentencing Guidelines (U.S.S.G.)*, aplicado a Estados Unidos.

También, en BARONA VILAR (2021b), se cuestiona la infalibilidad de la utilización de sistemas algorítmicos en el ámbito judicial. Siendo fundamental preguntarse acerca de los distintos parámetros que influyen en los mismos, es decir, qué datos se eligen, que patrones se escogen, etc.

Autores como VÉLEZ et al. (2022) plantean que, quizás uno de los principales problemas en la utilización de la IA sea la falta de transparencia de los procesos que se realizan dentro de la máquina. Tampoco deberíamos olvidar que la seguridad, o más bien la falta de ella, en estos sistemas podría acarrear serias consecuencias como resultado de ataques o manipulaciones de los mismos.

Y VALLS PRIETO (2017), muestra la preocupación por la utilización de los datos necesarios y su procesamiento para el funcionamiento de la IA de forma que cause grave perjuicio a derechos fundamentales de los ciudadanos. Esa preocupación se extiende al riesgo de utilizar los datos para la condena judicial de dichas personas, en base una posible predicción del suceso arrojada por una máquina que utilice la IA. Como ejemplo de ello expone el hecho del programa aplicado por Reino Unido en 2008, que pretendía actuar antes de que se hubiera producido el acto delictivo en jóvenes, basándose en aspectos tales como el coeficiente intelectual o resultados académico, lo que supuso un fracaso por implicar estigmatización.

Este autor presenta tres problemas, la utilización de datos o información personal de forma no controlada, la utilización en la justicia como método para adelantar la barrera punitiva a solo la previsión de posibles actos delictivos y la confianza excesiva en los resultados obtenidos sin control o análisis de los mismos.

La realidad de las decisiones y resultados obtenidos por un procedimiento de IA, como un elemento poco transparente se ve reforzada con el hecho de que, como bien indica MERCADER UGUINA (2022), la mayoría de los algoritmos utilizados en este campo pertenecen o son desarrollados por empresas privadas, lo que permite en nuestro país, la aplicación de la Ley 1/2019, de 20 de febrero, de Secretos Empresariales, donde se establecería la salvaguarda de aquellos conocimientos o información incluso de carácter tecnológico que reúna tres

condiciones: que esa información sea secreta (refiriéndose a que no sea conocido por la generalidad de las personas ni sea fácilmente accesible), que el empresario haya tomado medidas para que sea secreta y que dicho conocimiento o información tenga un valor empresarial potencial o real como consecuencia de que posee la cualidad de secreto. Esta realidad dificulta todavía más el hecho de conocer como es el proceso de toma de decisiones de cualquier sistema de IA y por lo tanto conocer el porqué de la decisión que es seleccionada por la máquina.

Asimismo, MIRÓ LLINARES (2020 b) en su trabajo identifica distintos problemas derivados de la aplicación de la IA al campo penal, entre ellos, destaca que el aumento de la complejidad de los sistemas de IA impide explicar el porqué de los resultados obtenidos. Para el autor, también sería problemática la falta de intervención en evaluación de los datos que determinadas herramientas como el famoso COMPAS, presentan, la cual analiza y evalúa los datos sin intervención humana. De hecho, la automatización de las valoraciones de riesgo dentro del mundo judicial supondría la existencia de un fuerte debate ético-jurídico.

Otra gran dificultad sería la que se refiere a la obtención de un gran volumen de datos para el entrenamiento de la IA. En estos casos el problema se plantearía en el conflicto entre derechos fundamentales, como el de privacidad, y la necesidad de mejorar la capacidad de predicción, aumentando los datos de entrenamiento. A este problema habría que añadirle la posibilidad de la existencia de sesgos desde el primer momento en la recogida de datos y que condicionan la respuesta de la IA.

La urgente necesidad de responder ante la fractura cada vez mayor entre la IA y la legislación, es la idea que introducen ALMACHE BARREIRO y ALBERT MÁRQUEZ (2023), debido a la oscuridad que impera dentro de los procesos de IA que podrían suponer un serio daño relativo a la discriminación. Los autores auguran que estas nuevas tecnologías causarán un fuerte impacto en el ámbito jurídico. Cualquier solución normativa debería pasar por una regulación lo más amplia posible a nivel internacional para que tuviera verdadera eficacia siendo muy interesante la creación de algún órgano o la capacitación de alguno de los ya existentes con funciones rectoras y de regulación en el ámbito de la IA.

2.5.3.1. Responsabilidad de los errores de la inteligencia artificial

Respecto a la posible responsabilidad judicial sobre las decisiones dañinas de la IA LLEDÓ (2022), no tiene duda, es absurdo plantearse calificar a un máquina como responsable de un delito debido a su falta de elementos básicos de imputabilidad, esto es, el elemento volitivo y el intelectual. Por lo que el responsable debería ser en todo caso, hoy en día e incluso en el futuro, el responsable humano encargado de su programación, fabricación o manejo.

Otra idea que cabe resaltar es la posibilidad, planteada por el mismo autor, de que la justicia sea en un futuro cercano, por máquinas que incluso sustituyan a los humanos, planteándose la disyuntiva de quien sería el responsable en caso de errar en su decisión. El peligro es bien cierto, toda vez que los prejuicios de los propios humanos se ven trasladados hacia los algoritmos que entrenan la IA, lo que supone un serio peligro de perpetuación de las discriminaciones y sesgos de la sociedad actual.

Sin embargo, la imputación de responsabilidad penal en casos en los que está implicada la IA no es un tarea sencilla, tal y como indica JANUÁRIO (2023). El entorno donde se lleva a cabo esta actividad es altamente complejo donde intervienen numerosos agentes, tanto personas físicas como jurídicas, ya sea en la fabricación, programación, desarrollo, etc. Lo que no contribuye a facilitar la tarea de atribuciones de responsabilidad. En el mismo sentido MIRÓ LLINARES (2020 b) Si bien se está todavía muy lejos de una IA totalmente autónoma capaz de generar responsabilidad por sus propias acciones, el autor, indica que la situación actual de la IA es suficientemente importante como para plantear ya problemas de responsabilidad cuando existe un error en las predicciones arrojadas por los algoritmos. Aunque MIRO LLINARES lo ve claro, no es todavía necesario un cambio en la atribución de responsabilidad desde el humano a la máquina, puesto que es el primero quien alimenta y gestiona las instrucciones que tiene que seguir la segunda.

2.5.3.2. Protección de datos

La necesidad de capturar una enorme cantidad de datos para su utilización tanto en el tratamiento con técnicas de *Big Data* como en el caso de su utilización para el desarrollo de IA o su aplicación dentro de nuestra sociedad, hace que esos datos tengan un importante valor que a su vez podrían ser objeto de la generación de hechos delictivos, es decir, que esos datos deberían ser protegidos, siendo este hecho fundamental a la hora del desarrollo de la IA, no

solo por los derechos reconocidos por las leyes a todos los ciudadanos, respecto del tratamiento de sus datos, sino también por el hecho de que si dichos datos no tuvieran la debida protección y el acceso a ellos pudiera darse por cualquier usuario estos podrían ser modificados o falsificados de forma que su uso podría implicar graves consecuencias en relación a los resultados y conclusiones que se obtendrían de su utilización dentro de los sistemas de IA.

Entre los problemas que la IA trae consigo MIRÓ LLINARES (2020 b) destaca como, como ya hemos expuesto anteriormente, el conflicto entre los derechos a la intimidad y la necesidad de datos de entrenamiento de la IA para su aprendizaje que serían tomados en muchas ocasiones de forma que el afectado no sería consciente de su cesión. Lo que plantearía, según el autor, un repunte de la preocupación, de parte de las administraciones públicas, por la necesidad de proteger la intimidad y la privacidad. Aunque como señala esta preocupación bien podría ser solo una máscara dada la legislación aprobada a nivel europeo (Reglamento (UE) 2016/679) donde se reglamentaría excepciones a estos derechos cuando los fines de las autoridades estuvieran dentro de las causas previstas. Efectivamente, en el artículo 2.2 de dicha directiva se indica cuáles son las excepciones donde no se aplica al tratamiento de datos personales, exponiendo es su subapartado d. «por parte de las autoridades competentes con fines de prevención, investigación, detección o enjuiciamiento de infracciones penales, o de ejecución de sanciones penales, incluida la de protección frente a amenazas a la seguridad pública y su prevención». También se encontraría la excepción en la Ley Orgánica 13/2015, de 5 de octubre, de modificación de la Ley de Enjuiciamiento Criminal en atención a nuevas formas delictivas, por lo que para realizar su persecución se permitiría por ejemplo el registro remoto de equipos informáticos.

Otros autores como MARTÍNEZ DEVIA (2019) también resaltan que las aplicaciones de IA copan cada vez más distintas áreas de la vida lo que supone la necesidad de reflexionar sobre sus implicaciones a nivel de los derechos fundamentales de las personas, concretamente su afectación sobre la protección de la privacidad, los datos personales y su tratamiento.

En la misma línea HUESO (2019) constata que es indudable la importancia de la IA, el *Big Data* y la robótica en nuestros días, lo que plantea serios retos relacionados con la discriminación o transparencia.

Podemos por tanto concluir que la propia naturaleza de la IA actual implica la necesidad de acceder a grandes cantidades de datos. La forma en que se consiguen dichos datos lleva implícita la necesidad de respetar derechos fundamentales relacionados con la privacidad y la vida íntima de las personas. En consecuencia la IA no puede disociarse del campo sobre la protección de datos.

2.6. Implicaciones ético-morales

Para MONTESINOS GARCÍA (2021) la cuestión ética, sobre la aplicación de la IA, sería ineludible. Esta versaría sobre si es o no ético adoptar medidas sobre aquellas personas que, aún no han delinquido, pero una herramienta predictiva ha indicado que existen altas probabilidades de que lo hagan. En definitiva, sería la extensión de la barrera penal a los actos previos o incluso internos del individuo.

En su Recomendación sobre la ética de la inteligencia artificial, adoptada el 23 de noviembre de 2021 la UNESCO (2022) reconoce que la tecnología por IA, tiene consecuencias negativas y puede contribuir a las divisiones y desigualdades presentes en el mundo. Sin embargo, también reconoce los beneficios de las mismas, así como, la necesidad de la elaboración de unas normas éticas que sean aceptadas a nivel mundial y que proporcionen una guía orientadora hacia como debe de desarrollarse la IA. Por lo tanto, con la elaboración de dicha recomendación la UNESCO (2022, p.14), pretende proporcionar «un marco universal de valores, principios y acciones para orientar a los Estados en la formulación de sus leyes».

Como valores la UNESCO propone los siguientes:

1. Respeto, protección y promoción de los derechos humanos, las libertades fundamentales y la dignidad humana
2. Prosperidad del medio ambiente y los ecosistemas
3. Garantizar la diversidad y la inclusión
4. Vivir en sociedades pacíficas, justas e interconectadas

Como principios se propone en sus recomendaciones:

1. Proporcionalidad e inocuidad
2. Seguridad y protección
3. Equidad y no discriminación
4. Sostenibilidad

5. Derecho a la intimidad y protección de datos
6. Supervisión y decisión humanas
7. Transparencia y explicabilidad
8. Responsabilidad y rendición de cuentas
9. Sensibilización y educación
10. Gobernanza y colaboración adaptativas y de múltiples partes interesadas

El propio PARLAMENTO EUROPEO (2017) en la ya citada resolución P8_TA (2017) 0076 resalta la importancia de cumplir estrictamente con unas normas éticas estrictas en la recogida y tratamiento de datos. Y la COMISIÓN EUROPEA (2020) señala, en su Libro blanco sobre la inteligencia artificial, que solo a través de una participación adecuada de las personas se puede alcanzar el objetivo de una IA ética y fiable.

En el mismo ámbito, el PARLAMENTO EUROPEO (2019), en la Resolución P8_TA (2019) 0081, pide la creación de una carta de principios éticos que deberán seguir expertos y empresas. Además, se indica que la tecnología deberá estar centrada en el ser humano, siendo necesario que los valores éticos estén incorporados desde el primer momento de diseño de la tecnología. Lo que implica imponer límites a la autonomía de la IA y la robótica, así como, garantizar la transparencia, la minimización del sesgo, la discriminación y la posibilidad de explicación de los algoritmos.

FERNÁNDEZ-FERNÁNDEZ (2021) resalta la idea del conflicto entre lo técnicamente posible y lo éticamente deseable incidiendo también, en el peligro que puede suponer la IA a la hora de acrecentar las diferencias y por lo tanto discriminación entre los propios humanos, ya que los algoritmos nunca son algo neutro. Lo que a vista del autor hace necesario la elaboración de un estudio ético de la cuestión a todos los niveles de la sociedad, incluyendo a profesionales de todas las disciplinas e incluso a la propia sociedad civil. Siendo, para dicho autor, esencial y urgente el desarrollo del aspecto ético de esta nueva tecnología. Habla también que, en la actualidad, existen muchas iniciativas con el objeto de elaborar unas guías éticas que superen los posibles problemas derivados de la aplicación de IA, tanto desde el ámbito público como desde el privado. Los principios morales que propone serían: la no maleficencia, la beneficencia, la autonomía, la justicia, la transparencia, la explicabilidad y la responsabilidad. Por otro lado, en España, podemos citar el Informe de 2023 sobre los avances en la Estrategia Nacional de Inteligencia Artificial (en adelante ENIA) elaborado por el gobierno de España. El

gobierno de España (2023), en la ENIA, apuesta por la elaboración un marco ético y legislativo que fortalezca el amparo de los derechos individuales y colectivos, asignando un presupuesto de 4,3 millones de euros para la consecución de tal fin. Este desarrollo prevé tres ámbitos de actuación: el jurídico, al objeto de proteger derechos y elaborar la legislación necesaria; el socio-tecnológico persiguiendo la elaboración de metodologías y estándares de los servicios automatizados; y finalmente el ámbito ético que pretender dar cobertura al uso de la IA respetando el bienestar y la inclusión de la sociedad.

Es interesante también, citar que en el informe ENIA del gobierno de España (2020) se establecían los principios éticos fundamentales, propugnando tres medidas principales que permitan la implantación los principios éticos. Estas serían: la supervisión humana, indicando que la IA debe ser sometida a supervisión humana y además existe la obligación de que sea comprensible; el gobierno de los datos y sistemas en relación con la necesidad de que los datos no deben ser utilizados de forma perjudicial hacia la sociedad o hacia la conculcación de derechos fundamentales; y finalmente la transparencia en el sentido de que se debe de poder permitir la trazabilidad de la IA, es decir, que las decisiones que se obtienen de los algoritmos utilizados deben poder ser auditados, evaluados y explicados por los responsables de los mismos.

En el mismo documento citado, se hace mención a la importancia de tener en cuenta el sesgo de los algoritmos, que a través de entrenamientos con datos que ya aportan un sesgo (falta de representatividad) pueden afectar seriamente a las conclusiones obtenidas. Lo que tiene serias implicaciones éticas en su utilización. Los principios éticos a tener en cuenta serían: la inclusión y no discriminación, el bienestar social contribuyendo al bien común y finalmente, la sostenibilidad en el ámbito ecológico.

En relación con los principios éticos CORTINA ORTS (2019), establece como fundamental para una ética de la IA, la consecución de una IA confiable donde el hombre se encuentre en el centro de la ecuación, es decir, que la nueva tecnología debería buscar la mejora de la vida y la naturaleza de las personas. Debiendo por ello asegurar que la IA está, en todo momento bajo control de la inteligencia humana. Bajo esta visión los principios éticos sobre los que debería basarse la IA, serían: la explicabilidad y rendición de cuentas, la beneficencia, la no maleficencia, el de autonomía de las personas y el de justicia.

En este campo de la ética, JOBIN, IENCA y VAYENA (2019) realizaron una recopilación de los principios éticos que deberían observarse en el desarrollo de la IA citados en numerosos trabajos. Estos principios se resumirían en la siguiente tabla:

Tabla 2. *Resumen principios éticos aplicables a la IA*

Principios éticos	Explicación principal
Transparencia	Entendida como un esfuerzo de dar mayor explicabilidad de las decisiones de la IA, al objeto de minimizar posibles daños.
Justicia y equidad	Entendida como respeto a la diversidad y derecho a reparación de posibles daños. Al objeto de evitar daños a derechos y mitigación de discriminación debido a sesgos.
No maleficencia	Como evitación de cualquier daño o riesgo de daño.
Responsabilidad	Como necesidad de determinar de forma clara y explícita quien debe ser el responsable de los daños o errores cometidos por la IA.
Privacidad	En tanto que la IA se basa en el tratamiento de datos, estos deben ser adecuadamente protegidos.
Beneficencia	Principalmente entendida como desarrollo del bienestar humano.
Libertad y autonomía	
Confianza	En la propia IA como confiable aunque no en exceso.
Sostenibilidad	Referida al medio ambiente y a la eficiencia energética.
Dignidad	Como derecho humano que no debe ser disminuido como consecuencia de la utilización de la IA.
Solidaridad	En relación con el mercado laboral, de forma que la IA no amenace la cohesión social y no se ponga en peligro los grupos más vulnerables.

Fuente: elaboración propia a partir de datos expuestos por JOBIN, IENCA y VAYENA (2019).

2.7. Presente y futuro del marco legal

Creemos que no hay duda de que, respecto a la IA, estamos en una etapa de transición en la que comienza a verse todo el potencial de la nueva tecnología. Es un momento clave, donde a través de distintos organismos nacionales e internacionales se pretende dar un marco adecuado que permita explotar todos sus beneficios reduciendo los peligros. En este contexto, en Europa, tenemos una nueva ley en ciernes, la Ley de IA de la UE, sobre la cual, el 9 de diciembre de 2023, el Parlamento llegó a un acuerdo con el Consejo para su aprobación definitiva. Terminados los trámites formales se convertirá en el nuevo marco legal de la UE. En consecuencia, nuestro presente está muy próximo a cambiar respecto al marco legal que regule todas las actividades con IA.

2.7.1. Presente

La importancia del cambio que estamos viviendo es recogida por organismos como la COMISIÓN EUROPEA (2020) en su Libro blanco, donde reconoce que la IA cambiará la vida de las personas con la mejora de muchos aspectos pero, también, supondrá un riesgo potencial debido a la posible discriminación provocada por estos sistemas, la afectación a la intimidad de las personas, nuevas formas de delincuencia y opacidad en la toma de decisiones y emisión de resultados. Es por ello necesario, que la legislación vele por el cumplimiento de las leyes en la UE, especialmente aquellas que afecta a los derechos fundamentales. De este modo se habla de generar un «ecosistema de confianza» (COMISIÓN EUROPEA, 2020, p.3) solo posible si se atiende a la colaboración entre Estados miembros, además de centrar esfuerzos en la investigación e innovación, creando centros de excelencia, desarrollo de habilidades necesarias para la utilización de la IA de profesionales y trabajadores, facilitar el acceso a la innovación en IA a las pymes y potenciar la asociación con el sector privado, entre otras. Dada las características de la IA, como puede ser su opacidad, esta puede provocar que la aplicación de la legislación actual sea complicada, por ello es necesario una revisión de toda la normativa actual al objeto de constatar si es o no, necesaria su adaptación.

Los distintos riesgos o problemas identificados, en dicho Libro blanco son: riesgo sobre la libertad de expresión, sobre la protección de datos, sobre la libertad de reunión o discriminación por raza, sexo u origen étnico, entre otros, todos ellos derechos

fundamentales; también identifica riesgos sobre la seguridad jurídica a la hora de la reclamación por responsabilidad civil o nuevos riesgos relacionados con la seguridad.

También, en la Declaración Europea (2023/C 23/01) sobre los Derechos y Principios Digitales para la Década Digital, se reconoce que la transformación digital afectará a todos los aspectos de la vida de las personas e incluye la inteligencia artificial como parte de esa transformación. Esta Declaración se compromete a garantizar el uso, la protección de los derechos que puedan verse afectados y el derecho a ser informados de que se está interactuando con una IA. Asimismo afirma su intención de que se convierta en una tecnología al servicio del ser humano, al objeto de mejorar su vida.

En relación con la legislación encargada de la IA, SIMÓN CASTELLANO (2022), destaca su fragmentación en muchas y distintos reglamentos, directivas o leyes, que complican la posibilidad de realizar una protección adecuada, así como una coordinación entre Estados.

Efectivamente, sin ser pormenorizados, si tomamos la diversa legislación que puede tratar el tema de la IA podríamos citar, por ejemplo: el Reglamento UE 2019/881 de 18 de abril de 2019, relativo a la Agencia de la Unión Europea para la Ciberseguridad; normativa europea de protección de datos representada por el Reglamento General de Protección de datos UE 2016/679 de 27 de abril de 2016; Reglamento UE 2018/1727 de 14 de noviembre de 2018 sobre la Agencia de la Unión Europea para la Cooperación Judicial Penal; así como, Reglamento UE 2016/794 de 11 de mayo de 2016 de la Agencia de UE para la Cooperación Policial.

Otros textos a nivel europeo que incluyen la IA serían:

1. Normas de Derecho civil sobre robótica. Resolución P8_TA (2017) 0051 del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL)).
2. Implicaciones de los macrodatos en los derechos fundamentales. Resolución P8_TA (2017) 0076 del Parlamento Europeo, de 14 de marzo de 2017, sobre las implicaciones de los macrodatos en los derechos fundamentales: privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)).
3. La inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales. Resolución P9_TA (2021) 0405 del

Parlamento Europeo, de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales (2020/2016(INI)).

4. Política industrial. Resolución P8_TA (2019) 0081 del Parlamento Europeo de 12 de febrero de 2019, sobre Una política industrial global europea en materia de inteligencia artificial y robótica (2018/2088(INI)).

2.7.2. Futuro

En colaboración con la Comisión Europea el Gobierno de España, ha iniciado un proyecto piloto para la implementación del primer entorno controlado de pruebas para evaluar cómo aplicar los requisitos a los sistemas de IA, de alto riesgo según la propuesta del Reglamento europeo de IA. Esta acción se enmarca en la estrategia de transformación digital de España, conocida como Agenda España Digital 2026, que forma parte del Plan de Recuperación y de la Estrategia Nacional de Inteligencia Artificial.

Los objetivos del entorno de pruebas incluyen fomentar la colaboración entre usuarios y proveedores de IA, acercar a las autoridades competentes a las empresas desarrolladoras de IA y lograr:

1. Claridad sobre los requisitos del futuro Reglamento de IA para los sistemas de IA.
2. Transferencia de conocimientos técnicos sobre el cumplimiento de la legislación.
3. Estímulo de la innovación y apoyo al desarrollo de nuevos sistemas confiables.
4. Evaluación de obligaciones y requisitos en un entorno controlado, seguido de la elaboración de guías técnicas y la recopilación de información para presentar resultados sobre todo el proyecto.

Este entorno se establece en el Real Decreto 817/2023 y pretende la participación de entidades privadas, administraciones públicas y entidades del sector público. Además, se seleccionarán sistemas de IA comercializados o en fase de pruebas para su inclusión en el banco de pruebas.

Se espera que este proyecto piloto, genere orientaciones para preparar a las empresas, de cara a la aprobación del Reglamento europeo de Inteligencia Artificial. El entorno de pruebas se enfoca en definir posibles formas de implementar los requisitos aplicables a sistemas de IA en áreas como seguridad de productos regulados, identificación biométrica, gestión de

infraestructuras críticas, educación, empleo, acceso a servicios, cumplimiento de la ley, gestión migratoria, entre otros.

El informe resultante de esta iniciativa contendrá, buenas prácticas, lecciones aprendidas y guías técnicas basadas en la experimentación práctica, que podrán ser utilizadas por la Comisión Europea en el desarrollo de directrices de la UE y contribuir al proceso de normalización y estandarización. Se establecerá un comité asesor de expertos para respaldar este proceso, integrado por profesionales independientes con prestigio y experiencia técnica en campos afines del conocimiento.

Como COTINO HUESO (2019) indica, no podemos pretender resolver los principales aspectos críticos que generan las tecnologías emergentes, a base de una acción legislativa reactiva que llega con retraso al verdadero problema. Por ello son necesarias acciones conjuntas de los diversos actores implicados, políticos, institucionales y del mundo del derecho para dar respuestas más adecuadas al núcleo de la cuestión.

Pretendiendo atender a lo que acabamos de exponer, el 9 de diciembre de 2023, el Parlamento llegó a un acuerdo provisional con el Consejo sobre la ley de AI. El texto acordado tendrá que ser adoptado formalmente ahora, tanto por el Parlamento, como por el Consejo para convertirse en legislación de la UE. Será la primera ley integral del mundo. Esta ley se basa en la Propuesta de Reglamento Del Parlamento Europeo y Del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley De Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM (2021) 206 final) (2021) y constituye un referente que pretende buscar un equilibrio entre las nuevas tecnologías y la protección de los derechos fundamentales.

Sin entrar a valorar en profundidad el texto, por claras cuestiones de espacio, podemos indicar que la ley dividiría las herramientas de IA en distintas categorías según su nivel de riesgo: las de riesgo inaceptable, las de riesgo alto, riesgo limitado y las de riesgo bajo. Esta clasificación sobre el riesgo de los sistemas y herramientas de IA implica a su vez que estos puedan, respectivamente, considerarse sistemas prohibidos, sistemas sometidos a regulación como sistemas de IA de alto riesgo, sometidos a requisitos de transparencia y aquellos que no sometidos a obligaciones pero sobre los que existirían unos códigos de conducta y requisitos mínimos.

Serán prohibidos los sistemas que utilicen técnicas subliminales para actuar sobre el comportamiento, los que utilicen vulnerabilidades de un grupo de personas en relación a su edad, discapacidad (física o mental) todos ellos susceptibles de causar perjuicios físicos o psicológicos. También, aquellos utilizados por las autoridades públicas que clasifiquen o evalúen a las personas físicas en función de sus características o comportamiento social, cuando cause un tratamiento perjudicial, aplicado fuera del ámbito en el que se generaron o recopilaban los datos, que resulta desmedido o injustificado en comparación con la conducta observada o su gravedad. Son prohibidos también sistemas de identificación biométrica remota en tiempo real en espacios de acceso público con fines policiales (con excepciones)

De alto riesgo serán diversos sistemas entre los que se encontraría aquellos destinados a la administración de justicia y en particular, aquellos destinados a ayudar en la investigación o incluso a la interpretación de hechos, siempre que puedan afectar a derechos y libertades de los individuos. Los que faciliten la vigilancia, la detención de personas o identificación biométrica, así como, aquellas que permitan la evaluación del riesgo o herramientas que detecten las emociones y aquellas similares al polígrafo. Evaluación de la fiabilidad de pruebas dentro del proceso penal, predicción de reiteración delictiva, perfilación criminal y análisis penales de individuos, serían otros de los sistemas considerados de alto riesgo.

La ley consciente de la importancia de los datos con los que se alimentan los sistemas de IA establece la necesidad de utilizar datos de calidad y establecer la obligatoriedad de la transparencia del proceso. Impone también que puedan ser fiscalizados por personas físicas.

Establece obligaciones para proveedores de datos al objeto de evitar sesgos en los mismos que puedan potenciarse con las tecnologías IA. Obliga también, a informar a las personas físicas cuando están interactuando con un sistema IA, salvo que sea evidente. Otras medidas de protección serían, el establecimiento de un registro público de aplicaciones IA de alto riesgo y la obligación de informar de incidentes o defectos graves.

Por supuesto, se prevén excepciones a la regla en situaciones donde los sistemas de inteligencia artificial estén autorizados por la ley para actividades relacionadas con la detección, prevención, investigación o enjuiciamiento de infracciones penales. Además, se establecen excepciones en casos en los que sea necesario para ejercer el derecho a la libertad de expresión, así como el derecho a la libertad en las artes y las ciencias, derechos garantizados por la Carta de los Derechos Fundamentales de la Unión Europea.

3. Resultados

Llegados a esta altura de nuestra exposición podemos presentar a continuación los resultados obtenidos de la investigación bibliográfica realizada durante todo el marco teórico.

3.1. Primera hipótesis. El mito de la malignidad de la IA

Respecto de esta primera hipótesis podemos establecer los siguientes resultados parciales:

1. Es evidente que la IA va a suponer una revolución en la vida diaria de los hombres. Diversos autores citados así lo avalan como JANUÁRIO (2023) o BARONA VILAR (2019). La propia UE se hace eco de esta revolución como en COMISIÓN EUROPEA (2020)
2. Los algoritmos y el *Big Data* constituyen parte esencial de la IA, tal y como hemos demostrado en los apartados 2.1 y 2.2. La IA necesita de datos y además de gran cantidad de datos y estos deben ser tratados adecuadamente para su uso en el aprendizaje de la máquina. Todo esto se lleva a cabo por medio de algoritmos que no serían más que las instrucciones del camino a seguir. Algoritmos y *Big Data* forman parte sustancial de la IA actual.
3. Atendiendo a lo expuesto en el apartado 2.3, es evidente que la IA no está libre de sesgos. Los propios algoritmos generan distintos tipos de errores o sesgos y estos serían de distintos tipos dependiendo de determinados factores. ANDREW NG (2018), CAÑADAS (2021) o PASQUINELLI (2022) dan muestras de ello. Esto es evidente si tenemos en cuenta la naturaleza real del proceso de aprendizaje, que no sería más que una aproximación estadística a un conjunto de datos facilitados a la máquina MC QUILLIAN (2018).
4. Todos los errores producidos por los algoritmos pueden ser controlados y mitigados. Aunque nunca eliminados del todo, debido a la misma naturaleza del proceso de aprendizaje. Su exposición más concreta se encontraría en el punto 2.3.3 a través de autores como ANDREW NG (2018), SURESH y GUTTAG (2021) o BAGNATO (2017). Además, los errores o sesgos algorítmicos estarían ligados, en muchos casos, con la intervención humana en alguna parte del proceso. Puntos 2.3 ,2.4 y 2.5. RUIZ-MORENO y TRUJILLO (2012), STANTON et al (2005).

5. La nueva legislación a nivel Europeo, es consciente de los problemas, beneficios y potencial de la IA. Sin embargo no renuncia a ella, sino que pretende con su regulación potenciarla. Este extremo queda expuesto en el punto 2.7 COMISIÓN EUROPEA (2020)

3.2. Segunda hipótesis. Falta de transparencia como principal problema de la IA

Los resultados respecto de esta hipótesis de trabajo serían los siguientes:

1. Es cierto que la complejidad de los sistemas de aprendizaje hace que los resultados de los algoritmos no sean fácilmente entendibles. Diversos autores como MONASTERIO ASTOBIZA (2017), MERCADER UGUINA (2022) o AGRAWAL et al. (2020), citados en el texto se han pronunciado en ese sentido.
2. La falta de transparencia supone un problema a la hora de controlar los resultados que arroja la IA. LEFEVRE CERVINI (2022) o VÉLEZ et al. (2022) defienden esa idea.
3. También, podemos afirmar que las IA son fuente de sesgos, errores y discriminación cuando se aplican al hombre, afectando a sus derechos fundamentales. Han dado fe de ello autores como CUATRECASAS MONFORTE (2023), BARONA VILAR (2019) o la propia COMISIÓN EUROPEA (2020).
4. Por último, es evidente la existencia de un problema ético en la aplicación de los sistemas de IA en el ámbito policial y judicial, toda vez que es, en estos campos, donde los derechos fundamentales pueden verse más afectados de forma grave. En el apartado 2.6 se ha constatado este hecho, a través de autores como MONTESINOS GARCÍA (2021) o FERNÁNDEZ-FERNÁNDEZ (2021).

4. Conclusiones

4.1. Primera hipótesis

En vista a lo expuesto hasta el momento, nuestra conclusión es clara respecto de la primera hipótesis; la IA, en su aplicación a la investigación criminal o judicial no supone el fin catastrofista de los derechos fundamentales ni el advenimiento de una sociedad controlada por el ojo que todo lo ve y todo lo sabe. La realidad de la cuestión es otra distinta; la IA no es más que un sistema tecnológico más, con sus beneficios y limitaciones.

Concretamente, en el campo de la investigación criminal, la IA todavía puede y debe mejorar, ya que supondrá un claro beneficio sobre la ciudadanía, al permitir combatir la delincuencia más eficazmente. Esto conllevará para los ciudadanos vivir en una sociedad más segura, pero también, más justa al no ver conculcados sus derechos fundamentales tan fácilmente.

Sin embargo, no debemos ser ingenuos, claramente existen peligros inherentes a su desarrollo. Peligros que, dentro del ámbito al que nos referimos pueden implicar serias consecuencias sobre los derechos de las personas reconocidos por la ley. Esto, sin embargo, no es muy distinto de cualquier otra tecnología creada por el hombre. Un arma de fuego puede ser empleada con un fin noble o no, al igual que el dominio de la energía nuclear. En la IA se repite el mismo patrón descrito, puede ser beneficiosa o perjudicial dependiendo de cómo la usemos. Somos nosotros quien le damos el valor, en un sentido o en otro.

Esta conclusión se apoya en los resultados obtenidos, los cuales nos indican que a pesar de la existencia indubitada de problemas o riesgos en el desarrollo de la IA; tales como existencia de sesgos, que pueden desencadenar situaciones de discriminación y de afectación sobre los derechos fundamentales de los individuos; estos, pueden ser controlados a través de distintas técnicas matemáticas y estadísticas.

Además, la idea de atribuir todos los errores a la IA es falaz. El hombre forma parte fundamental en la aparición de sesgos y errores, en una suerte de proyección sobre la tecnología de sus propias imperfecciones. Por lo tanto, la existencia de sesgo es consustancial con todos los aspectos de la vida humana. Y los ámbitos jurídico y de investigación policial no son distintos.

A pesar de ello es fácil dejarse llevar por el temor atávico de lo desconocido, imprimiendo sobre la IA facultades humanas, éticas o morales que realmente no tienen.

No cabe vuelta atrás. Esto ha sido entendido muy bien desde las instancias europeas que han visto la necesidad de legislar cuanto antes, pero no a cualquier precio. Es cierto que la nueva legislación, a punto de nacer, supone limitaciones y prohibiciones (siempre con excepciones), pero desde un punto de vista positivo, dota de un marco legal necesario para proteger a los ciudadanos de las consecuencias negativas o maliciosas que pudieran derivarse de la utilización de los sistemas de IA (siempre de la mano del hombre, no lo olvidemos).

Es necesario aprovechar esta oportunidad que nos brinda la nueva tecnología para comprender mejor la naturaleza humana y como nuestra visión parcial afecta a la sociedad donde se desarrolla. La IA es un instrumento tecnológico creado por nosotros; sus aciertos y sus desaciertos serán los nuestros.

4.2. Segunda hipótesis

Por último concluimos, que la falta de transparencia y la complejidad de los sistemas IA suponen el principal problema de su aplicación en los campos judiciales y de investigación delictual. Impidiendo que pueda apreciarse fácilmente la generación de errores. Ambos acrecientan la probabilidad de que la discriminación y la negación de derechos fundamentales hagan su aparición en la vida de las personas por motivos simplemente circunstanciales, como nacimiento en una determinada familia o vivir en un determinado barrio. No obstante, la implantación de equipos multidisciplinares que intervengan en la creación, prueba, puesta en marcha y control de estas herramientas, puede suponer el verdadero punto de inflexión en la problemática que se presenta.

En otro orden de las cosas el conflicto ético es claro, pero no muy distinto de los planteados por los fenómenos de los últimos años en nuestra sociedad que han supuesto un acrecentamiento del rigor punitivo y un adelanto de la barrera penal. Lo que cambia es el objeto de discusión. Aquí se plantea la naturaleza artificial del que genera la decisión. Sin embargo, no se debe olvidar, que de momento y por mucho tiempo, detrás de cada máquina hay un ser humano responsable de los resultados que arroja.

4.3. Algunas consideraciones finales

Desde nuestra perspectiva, el problema principal que hay que abordar no se trataría de una lucha hombre contra máquina, sino más bien la búsqueda de una simbiosis, de forma que el resultado final se constituya como algo superior a la suma de sus partes de sus partes.

Por lo tanto el hombre debe trabajar para mejorar su vida y sus características a través de la máquina.

4.4. Dificultades encontradas

Antes de finalizar queremos señalar que la principal dificultad encontrada, a la hora de la elaboración de esta investigación, ha sido la limitación de extensión en páginas. Ante un tema tan complejo y extenso, hubiera sido más adecuado poder elaborar un estudio de mayor calado, que nos permitiera explorar con profundidad algunos de los aspectos expuestos tan brevemente, en estas páginas.

Igualmente hubiera sido muy interesante poder desarrollar más ampliamente comentarios sobre la nueva Ley de IA europea, recién acordada y a punto de ser publicada. Indagando en sus luces y sombras.

4.5. Propuestas para líneas futuras de investigación

Atendiendo a la posibilidad de ampliar la investigación realizada, sería interesante, explorar la posibilidad de hacer una valoración profunda de las distintas herramientas aplicadas en nuestro país en relación a su impacto y efectividad en la investigación criminal, tanto en el ámbito del perfilado geográfico como en el caso del perfilado criminal.

Otra posibilidad sería investigar todas las herramientas de valoración de riesgo que se emplean hoy en día en nuestro entorno desde una perspectiva crítica que arroje mayor transparencia sobre sus resultados. Así permanecer expectantes ante la introducción de nuevas herramientas de IA en la administración de justicia al objeto de poder realizar los estudios correspondientes de su afectación sobre los derechos fundamentales de las personas.

Por último, como futura, línea de investigación propondríamos el análisis en profundidad de la nueva Ley sobre IA de la UE a punto de ser publicada, que permita establecer las posibles consecuencias que tendrá esta en el campo de la investigación criminal.

Referencias bibliográficas

Bibliografía básica

ABELIUK, A.; GUTIÉRREZ, C. «Historia y evolución de la inteligencia artificial». *Revista Bits de Ciencia*. 2021, núm. 21, pp. 14-21. Disponible en:

<https://revistasdex.uchile.cl/index.php/bits/article/download/2767/2700>

AGENCIA DE LOS DERECHOS FUNDAMENTALES DE LA UNIÓN EUROPEA (FRA). *Bias in algorithms: artificial intelligence and discrimination*. Oficina de Publicaciones de la Unión Europea, 2022. Disponible en: <https://data.europa.eu/doi/10.2811/25847>

AGENCIA EFE. «El alto comisionado de la ONU advierte sobre los peligros de la IA para los derechos humanos». *Agencia EFE*. 30 de noviembre de 2023. Disponible en:

<https://www.msn.com/es-es/noticias/internacional/el-alto-comisionado-de-la-onu-advierte-sobre-los-peligros-de-la-ia-para-los-derechos-humanos/ar-AA1kN5om>

AGRAWAL, A. et al. *Máquinas predictivas: La sencilla economía de la inteligencia artificial* [en línea]. Barcelona: Editorial Reverté, 2020 [consulta: 21 de noviembre de 2023]. Disponible en:

<http://ebookcentral.proquest.com/lib/univunirsp/detail.action?docID=29197643>.

ALEXANDER, T. M. *Human error assessment and reduction technique (HEART) and Human factor analysis and classification system (HFACS)*. NASA/Kennedy Space Center, John F. Kennedy Space Center in Collaboration on Quality in the Space and Defense Industries Forum (CQSDI), 2017. Disponible en: <https://core.ac.uk/download/pdf/141519682.pdf>

ALMACHE BARREIRO, J.C. y J. ALBERT MÁRQUEZ. «Implicaciones éticas de la IA y su potencial impacto en el derecho internacional». *Revista San Gregorio*. 2023, núm. 54, pp. 209-231 [consulta: 21 de noviembre de 2023]. ISSN 2528-7907. Disponible en:

<http://scielo.senescyt.gob.ec/pdf/rsan/v1n54/2528-7907-rsan-1-54-00209.pdf>

AMADOR HIDALGO, L. *Inteligencia artificial y sistemas expertos*. Córdoba: Universidad de Córdoba, Servicio de Publicaciones, 1996.

ANDREW NG. *Machine Learning. Technical Strategy for AI Engineers, In the Era of Deep Learning*. Self-publishing. Machine Learning Yearning-Draft, 2018.

BARONA VILAR, S. «Inteligencia Artificial o La Algoritmización de la Vida y de La Justicia: ¿Solución o Problema?». *Revista boliviana de Derecho de Derecho*. 2019, núm. 28, pp. 18-49 [consulta: 15 de octubre de 2023]. ISSN: 2070-8157. Disponible en:

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3510300

ASIMOV, I. *Yo, Robot*. 1ª ed. Barcelona: Edhasa, 2010.

BARONA VILAR S. «Una justicia “digital” y “algorítmica” para una sociedad en estado de mudanza», pp. 21-58. En BARONA VILAR S. (ed.) *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar* [en línea]. Valencia: Tirant lo Blanch, 2021a. [Consulta: 21 de noviembre de 2023]. Disponible en:

<https://bv.unir.net:4255/cloudLibrary/ebook/info/9788413972022>

BARONA VILAR, S. *Algoritmización del Derecho y de la Justicia. De la Inteligencia Artificial a la Smart Justice* [en línea]. Valencia: Tirant lo Blanch, 2021b. [Consulta: 21 de noviembre de 2023]. Disponible en: <https://bv.unir.net:4255/cloudLibrary/ebook/info/9788413786667>

BBC NEWS MUNDO. «La inteligencia artificial puede llevar a la extinción de la humanidad: la advertencia de los grandes expertos en IA». *BBC News Mundo*. 2023. Disponible en: <https://www.bbc.com/mundo/noticias-65759254>

BECK, U. *La sociedad del riesgo. Hacia una nueva modernidad*. Barcelona: Paidós, 2002.

BELLOSO MARTÍN, N. «La problemática de los sesgos algorítmicos (con especial referencia a los de género). ¿Hacia un Derecho a la protección contra los sesgos?» 45-78. En GARRIDO MARTÍN, J. y VALDIVIA JIMÉNEZ, R. (coord.) *Inteligencia artificial y filosofía del derecho*. Murcia: Laborum ediciones, 2022.

BERLANGA, A. «El camino desde la inteligencia artificial al Big Data». *Revista de Estadística y Sociedad*. 2016, núm. 68, pp. 9-11 [consulta: 10 de octubre de 2023]. ISSN-e 1696-9359. Disponible en: <http://www.revistaindice.com/numero68/p9.pdf>

BODEN, M. A. *Inteligencia artificial*. Madrid: Turner publicaciones, 2017.

CABALLERO, R. y E. MARTÍN. *Las bases de big data y de la inteligencia artificial* [en línea]. Madrid: Los libros de la Catarata, 2022. [Consulta: 25 de octubre de 2023]. Disponible en:

<https://bv.unir.net:2769/es/ereader/unir/233521?page=93>

CASTELLANO, P. S. y V. MAGRO SERVET. *Justicia cautelar e inteligencia artificial: la alternativa a los atávicos heurísticos judiciales* [en línea]. Barcelona: J M BOSCH EDITOR, 2021. [Consulta: 25 de octubre de 2023]. Disponible en: <https://bv.unir.net:2769/es/lc/unir/titulos/179962>

CARRILLO-GUTIÉRREZ, T. et al. «Análisis del error humano y la calidad del producto en la industria de manufactura de dispositivos médicos. Estudio de caso». *3C Tecnología*. 2021, vol. 10, núm. 1, pp. 73-91. [Consulta: 08 de enero de 2024]. ISSN: 2254 – 4143. Disponible en: https://3ciencias.com/wp-content/uploads/2021/03/art.-4_3C-Tecnologi%CC%81a-Ed.37-vol.10-n.1-2.pdf

CASAS ROMA, J., J. NIN GUERRERO y F. JULBE LÓPEZ. *Big data: análisis de datos en entornos masivos* [en línea]. Barcelona: Editorial UOC, 2019. [Consulta: 10 de octubre de 2023]. Disponible en: <https://bv.unir.net:2769/es/ereader/unir/117744?page=23>.

CCBE-COUNCIL OF BARS AND LAW SOCIETIES OF EUROPE. *Considerations on the legal aspects of Artificial Intelligence*. Bruselas: CCBE, 2020. Disponible en:

https://www.ccbe.eu/fileadmin/speciality_distribution/public/documents/IT_LAW/ITL_Guides_recommendations/EN_ITL_20200220_CCBE-considerations-on-the-Legal-Aspects-of-I.pdf

CEPEJ–EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE. *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment*. Pleno del CEPEJ 31ª sesión, 2018. Disponible en: <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>

CINELLI, V. y A.G. MANRIQUE. «El uso de programas de análisis predictivo en la inteligencia policial: una comparativa europea». *Revista de Estudios en Seguridad Internacional*, 2019, vol. 5, núm. 2, pp. 1-19. [Consulta: 20 de diciembre de 2023]. ISSN: 2444-6157. Disponible en: <http://www.seguridadinternacional.es/resi/index.php/revista/article/view/174>

COBO CANO, M. y L. LLORET IGLESIAS. *Inteligencia artificial y medicina* [en línea]. Madrid: Editorial CSIC. Los Libros de la Catarata, 2023. [Consulta: 29 de octubre de 2023]. Disponible en: <https://bv.unir.net:2769/es/ereader/unir/230641>

COLINA, C. «Manipulación algorítmica y sesgo psicosocial en redes sociales». *Temas de Comunicación*. 2023, núm. 46, pp. 6-26. [Consulta: 10 de diciembre de 2023]. Disponible en: <https://revistasenlinea.saber.ucab.edu.ve/index.php/temas/article/view/6219>

COMISIÓN EUROPEA. COM (2018) 237 final. *Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Inteligencia artificial para Europa*. Comisión Europea, 2018. Disponible en:

<https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52018DC0237>

COMISIÓN EUROPEA. COM (2020) 65 final. *Libro Blanco sobre la inteligencia artificial. Un enfoque europeo orientado a la excelencia y la confianza*. Comisión Europea, 2020. Disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065>

COMISIÓN EUROPEA. COM (2021) 206 final. *Propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de Inteligencia Artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión*. Comisión Europea, 2021. Disponible en:

https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0008.02/DOC_1&format=PDF

CORTADA DE KOHAN, Nuria; MACBETH, Guillermo. «Los sesgos cognitivos en la toma de decisiones». *Revista de Psicología*. 2006, vol. 2 núm. 3, pp. 55-70. [Consulta: 10 de octubre de 2023]. ISSN: 1669-2438 Disponible en:

<https://repositorio.uca.edu.ar/handle/123456789/6131>

CORTINA ORTS, A. «Ética de la inteligencia artificial». *Anales de la Real Academia de Ciencias Morales y Políticas*. Ministerio de Justicia, 2019, núm. 96, pp. 379-394. [Consulta: 10 de noviembre de 2023]. ISSN: 0210-296X. Disponible en:

https://www.boe.es/biblioteca_juridica/anuarios_derecho/abrir_pdf.php?id=ANU-M-2019-10037900394

COTINO HUESO, L. «Riesgos e impactos del Big Data, la inteligencia artificial y la robótica: enfoques, modelos y principios de la respuesta del derecho». *Revista general de Derecho administrativo*. 2019, núm. 50, pp. 1-37. [Consulta: 10 de noviembre de 2023]. ISSN: 1696-9650. Disponible en:

https://scholar.google.com/scholar?cluster=4165292729244863048&hl=es&as_sdt=0,5

COTINO HUESO, L. «Transparencia y explicabilidad de la inteligencia artificial y “compañía” (comunicación, interpretabilidad, inteligibilidad, auditabilidad, testabilidad, comprobabilidad,

simulabilidad...). Para qué, para quién y cuánta», pp. 25-70. En COTINO HUESO, L y J. CASTELLANOS CLARAMUNT (ed.). *Transparencia y explicabilidad de la inteligencia artificial* [en línea]. Valencia: Tirant lo Blanch, 2022. [Consulta: 10 de noviembre de 2023]. Disponible en: <https://bv.unir.net:4255/cloudLibrary/ebook/info/9788411471619>

CUATRECASAS MONFORTE, C. «La inteligencia artificial y la investigación de delitos». *Logos Guardia Civil, Revista Científica del Centro Universitario de la Guardia Civil*. 2023, núm. 1, pp. 61-84 [consulta: 10 de noviembre de 2023]. ISSN: 2952-394X. Disponible en: <https://revistacugc.es/article/view/5912>

DANAHER, J. «The Threat of Algocracy: Reality, Resistance and Accommodation». *Philosophy and Technology*. 2016, núm. 29, pp. 245-268 [consulta: 12 de octubre de 2023]. DOI 10.1007/s13347-015-0211-1. Disponible en: <https://philpapers.org/rec/DANTTO-13>

Declaraciones comunes Parlamento Europeo, Consejo, Comisión Europea. Declaración Europea sobre los Derechos y Principios Digitales para la Década Digital 2023/C 23/01. *Diario Oficial de la Unión Europea*, 23 de enero de 2023. Disponible en: [https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32023C0123\(01\)](https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32023C0123(01))

DIEBOLD, F. X. «On the Origin (s) and Development of the Term 'Big Data'». PIER Working Paper. 2012, núm. 12-037 [consulta: 10 de diciembre de 2023]. ISSN: 1556-5068 Disponible en: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2152421

DROR, I. E. «Cognitive and human factors in expert decision making: six fallacies and the eight sources of bias». *Analytical Chemistry*. 2020, vol. 92, núm. 12, pp. 7998-8004 [consulta: 10 de noviembre de 2023]. DOI: 10-1021. Disponible en: <https://pubs.acs.org/doi/full/10.1021/acs.analchem.0c00704>

ESTRATEGIA NACIONAL DE INTELIGENCIA ARTIFICIAL (ENIA). *Estrategia Nacional de Inteligencia Artificial*. Gobierno de España, 2020.

ESTRATEGIA NACIONAL DE INTELIGENCIA ARTIFICIAL (ENIA). *Informe sobre los avances en la Estrategia Nacional de Inteligencia Artificial*. Gobierno de España, 2023.

FALIERO, J. C. «Limitar la dependencia algorítmica: Impactos de la inteligencia artificial y sesgos algorítmicos». *Nueva Sociedad*. 2021, núm. 294, pp. 120-129 [consulta: 10 de

septiembre de 2023]. ISSN 0251-3552. Disponible en: <https://nuso.org/articulo/limitar-la-dependencia-algoritmica/>

FERRANTE, E. «Inteligencia artificial y sesgos algorítmicos ¿Por qué deberían importarnos?». *Nueva sociedad*. 2021, núm. 294, pp. 27-36 [consulta: 10 de noviembre de 2023]. ISSN 0251-3552, Disponible en: <https://biblat.unam.mx/hevila/Nuevasociedad/2021/no294/3.pdf>

FERNÁNDEZ-FERNÁNDEZ, J.L. «Hacia el Humanismo Digital desde un denominador común para la Ciber Ética y la Ética de la Inteligencia Artificial». *Disputatio : Philosophical Research Bulletin*. 2021, vol. 10, núm. 17, pp. 107-130 [consulta: 10 de noviembre de 2023]. ISSN 2254-0601. Disponible en: <https://doi.org/10.5281/zenodo.5136247>

GARCÍA-CAMPOS, J., S. SARABIA-LÓPEZ y P. HERNÁNDEZ-CHÁVEZ. «Tres grandes enigmas de los sesgos cognitivos». *SCIO: Revista de Filosofía*. 2022, núm. 22, pp. 99-125 [consulta: 10 de noviembre de 2023]. ISSN 2603-6924. Disponible en: <https://revistas.ucv.es/scio/index.php/scio/article/view/1031>

GÓMEZ COLOMER J.L. *El Juez Robot. La independencia judicial en peligro* [en línea]. Valencia: Tirant lo Blanch, 2023. [Consulta: 10 de noviembre de 2023]. Disponible en: <https://bv.unir.net:4255/cloudLibrary/ebook/info/9788411691291>

GONZÁLEZ-ÁLVAREZ, J. L., J. S. HERMOSO y M. CAMACHO-COLLADOS. «Policía predictiva en España. Aplicación y retos futuros». *Behavior & Law Journal*. 2020, vol. 6, núm. 1, pp. 26-41 [consulta: 10 de noviembre de 2023]. ISSN 2444-4170. Disponible en: <https://www.behaviorandlawjournal.com/BLJ/article/view/75>

GONZÁLEZ, J. «La inteligencia artificial se forma como activista del clima». ABC. 30 de noviembre de 2023. Disponible en: <https://www.abc.es/antropia/inteligencia-artificial-forma-activista-clima-20231130065747-nt.html>

GONZÁLEZ FUSTER, G. *Artificial Intelligence and law enforcement impact on fundamental rights*. Parlamento Europeo, 2020. Disponible en: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU\(2020\)656295_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/656295/IPOL_STU(2020)656295_EN.pdf)

GRAJALES ARTURO, A. «Error judicial y verdad. Una mirada transversal entre el proceso y la teoría del derecho», pp. 31-55. En CARBONELL, F., A. MEROI y A. USCANGA (coord.). *El error judicial. Problemas y regulaciones* [en línea]. Valencia: Tirant lo Blanch, 2023. [Consulta: 10 de noviembre de 2023]. Disponible en:

<https://bv.unir.net:4255/cloudLibrary/ebook/info/9788411695688>

HERNÁNDEZ-AVILA, M., F. GARRIDO y E. SALAZAR-MARTÍNEZ. «Sesgos en estudios epidemiológicos». *Salud pública de México*. 2000, vol. 42, núm. 5, pp. 438-446 [consulta: 10 de noviembre de 2023]. Disponible en: <https://www.scielosp.org/pdf/spm/v42n5/3995.pdf>

HERNÁNDEZ-LEAL, E. J., N. D. DUQUE-MÉNDEZ y J. MORENO-CADAVID. «Big Data: una exploración de investigaciones, tecnologías y casos de aplicación». *TecnoLógicas*. 2017, vol. 20, núm. 39, pp. 17-24 [consulta: 05 de septiembre de 2023]. ISSN 2256-5337. Disponible en: <http://www.scielo.org.co/pdf/teclo/v20n39/v20n39a02.pdf>

JANUÁRIO, T. F. X. «Inteligencia artificial y responsabilidad penal de personas jurídicas: un análisis de sus aspectos materiales y procesales». *Estudios Penales y Criminológicos*. 2023, vol. 44, núm. Especial “Inteligencia artificial y sistema penal”, pp. 1-39 [consulta: 10 de octubre de 2023]. ISSN 2340-0080. Disponible en: <https://estudogeral.uc.pt/handle/10316/109996>

JOBIN, A., M. IENCA y E. VAYENA. «The global landscape of AI ethics guidelines». *Nature machine intelligence*. 2019, vol. 1, núm. 9, pp. 389-399 [consulta: 10 de octubre de 2023]. ISSN 2522-5839. Disponible en: <https://www.nature.com/articles/s42256-019-0088-2>

JRC-JOIN RESEARCH CENTRE. JRC Technical Reports. *AI Watch Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*. Comisión Europea, 2020. Disponible en: <https://publications.jrc.ec.europa.eu/repository/handle/JRC118163>

KAI-FU, L. «La verdadera amenaza de la inteligencia artificial». *The New York Times*. 27 de junio de 2017. Disponible en:

<https://www.nytimes.com/es/2017/06/27/espanol/opinion/la-verdadera-amenaza-de-la-inteligencia-artificial.html>

LANCHO PEDRERA, F. «Los sistemas expertos en el Derecho». *Anuario De La Facultad De Derecho, Universidad De Extremadura*. 2003, núm. 21, pp. 629-636 [consulta: 10 de octubre de 2023]. ISSN 2695-7728. Disponible en:

<https://dialnet.unirioja.es/servlet/articulo?codigo=854374>

LEFEVRE CERVINI, E. *Uso estratégico de datos e inteligencia artificial en la justicia*. CAF- banco de desarrollo de América Latina, 2022. Disponible en:

<https://scioteca.caf.com/handle/123456789/1945>

LÓPEZ DE MÁNTARAS BADIA, R. y P. MESEGUER GONZÁLEZ. *Inteligencia artificial* [en línea]. Madrid: Editorial CSIC, 2017. [Consulta: 12 de octubre de 2023]. Disponible en:

<https://bv.unir.net:2769/es/lc/unir/titulos/42319>

LLEDÓ, B. I. *El Derecho Penal, Robots, IA y Cibercriminalidad: Desafíos éticos y Jurídicos. ¿Hacia una Distopía?*, [en línea]. Madrid: Dykinson, S.L., 2022. [Consulta: 15 de diciembre de 2023]

Disponible en:

<http://ebookcentral.proquest.com/lib/univunirsp/detail.action?docID=29195640>.

MANTEROLA, C. y T. OTZEN. «Los sesgos en investigación clínica». *International Journal of Morphology*. 2015, vol. 33, núm. 3, pp. 1156-1164 [consulta: 15 de diciembre de 2023]. ISSN 0717-9502. Disponible en: <http://dx.doi.org/10.4067/S0717-95022015000300056>

MARR, B. «Estos son los 15 mayores riesgos de la IA». *Forbes10years*. 28 de julio de 2023. Disponible en: <https://forbes.es/tecnologia/316482/estos-son-los-15-mayores-riesgos-de-la-ia/>

MARTÍNEZ DEVIA, A. «La Inteligencia Artificial, el Big Data y la Era Digital: Una Amenaza para los Datos Personales». *Rev. Prop. Inmaterial*. 2019, vol. 27, pp. 5-23 [consulta: 15 de diciembre de 2023]. ISSN 1657-1959. Disponible en: <https://doi.org/10.18601/16571959.n27.01>

MAYBIN, S. « ¿Cómo en Estados Unidos las matemáticas te pueden meter en prisión?». *BBC News Mundo*. 17 de octubre de 2016. Disponible en: <https://www.bbc.com/mundo/noticias-37679463>

MEDINA LA PLATA, E. H. *Big data: los datos como generadores de valor* [en línea]. Lima: Editorial Universidad Peruana de Ciencias Aplicadas, 2023. [Consulta: 15 de diciembre de 2023]. Disponible en: <https://bv.unir.net:2769/es/ereader/unir/231338?page=22>.

MERCADER UGUINA, J. R. *Algoritmos e inteligencia artificial en el derecho digital del trabajo* [en línea]. Valencia: Tirant lo Blanch, 2022. [Consulta: 15 de diciembre de 2023]. Disponible en: <https://bv.unir.net:4255/cloudLibrary/ebook/info/9788411306485>

MEHRABI, N. et al. «A survey on bias and fairness in machine learning». *ACM computing surveys (CSUR)*. 2021, vol. 54, núm. 6, pp. 1-35 [consulta: 15 de diciembre de 2023]. ISSN 0360-0300. Disponible en:

https://scholar.google.com/scholar?cluster=10847217828384774171&hl=es&as_sdt=0,5

MIRÓ LLINARES, F. «Inteligencia artificial y justicia penal: más allá de los resultados lesivos causados por robots». *Revista De Derecho Penal Y Criminología*. 2020a, núm. 20, pp. 87–130 [consulta: 15 de diciembre de 2023]. ISSN 1132-9955. Disponible en:

<https://doi.org/10.5944/rdpc.20.2018.26446>

MIRÓ LLINARES, F. «Predictive policing: utopia or dystopia? On attitudes towards the use of big data algorithms for law enforcement». *IDP: Revista De Internet, Derecho y Política*. 2020b, núm. 30 [consulta: 15 de diciembre de 2023]. ISSN 1699-8154. Disponible en:

<https://www.raco.cat/index.php/IDP/article/view/373608>

MONASTERIO ASTOBIZA, A. «Ética algorítmica: Implicaciones éticas de una sociedad cada vez más gobernada por algoritmos». *Dilemata*, 2017, núm. 24, pp. 185-217 [consulta: 21 de noviembre de 2023]. ISSN 1989-7022. Disponible en:

<https://www.dilemata.net/revista/index.php/dilemata/article/view/412000107>

MONTESINOS GARCÍA, A. «Justicia penal predictiva. Una justicia “digital” y “algorítmica” para una sociedad en estado de mudanza», pp. 423-454. En BARONA VILAR, S. (ed.). *Justicia algorítmica y neuroderecho. Una mirada multidisciplinar* [en línea]. Valencia: Tirant lo Blanch, 2021. [Consulta: 21 de noviembre de 2023]. Disponible en:

<https://bv.unir.net:4255/cloudLibrary/ebook/show/9788411135719>

MUÑOZ GUTIÉRREZ, C. «La discriminación en una sociedad automatizada: Contribuciones desde América Latina». *Revista chilena de derecho y tecnología*. 2021, vol. 10, núm. 1, pp. 271-307 [consulta: 21 de noviembre de 2023]. ISSN 0719-2584. Disponible en:

<https://rchdt.uchile.cl/index.php/RCHDT/article/view/58793>

MUÑOZ RODRÍGUEZ, A. B. «El impacto de la inteligencia artificial en el proceso penal». *Anuario de la Facultad de Derecho. Universidad de Extremadura*. 2020, núm. 36, pp. 695-728 [consulta: 21 de noviembre de 2023]. ISSN 2695-7728. Disponible en:

<https://publicaciones.unex.es/index.php/AFD/article/view/489>

MURGAS, C. «Pausar la Inteligencia artificial: ¡tu seguridad está en juego!». *Forbes10years*. 10 de abril de 2023. Disponible en:

<https://forbes.es/forbes-funds/260664/pausar-la-inteligencia-artificial-tu-seguridad-esta-en-juego/>

NÓHPAL RODRÍGUEZ, V. S. «La posibilidad de la policía predictiva». *Revista Mexicana de Ciencias Penales*. 2021, vol. 4, núm. 14, pp. 135-144 [consulta: 21 de noviembre de 2023]. ISSN 2954-4963. Disponible en:

<https://revistaciencias.inacipe.gob.mx/index.php/02/article/view/438>

ODDI, F. J., et al. «Ambigüedades en términos científicos: El uso del “error” y el “sesgo” en estadística». *Ecología austral*. 2018, vol. 28, núm. 3, pp. 525-536 [consulta: 21 de noviembre de 2023]. ISSN 1667-782X. Disponible en:

http://www.scielo.org.ar/scielo.php?pid=S1667-782X2018000400012&script=sci_arttext

PASQUINELLI, M. «Cómo una máquina aprende y falla: una gramática del error para la Inteligencia Artificial (Traducción de Emilio Cafassi, Carolina Monti, Hernán Peckaitis y Graciana Zarauza)». *Hipertextos*. 2022, vol. 10, núm. 17, pp. 13-29 [consulta: 21 de noviembre de 2023]. ISSN 2314-3924. Disponible en:

<https://revistas.unlp.edu.ar/hipertextos/article/view/14028>

PARLAMENTO EUROPEO. *Resolución P8_TA (2017) 0051 del Parlamento Europeo, de 16 de febrero de 2017, con recomendaciones destinadas a la Comisión sobre normas de Derecho civil sobre robótica (2015/2103(INL))*. Parlamento Europeo, 2017. Disponible en:

https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_ES.pdf

PARLAMENTO EUROPEO. *Resolución P8_TA (2017) 0076 del Parlamento Europeo de 14 de marzo de 2017 sobre las implicaciones de los macrodatos en los derechos fundamentales:*

privacidad, protección de datos, no discriminación, seguridad y aplicación de la ley (2016/2225(INI)). Parlamento Europeo, 2017. Disponible en:

https://www.europarl.europa.eu/doceo/document/TA-8-2017-0076_ES.pdf

PARLAMENTO EUROPEO. *Resolución P8_TA (2019) 0081 del Parlamento Europeo de 12 de febrero de 2019, sobre Una política industrial global europea en materia de inteligencia artificial y robótica (2018/2088(INI))*. Parlamento Europeo, 2019. Disponible en:

https://www.europarl.europa.eu/doceo/document/TA-8-2019-0081_ES.pdf

PARLAMENTO EUROPEO. *Resolución P9_TA (2021) 0405 del Parlamento Europeo, de 6 de octubre de 2021, sobre la inteligencia artificial en el Derecho penal y su utilización por las autoridades policiales y judiciales en asuntos penales (2020/2016(INI))*. Parlamento Europeo, 2020. Disponible en:

https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_ES.pdf

PERALTA GUTIÉRREZ, A. P. y L. S. TORRES LÓPEZ. «Herramientas de inteligencia artificial en el ámbito jurídico comparado», pp. 295-330. En HERRERA TRIGUERO, F., A. PERALTA GUTIERREZ y L. S. TORRES LÓPEZ (coord.). *El derecho y la inteligencia artificial [en línea]*. Granada: Editorial Universidad de Granada, 2022. [Consulta: 10 de noviembre de 2023]. Disponible en:

<https://bv.unir.net:2689/viewer/9788433870506/1>

PERRY, W. L. et al. *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Santa Monica, CA: RAND Corporation, 2013.

PINEDA PERTUZ, C. *Aprendizaje automático y profundo en Python: una mirada hacia la inteligencia artificial [en línea]*. Madrid: RA-MA Editorial, 2022. [Consulta: 10 de noviembre de 2023]. Disponible en: <https://bv.unir.net:2769/es/ereader/unir/230579>

PINO V., L. E., A. E. RICO-CARRILLO y A. HERNÁNDEZ-ARANGO. «Del ábaco a las redes neuronales o la breve historia de la inteligencia artificial en salud». *Medicina*. 2022, vol. 43, núm. 4, pp. 514-526 [consulta: 10 de septiembre de 2023]. ISSN 0120-5498. Disponible en: <https://revistamedicina.net/index.php/Medicina>

RATTINGER, A. «Las Consecuencias de la Inteligencia Artificial en el Trabajo de Marketing». *Merca2.0*. 30 de noviembre de 2023. Disponible en:

<https://www.merca20.com/las-consecuencias-de-la-inteligencia-artificial-en-el-trabajo-de-marketing/>

REBOLLO DELGADO, L. *Inteligencia artificial y derechos fundamentales* [en línea]. Madrid: Dykinson, 2023. [Consulta: 10 de noviembre de 2023]. Disponible en:

<https://bv.unir.net:2769/es/lc/unir/titulos/232079>

ROJAS, M. E. A. «Una mirada a la inteligencia artificial». *Revista Ingeniería, Matemáticas y Ciencias de la Información*. 2015, vol. 2, núm. 3, pp. 27-31 [consulta: 10 de noviembre de 2023]. ISSN 2357-3716. Disponible en:

<http://ojs.urepublicana.edu.co/index.php/ingenieria/article/view/234>

ROSE, I. «Los trabajadores que ya han sido sustituidos por la inteligencia artificial». *BBC News Mundo*. 09 de julio 2023. Disponible en: <https://www.bbc.com/mundo/articles/cn42glz5q8ro>

ROUHIAINEN, L. *Inteligencia artificial. 101 cosas que debes saber hoy sobre nuestro futuro* [en línea]. Madrid: Alienta Editorial, 2018. [Consulta: 10 de noviembre de 2023]. Disponible en:

https://planetadelibrosec0.cdnstatics.com/libros_contenido_extra/40/39308_Inteligencia_artificial.pdf

RUIZ-MORENO, J. M. y H. M. TRUJILLO. «Modelos para la evaluación del error humano en estudios de fiabilidad de sistemas». *Anales de Psicología/Annals of Psychology*. 2012, vol. 28, núm. 3, pp. 962-976 [consulta: 10 de noviembre de 2023]. ISSN 1695-2294. Disponible en:

<https://revistas.um.es/analesps/article/view/analesps.28.3.148941>

RUSSELL, S. J., P. NORVIG y J. M. CORCHADO RODRÍGUEZ. *Inteligencia artificial: un enfoque moderno* [en línea]. México, D.F: Pearson Educación, 2008. [Consulta: 11 de septiembre de 2023]. Disponible en:

https://bv.unir.net:2522/ib/NPcd/IB_Escritorio_Visualizar?cod_primaria=1000193&libro=2864

SEARLE, J. R. «Minds, brains, and programs». *Behavioral and brain sciences*. 1980, vol. 3, núm. 3, pp. 417-424 [consulta: 10 de noviembre de 2023]. ISSN 1469-1825. Disponible en:

<https://web-archive.southampton.ac.uk/cogprints.org/7150/1/10.1.1.83.5248.pdf>

SIMÓN CASTELLANO, P. «Inteligencia artificial y Administración de Justicia: ¿Quo vadis, justicia?». *IDP: Revista De Internet, Derecho y Política*. 2021, núm. 33, pp. 1-14 [consulta: 10 de noviembre de 2023]. ISSN 1699-8154. Disponible en:

<https://www.raco.cat/index.php/IDP/article/view/373817/472448>

SIMÓN CASTELLANO, P. *La prisión algorítmica* [en línea]. Valencia: Tirant lo Blanch, 2022. [Consulta: 10 de noviembre de 2023]. Disponible en:

<https://bv.unir.net:4255/cloudLibrary/ebook/info/9788411138086>

SOLETO, H. *Garantías y Errores en la Investigación Penal. Ciencia Versus Memoria* [en línea]. Valencia: Tirant lo Blanch, 2016. [Consulta: 10 de noviembre de 2023]. Disponible en:

<https://bv.unir.net:4255/cloudLibrary/ebook/info/9788491195498>

SORIA OLIVAS, E. *Inteligencia Artificial: casos prácticos con aprendizaje profundo* [en línea]. Madrid: RA-MA Editorial, 2022. [Consulta: 01 de septiembre de 2023]. Disponible en:

<https://bv.unir.net:2769/es/lc/unir/titulos/222670>

STANTON, N. A. et al. *Human Factors Methods: A Practical Guide for Engineering and Design* [en línea]. Hampshire: Taylor & Francis Group, 2005. [Consulta: 15 de noviembre de 2023]. Disponible en:

<https://bv.unir.net:2056/lib/univunirsp/reader.action?docID=429580>

STEINER, C. *Automate this. How algorithms came to rule our world*. New York: Portfolio, Penguin, 2012.

SURESH, H. y J. V. GUTTAG. «A Framework for Understanding Sources of Harm Throughout the Machine Learning Life Cycle», artículo núm. 17 pp. 1-9. En *EAAMO '21: Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* [en línea]. New York: Association for Computing Machinery, 2021. [Consulta: 10 de noviembre de 2023]. Disponible en: <https://doi.org/10.1145/3465416.3483305>

TABARES, L. F. y J. F. HERNÁNDEZ. «Big Data analytics: Oportunidades, retos y tendencias». *Universidad de San Buenaventura*. 2014, vol. 20 [consulta: 10 de noviembre de 2023]. Disponible en:

<https://gc.scalahed.com/recursos/files/r161r/w25569w/Big%20Data%20Analytics.pdf>

The Terminator. Dirigida por James CAMERON. EE.UU: Hemdale Film Corporation, 1984.

ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA EDUCACIÓN, LA CIENCIA Y LA CULTURA (UNESCO). *Recomendación sobre la ética de la inteligencia artificial*. Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura, 2022.

VALENZUELA, L. A. «Los algoritmos digitales en el trabajo. Brechas y sesgos». *Relaciones Laborales y Derecho del Empleo*. 2021, vol. 9, núm.4, pp.120-146 [consulta: 10 de noviembre de 2023]. ISSN 2282-2313. Disponible en:

file:///C:/Users/casa_1/Downloads/araguez_revista_n4_2021_def-1.pdf

VALLS PRIETO, J. *Problemas Jurídico Penales Asociados a Las Nuevas Técnicas de Prevención y Persecución Del Crimen Mediante Inteligencia Artificial* [en línea]. Madrid: Dykinson, S.L., 2017. [Consulta: 10 de noviembre de 2023]. Disponible en:

<https://bv.unir.net:2056/lib/univunirsp/detail.action?docID=6514512>.

VÉLEZ, M. I., C. GÓMEZ SANTAMARÍA y M. A. OSORIO SANABRIA. *Conceptos fundamentales y uso responsable de la inteligencia artificial en el sector público. Informe 2*. Banco de Desarrollo de América Latina (CAF), 2022. [Consulta: 17 de noviembre de 2023]. Disponible en: <https://scioteca.caf.com/handle/123456789/1921>

VICENTE, L. y MATUTE, H. «Humans inherit artificial intelligence biases». *Scientific Reports*. 2023, vol. 13, núm. 15737 [consulta: 10 de noviembre de 2023]. ISSN 2045-2322. Disponible en: <https://doi.org/10.1038/s41598-023-42384-8>

VILLANUEVA, C. M. «Sesgos Implícitos y Juicios por Jurados». *Revista Jurídica Cesumar*. 2021, vol. 21, núm. 3, pp. 835-863 [consulta: 10 de noviembre de 2023]. ISSN 1677-6402. Disponible en: <https://periodicos.unicesumar.edu.br/index.php/revjuridica>

ZARZA, G., y MURPHY, J. J. L. *La ingeniería del big data: Cómo trabajar con datos* [en línea]. Barcelona: Editorial UOC, 2017. [Consulta: 10 de octubre de 2023]. Disponible en:

<https://bv.unir.net:2056/lib/univunirsp/detail.action?docID=7051391&query=inteligencia%20artificial#>

Bibliografía complementaria

CASABONA, C. M. R. «Riesgo, procedimientos actuariales basados en inteligencia artificial y medidas de seguridad». *Revista de Derecho, Empresa y Sociedad (REDS)*. 2018, núm. 13, pp. 39-55 [consulta: 10 de noviembre de 2023]. ISSN 2340-4647. Disponible en: <https://dialnet.unirioja.es/descarga/articulo/6859383.pdf>

CEBALLOS-ESPINOZA, F. «De la criminología clásica a la criminología moderna: La investigación criminal multifactorial en la era digital-De la Criminología Clásica». *Formación y Desarrollo Policial*. 2021, vol. 3, núm. 1, pp. 55-85 [consulta: 10 de noviembre de 2023]. ISSN 2452-4271. Disponible en: <https://www.aacademica.org/fceballose/22/1.pdf>

COMISIÓN EUROPEA. COM (2018) 795 final. Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Plan coordinado sobre la inteligencia artificial. Comisión Europea, 2018. Disponible en:

<https://www.prodetur.es/prodetur/AlfrescoFileTransferServlet?action=download&ref=ff2d375b-9f69-4c49-bece-3a6e39e135b5>

CORNEJO, M. A. N., S. V. E. DESIDERIO y O. E. O. AGREDA. «Estilos de aprendizajes e inteligencia artificial». *Polo del Conocimiento: Revista científico-profesional*. 2020, vol. 5, núm. 9, pp. 222-253 [consulta: 10 de noviembre de 2023]. ISSN 2550-682X. Disponible en:

<https://dialnet.unirioja.es/servlet/articulo?codigo=7554412>

DELLUNDE I CLAVÉ, P., O. PUJOL VILA y J. VITRIÀ I MARCA. «Cerrando una brecha: una reflexión multidisciplinar sobre la discriminación algorítmica». *Daimon. Revista Internacional de Filosofía*. 2023, núm. 90, pp. 63-80 [consulta: 10 de diciembre de 2023]. ISSN 1989-4651. Disponible en: <https://revistas.um.es/daimon/article/view/562811>

FERNÁNDEZ, R. C. *Una ética para Big Data : Introducción a la Gestión ética de Datos Masivos* [en línea]. Barcelona: Editorial UOC, 2017. [Consulta: 10 de diciembre de 2023]. Disponible en: <https://bv.unir.net:2056/lib/univunirsp/detail.action?docID=7025875>.

HERRERA DE LAS HERAS, R. *Aspectos legales de la inteligencia artificial: personalidad jurídica de los robots, protección de datos y responsabilidad civil* [en línea]. Madrid: Dykinson, 2022. [Consulta: 10 de diciembre de 2023]. Disponible en:

<https://bv.unir.net:2769/es/lc/unir/titulos/219335>

LLAMAS COVARRUBIAS, J. Z., O. A. MENDOZA ENRÍQUEZ y M. G. GUERRERO. «Enfoques regulatorios para la Inteligencia Artificial (IA)». *Revista Chilena De Derecho*. 2022, vol. 49, núm. 3, pp. 31-62 [consulta: 10 de diciembre de 2023]. ISSN 0718-3437. Disponible en:

<https://dialnet.unirioja.es/servlet/articulo?codigo=9114760>

MENDOZA, D. «Racismo y roles de género, conductas perpetuadas en algoritmos de inteligencia artificial». *Coloquio*. 2020, núm. 65, pp. 131-135 [consulta: 10 de diciembre de 2023]. ISSN 1390-2865. Disponible en:

<https://revistas.uazuay.edu.ec/index.php/coloquio/issue/view/48/PDF>

NÚÑEZ ZORRILLA, M. C. *Inteligencia artificial y responsabilidad civil: régimen jurídico de los daños causados por robots autónomos con inteligencia artificial* [en línea]. Madrid: Reus Editorial, 2019. [Consulta: 10 de diciembre de 2023]. Disponible en:

<https://bv.unir.net:2769/es/lc/unir/titulos/127671>

OLIVER, N. *Inteligencia artificial, naturalmente: un manual de convivencia entre humanos* [en línea]. Madrid: Ministerio de Asuntos Económicos y Transformación Digital, Secretaria General Técnica, 2020. [Consulta: 10 de diciembre de 2023]. Disponible en:

<https://bv.unir.net:2769/es/lc/unir/titulos/175967>

YANKE, R. « "La tecnología no tiene ética, pero la humanidad depende de ella"». *El Mundo*. 2023. Disponible en: <https://lab.elmundo.es/inteligencia-artificial/riesgos.html>

Legislación citada

España. Ley Orgánica 13/2015, de 5 de octubre, de modificación de la Ley de Enjuiciamiento Criminal para el fortalecimiento de las garantías procesales y la regulación de las medidas de investigación tecnológica. *Boletín Oficial del Estado*, 06 de octubre de 2015, núm. 239, pp. 90192-90219. Disponible en: <https://www.boe.es/eli/es/lo/2015/10/05/13>

España. Ley 1/2019, de 20 de febrero, de Secretos Empresariales. *Boletín Oficial del Estado*, 21 de febrero de 2019, núm. 45, pp. 16713-16727. Disponible en:

<https://www.boe.es/eli/es/l/2019/02/20/1>

España. 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 por el que se establecen normas armonizadas en materia de inteligencia artificial. *Boletín Oficial del Estado*, 9 de noviembre de 2023, núm. 268, pp. 149138-149168. Disponible en:

https://www.boe.es/diario_boe/txt.php?id=BOE-A-2023-22767

Reglamento (UE) 2019/881 del Parlamento Europeo y del Consejo, de 17 de abril de 2019, relativo a ENISA (Agencia de la Unión Europea para la Ciberseguridad) y a la certificación de la ciberseguridad de las tecnologías de la información y la comunicación y por el que se deroga el Reglamento (UE) nº 526/2013 («Reglamento sobre la Ciberseguridad»). *Diario Oficial de las Comunidades Europeas*, 7 de junio de 2019, núm. 151. Disponible en:

<https://www.boe.es/buscar/doc.php?id=DOUE-L-2019-80998>

Reglamento (UE) 2018/1727 del Parlamento Europeo y del Consejo, de 14 de noviembre de 2018, sobre la Agencia de la Unión Europea para la Cooperación Judicial Penal (Eurojust) y por el que se sustituye y deroga la Decisión 2002/187/JAI del Consejo. *Diario Oficial de las Comunidades Europeas*, 21 de noviembre de 2018, núm. 138. Disponible en:

<https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32018R1727>

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). *Diario Oficial de las Comunidades Europeas*, 4 de mayo de 2016, núm. 119. Disponible en:

<https://www.boe.es/buscar/doc.php?id=DOUE-L-2016-80807>

Reglamento (UE) 2016/794 del Parlamento Europeo y del Consejo, de 11 de mayo de 2016, relativo a la Agencia de la Unión Europea para la Cooperación Policial (Europol) y por el que se sustituyen y derogan las Decisiones 2009/371/JAI, 2009/934/JAI, 2009/935/JAI, 2009/936/JAI y 2009/968/JAI del Consejo. *Diario Oficial de las Comunidades Europeas*, 24 de mayo de 2016, núm. 135. Disponible en: <https://www.boe.es/doue/2016/135/L00053-00114.pdf>

Listado de abreviaturas

CCBE: Council Of Bars And Law Societies Of Europe

CEPEJ: Comisión europea para la eficiencia de la justicia

ENIA: Estrategia Nacional de Inteligencia Artificial

FRA: Agencia de los Derechos Fundamentales de la Unión Europea

IA: Inteligencia Artificial

JRC: Centro Común de Investigación

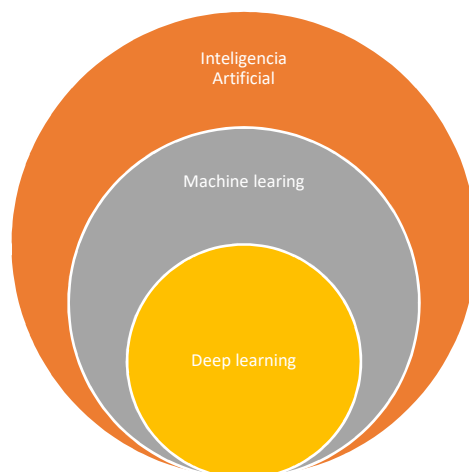
RAE: Real Academia de la Lengua Española

UNESCO: Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

Anexo A. Figuras 1 y 2

A continuación se exponen dos figuras que esquematizan el apartado 2.1.4 Tipos de inteligencia Artificial.

Figura 1: Inteligencia artificial.



Fuente: elaboración propia a partir de a partir de datos presentados en marco teórico

Figura 2: Tipos de aprendizaje automático



Fuente: elaboración propia a partir de a partir de clasificación realizada por ROUHAINEN, 2018