



Facultad de Ciencias Económicas y Empresariales

IMPLICACIONES ÉTICO-JURÍDICAS DEL USO DE ALGORITMOS DE APRENDIZAJE AUTOMATIZADO EN EL ÁMBITO DE LA BIOMEDICINA

Autor: Victoria Ruiz-Gallardón López- Monís

Director: Francisco Javier Fuertes Pérez

Resumen:

En la actualidad, el uso de la inteligencia artificial (IA) se ha expandido a muchos ámbitos de la sociedad, y la medicina no ha sido la excepción. Los algoritmos de aprendizaje automatizado se han convertido en una herramienta cada vez más utilizada en el ámbito biomédico, debido a sus múltiples aplicaciones, como el diagnóstico y tratamiento de enfermedades, desarrollo de medicamentos, y la gestión hospitalaria, entre otros. Sin embargo, este uso creciente de la IA en la medicina ha dado lugar a numerosas problemáticas éticas y jurídicas, que deben ser abordadas para garantizar un uso responsable y beneficioso de esta tecnología. Por ello, en este trabajo, se realizará una visión actual sobre el impacto de la IA en la biomedicina y se analizan las estrategias y legislaciones existentes para hacer frente a estas problemáticas.

Palabras clave: Inteligencia Artificial, aprendizaje automatizado, biomedicina, ley, ética

Abstract:

Nowadays, the use of artificial intelligence (AI) has expanded to many areas of society, and medicine has been no exception. Automated learning algorithms have become an increasingly used tool in the biomedical field, due to their multiple applications, such as disease diagnosis and treatment, drug development, and hospital management, among others. However, this growing use of AI in medicine has given rise to numerous ethical and legal issues, which must be addressed to ensure a responsible and beneficial use of this technology. Therefore, this paper will provide a current overview of the impact of AI in biomedicine and analyze existing strategies and legislation to address these issues.

Key Words: Artificial intelligence, machine learning, biomedicine, law, ethics.

Índice de contenido

1.	Introducción.....	4
1.1.	Justificación del tema	4
1.2.	Objetivos.....	4
1.3.	Metodología.....	5
1.4.	Estructura.....	5
2.	Marco conceptual	6
2.1.	Definición y evolución de la Inteligencia Artificial	6
2.2.	Algoritmos de aprendizaje automatizado e Inteligencia Artificial	8
2.2.1	Tipos de APAs.....	8
2.2.2	Técnicas de los APAs	9
2.2.3.	Avances de los APAs en ámbito de la biomedicina	12
3.	Desafío jurídico-Imputación de la responsabilidad	18
3.1.	Imputación de la responsabilidad de manera individualizada	20
3.1.1.	Creadores del algoritmo.....	20
3.1.2.	Médico que los utiliza.....	21
3.1.3.	Hospital que los aplica.....	22
3.1.4.	Estado	23
3.1.5.	Paciente.....	25
3.2.	Responsabilidad compartida-Propuesta UE	26
4.	Parte ética	30
4.1.	Libertad.....	32
4.2.	Igualad	40
4.2.1.	Algoritmos opacos.....	40
4.2.2	APAs e inclusión social.....	41
4.2.3.	Posibles soluciones para algoritmos menos discriminatorios.....	44
4.3.	Fraternidad.....	46
4.3.1.	. APAs y aumento brecha social en el acceso a la medicina	47
4.3.2.	APAs y disminución brecha social en el acceso a la medicina	49
5.	Conclusiones.....	50
6.	Bibliografía.....	53

1. Introducción

1.1. Justificación del tema

La inteligencia artificial y los algoritmos de aprendizaje automatizado han tenido un impacto significativo en nuestra sociedad, transformando la forma en que trabajamos y tomamos decisiones. Desde la regulación de piensos en la ganadería hasta el marketing personalizado, estas tecnologías han demostrado ser herramientas poderosas para resolver una amplia variedad de problemas.

Sin embargo, a medida que las aplicaciones de la inteligencia artificial y los algoritmos de aprendizaje automatizado se vuelven más sofisticadas, es esencial ser conscientes de sus limitaciones y de las consecuencias negativas que pueden surgir de un uso irresponsable. En particular, el uso de estos algoritmos en el ámbito de la salud es especialmente preocupante debido a las posibles consecuencias negativas para la salud de las personas.

Este trabajo tiene como objetivo realizar una visión general de las limitaciones de los algoritmos de aprendizaje automatizado en el ámbito de la salud y reflexionar sobre su situación actual en la legislación y los problemas éticos que pueden surgir en su aplicación.

1.2. Objetivos

En este trabajo se pretende profundizar en los siguientes tres objetivos:

- Investigar sobre los principales avances que ha supuesto el uso de los algoritmos de aprendizaje en el ámbito de la medicina
- Realizar un análisis de las problemáticas jurídicas y sus posibles soluciones que supone el uso de los algoritmos de aprendizaje automatizado en el ámbito de la

medicina, destacando el problema de imputación de la responsabilidad jurídica y el papel de la Unión Europea.

- Realizar un análisis de las problemáticas éticas y sus posibles soluciones que supone el uso de los algoritmos de aprendizaje automatizado en el ámbito de la medicina, desde los tres grandes valores morales de la moralidad: la libertad, la igualdad y la fraternidad.

1.3. Metodología

A fin lograr los objetivos previamente establecidos se ha realizado una investigación bibliográfica exhaustiva, utilizando principalmente tesis académicas, artículos de investigación y libros. Además, para las implicaciones jurídicas de la IA en el mundo de la biomedicina, se ha prestado especial intención al papel de la UE, analizando las directrices, reglamentos, propuestas de reglamentos e informes que esta ha realizado.

1.4. Estructura

Una vez han sido establecido, la motivación y relevancia del presente trabajo, junto con los objetivos que se pretenden cumplir y la metodología utilizada para ello, se procederá a explicar la estructura seguida por el presente trabajo.

El presente trabajo, está dividido en tres capítulos, cada uno de los cuales corresponde a un objetivo. En el **primer capítulo** de este trabajo consistirá en un marco conceptual. En él, se abordará la aclaración de los términos técnicos utilizados en el campo de la IA, y se evaluarán los principales avances que han supuesto el uso de los algoritmos de aprendizaje automatizado en la medicina

En el **segundo capítulo**, se abordará una de las principales problemáticas jurídicas en torno al uso de APAs en la medicina: la imputación de responsabilidad jurídica. Para ello, se analizarán las diferentes implicaciones de imputar la responsabilidad de manera individual a cada participante involucrado en la creación y aplicación de APAs en la biomedicina. Además, se examinará la solución propuesta por la Unión Europea, sobre una responsabilidad compartida y el impacto de esta solución en nuestra legislación.

Posteriormente, en el **tercer capítulo** se abordarán las problemáticas éticas y sus posibles soluciones relacionadas con el uso de APAs en la biomedicina. Para ello, se evaluará el impacto que han tenido los APAs en los tres grandes valores morales de la moralidad: la libertad, la igualdad y la fraternidad. Además, se examinarán las estrategias y legislaciones existentes para mitigar sus consecuencias negativas. Por ello, en este capítulo también se evaluarán implicaciones jurídicas como la privacidad de las personas, el tratamiento de la información personal o el derecho a la no discriminación.

Finalmente, se realizarán las **conclusiones** de este trabajo, con el objetivo de evaluar el cumplimiento de los objetivos propuestos, establecer los principales obstáculos en la elaboración de este trabajo y posibles líneas de investigación.

2. Marco conceptual

2.1. Definición y evolución de la Inteligencia Artificial

La Unión Europea (2018) define la Inteligencia Artificial (IA) como aquellos “sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción –con cierto grado de autonomía– con el fin de alcanzar objetivos específicos” (Unión Europea, COM(2018) 237 final). Esta definición hace referencia a IA específica, la cual utiliza técnicas y algoritmos para realizar tareas limitadas y automatizar procesos (Porcelli, 2020). Actualmente toda la IA se considera específica y por ende es donde se han producido todos los avances. Entre ellos destacan programas como *Google Translate*, *Alpha*

Go, coches autónomos o aplicaciones para predecir enfermedades con mayor precisión que el médico promedio (Porcelli, 2020).

Por su parte la Inteligencia Artificial General hace referencia a la existencia de una tecnología con capacidad de observar, analizar y reaccionar de manera similar o superior al ser humano en todas las áreas, incluyendo el pensamiento abstracto, la resolución de problemas o la toma de decisiones. Aunque puede parecer ciencia ficción, las predicciones actuales apuntan a que la Inteligencia Artificial General será alcanzada en las próximas décadas, en concreto entre los años 2030 y 2050 (Porcelli, 2020).

Es relevante tener en cuenta que los avances en el campo de la inteligencia artificial son continuos y rápidos, lo que significa que la transición de la IA específica a la general no será instantánea. Por lo tanto, se necesitará una actualización constante de la legislación que regula el papel de la inteligencia artificial en nuestra sociedad, tal y como han señalado Haenlein y Kaplan (2019).

En resumen, según la Unión Europea la IA son aquellos sistemas capaces de tomar decisiones autónomas y de manifestar un comportamiento inteligente para alcanzar objetivos específicos. Aunque actualmente toda la IA se considera específica, la cantidad de avances diarios no descarta la posibilidad de una transición a una IA general en un futuro cercano. Sin embargo, para estar preparados ante estos cambios, se necesitará una legislación que pueda recoger y regular todas las implicaciones que conlleva el avance de la Inteligencia Artificial en nuestra sociedad.

2.2. Algoritmos de aprendizaje automatizado e Inteligencia Artificial

Debido a la amplitud del campo la inteligencia artificial, excesiva para ser materia de un trabajo fin de grado, el presente trabajo se centrará en aspectos concretos y más reducidos de ese campo, como son los algoritmos de aprendizaje automatizado (o automático, en adelante “APAs”) en el mundo de la biomedicina.

Los algoritmos consisten en una serie de pasos o rutinas utilizados para el entrenamiento de un modelo, el cuál debe estar previamente definido (Aracena et al., 2022). Por su parte, el aprendizaje automatizado o *machine learning* es una rama de la IA que consiste en aquellos algoritmos que “permiten a las máquinas aprender a partir de los datos y mejoran su rendimiento en una tarea específica sin ser explícitamente programadas para hacerlo” (Müller y Bostrom, 2016, p.15). En otras palabras, consisten en aquellas técnicas que permiten a las máquinas mejorar de forma autónoma con la experiencia y tener una mayor capacidad para imitar comportamiento humano. Los APAs han sido usados para la mayoría de la IA, aunque existen determinadas áreas como ciertos aspectos de la robótica y la telemedicina en las que no es precisas de su uso (Aracena et al., 2022).

2.2.1 Tipos de APAs

Los APAs pueden clasificarse en dos grupos: de aprendizaje supervisado y algoritmos de aprendizaje no supervisado.

Los algoritmos de aprendizaje supervisado son utilizados cuando es conocido el resultado que debe predecir el modelo y es entrenado con un conjunto de datos establecidos a tal efecto. El objetivo es que el modelo encuentre los patrones de los datos, para así poder crear un modelo que realice predicciones más precisas (Salehi & Burgueño, 2018). En medicina poseen una gran utilidad ya que su uso ha permitido detectar determinadas enfermedades con alta sensibilidad y especificidad. Un ejemplo de ello es un modelo de aprendizaje supervisado

entrenado a partir de imágenes de fondo de ojo que ha detectado la retinopatía diabética con una sensibilidad de 94,6% y una especificidad del 74,3% (Aracena et al., 2022).

Los algoritmos de aprendizaje no supervisado se utilizan cuando el modelo es alimentado con una cantidad masiva de datos sin etiquetar. Este aprendizaje realiza un reconocimiento de los patrones ocultos que poseen los datos, sin tener una respuesta conocida que pueda utilizarse para guiar el proceso. El aprendizaje no supervisado se utiliza a menudo para agrupar los datos en categorías (Salehi & Burgueño, 2018). Ejemplo de ello es su uso para la identificación de patrones en los datos genómicos y la agrupación de los pacientes en subtipos de enfermedades, lo que puede ser útil para personalizar el tratamiento y mejorar los resultados clínicos (Hajirasouliha & Elemento, 2020).

En resumen, los **APAs** se pueden clasificar en supervisados y no supervisados en función de si tienen conocimiento previo del resultado que deben predecir o no. Si lo tienen serán considerados de aprendizaje supervisado y en caso contrario de aprendizaje no supervisado.

2.2.2 Técnicas de los APAs

Independientemente de que los **APAs** precisen de supervisión o no, podrán seguir diferentes técnicas en su funcionamiento. Cada técnica tiene sus propias ventajas y limitaciones, y la elección adecuada dependerá de las características de los datos y los objetivos de la tarea. Además, un mismo algoritmo puede seguir diferentes técnicas, por lo que en general no son excluyentes entre sí (Aracena et al., 2022). A continuación, explicaremos brevemente aquellas técnicas más relevantes para nuestro trabajo.

Los árboles de decisión o *decision trees (DTs)* son una técnica utilizada en el aprendizaje automático para la toma de decisiones, que consiste en realizar una clasificación de los datos a partir de una serie de preguntas sobre sus características. Los árboles de decisión están formados por nodos que pueden ser internos o de hoja y que se utilizan para la estructuración

de datos. Cada nodo interno representa una pregunta sobre una característica de los datos y cada hoja representa una clasificación. El algoritmo toma el conjunto de datos y busca la mejor característica/pregunta para dividir los datos en dos conjuntos más pequeños y homogéneos. Luego, sigue dividiendo los datos hasta que se llega a un punto donde se puede tomar una decisión. El objetivo es construir un árbol que sea lo suficientemente complejo como para representar los patrones en los datos, pero no tanto como para ser demasiado específico y ajustado a los datos de entrenamiento (Zhang et al., 2017).

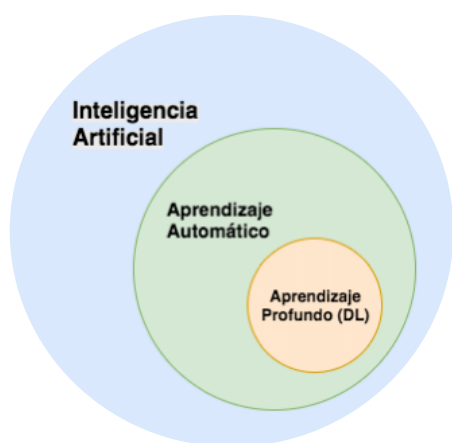


Figura adaptada de: Hernán Ordiales & Barrera Gabriel, 2018

Por su parte los bosques aleatorios o *random forest (RF)* son técnicas por las que se construyen múltiples árboles de decisión y se combinan los resultados de cada uno para dar la predicción final. Cada árbol de decisión toma una decisión basada en un subconjunto aleatorio de características del conjunto de datos de entrenamiento (Aracena et al., 2022). Cuando se combinan los resultados de todos los árboles, se reduce el riesgo de que el modelo se ajuste demasiado a los datos de entrenamiento. De esta manera, el Random Forest ayuda a prevenir el sobreajuste y realiza predicciones más precisas (Zhang et al., 2017). Un ejemplo de su aplicación en biomedicina fue su uso para modelar la afinidad de unión proteína-ligando¹ en diferentes complejos (Zhang et al., 2017).

Los APAs también pueden utilizar la técnica de *Naive Bayes* para la clasificación de datos. Realizada una clasificación probabilística, basada en el teorema de Bayes, donde la clasificación de las nuevas observaciones en diferentes categorías se realiza en función de la probabilidad de que pertenezcan a cada categoría. La técnica es ingenua o *naive* porque

¹ La afinidad de unión proteína ligando es la fuerza de interacción entre una proteína y un ligando (una sustancia que se une específicamente a la proteína) en un complejo.

supone que todas las características son independientes entre sí, lo que puede no ser cierto en la realidad (Berrar, 2018). Actualmente ha demostrado su potencial para la predicción de inasistencia hospitalaria (Aracena et al., 2022).

Por último, los algoritmos que emplean aprendizaje profundo o *deep learning (DL)* se enfocan en el entrenamiento de redes neuronales artificiales para resolver problemas complejos de forma eficiente. Para ello las redes neurales artificiales pretenden imitar el funcionamiento del cerebro humano (Aguirre, et al., 2021). El cerebro humano está formado por miles de millones de células llamadas neuronas, que se conectan entre sí formando redes neuronales. Las neuronas se comunican entre sí a través de sinapsis, que son pequeños espacios entre las células. Estas conexiones sinápticas forman redes neuronales complejas que permiten al cerebro procesar información y tomar decisiones (Alcami et al., 2019).

De manera similar, en las redes neuronales artificiales se utilizan nodos artificiales que se conectan entre sí para formar una red. Cada nodo se comporta como una neurona y procesa la información, a través de un cálculo matemático, que recibe de los nodos conectados a él. La información se transmite de un nodo a otro a través de conexiones ponderadas, similares a las sinapsis del cerebro humano (Salehi & Burgueño, 2018). Cada conexión entre dos nodos tiene un peso asociado que indica la importancia de la conexión. A través del aprendizaje de estos pesos, la red neuronal puede ajustar su comportamiento para realizar tareas específicas, como la clasificación de la información o la detección de enfermedades por imágenes (Aguirre, et al., 2021). Esta técnica posee un alto potencial y ha sido utilizada en la mayoría de los avances de la IA en las últimas décadas. Ejemplo de ello son los algoritmos utilizados para el Procesamiento del Lenguaje Natural, los cuales poseen valor para extraer la información relevante de cada paciente de los registros clínicos electrónicos (Aracena et al., 2022).

En resumen, los algoritmos de aprendizaje automático (APAs) pueden emplear diversas técnicas para llevar a cabo la clasificación y predicción de datos. Algunas de las técnicas más

destacadas son los árboles de decisión, los bosques aleatorios, *Naive Bayes* y el aprendizaje profundo. Su correcta elección es relevante, ya que puede tener un gran impacto en la eficacia y precisión de los resultados obtenidos (Aracena et al., 2022).

2.2.3. Avances de los APAs en ámbito de la biomedicina

Actualmente los APAs han sido utilizados en la mayoría de los ámbitos de la sociedad, desde su uso para la regulación del consumo de pienso en la ganadería, hasta la publicidad personalizada en el marketing (Unión Europea, COM(2018) 237 final). Aun así, debido al interés de este trabajo y a la imposibilidad de abarcar en profundidad todas sus implicaciones, nos centraremos en sus aplicaciones en el ámbito de la biomedicina. Para ello, analizaremos el alcance que ha tenido el uso de los APAs en la medicina personalizada, en el desarrollo de medicamentos, en la detección de enfermedades y en la gestión hospitalaria, para determinar el alcance que tienen en la salud de las personas.

2.2.3.1. Medicina personalizada y desarrollo de medicamentos.

Tradicionalmente, los nuevos tratamientos y diagnósticos de enfermedades han sido desarrollados para la población en general (Hajirasouliha & Elemento, 2020). Sin embargo, la medicina personalizada supone un nuevo enfoque que considera al paciente de manera individual y tiene en cuenta su variabilidad genómica, epigenómica y metabólica, así como su ambiente y estilo de vida. Este tipo de medicina permite ofrecer a cada individuo el mejor tratamiento basado en su singularidad genética y ha supuesto un avance en la mayoría de los campos de la medicina. Ahora bien, el uso de inteligencia artificial y algoritmos de aprendizaje automático permite considerar de manera más eficiente todas las variables específicas de un individuo (Liang et al., 2020). Ejemplo de ello es cómo el uso de los APAs, especialmente aquellos que emplean la técnica del bosque aleatorio, ha permitido determinar la sensibilidad a los fármacos de pacientes con cáncer de ovario y cáncer de endometrio. Además, se ha demostrado la eficacia de los modelos de aprendizaje automatizado para determinar cuál es la dosis óptima de quimioterapia según los datos de cada paciente (Xu et

al., 2019). Sin embargo, los APAs no poseen una fiabilidad completa y su uso inadecuado y sin supervisión médica, puede implicar consecuencias y afectar a la salud de las personas (Aracena et al., 2022).

Los APAs también se han utilizado para el desarrollo de medicamentos. El desarrollo de un nuevo medicamento implica crear una ruta que permita, a partir de otras moléculas más simples llegar a una molécula objetivo. Para lograrlo, es necesario analizar grandes cantidades de datos químicos, lo que puede ser complicado y llevar mucho tiempo. Sin embargo, el uso de los APAs permite procesar grandes cantidades de información química y encontrar las nuevas rutas para llegar a la molécula objetivo, lo cual acelera el proceso de descubrimiento de nuevos medicamentos (Hajirasouliha & Elemento, 2020).

En este sentido, cabe destacar el caso de *AlphaFold2*. *AlphaFold2* es un modelo desarrollado por la compañía *DeepMind* que utiliza algoritmos de aprendizaje profundo para predecir la estructura tridimensional de las proteínas con un alto grado de precisión, lo cual ha ayudado a avanzar en el problema del plegamiento de proteínas. El problema del plegamiento de proteínas se refiere a la dificultad que existe para predecir cómo una cadena de aminoácidos, estructura primaria de una proteína, se plegará para formar su estructura final, estructura terciaria de la proteína, la cual determinará su función biológica. Esta tarea es compleja debido a la gran cantidad de variables que intervienen en el proceso de plegamiento, incluyendo las interacciones químicas y la influencia del entorno celular (Alcami et al., 2019, pp.70-86). A partir de un enfoque de aprendizaje profundo basado en la predicción de la distancia entre pares de aminoácidos, *AlphaFold2* predice la estructura tridimensional con una precisión del 92,4% equiparable a la de los métodos tradicionales. Además, realiza la predicción de manera más rápida y menos costosa en comparación a los métodos tradicionales, al no requerir tanto personal ni instalaciones para la cristalografía de rayos X". Sin embargo, es relevante tener en cuenta que *AlphaFold2* es sólo una herramienta y no ofrece una respuesta al problema del plegamiento de proteínas o asegura de manera certera un desarrollo de nuevos medicamentos sin efectos secundarios (Jumper et al., 2021).

En resumen, el uso APAs y su desarrollo ha permitido tratar a cada paciente de manera individualizada y acelerar el desarrollo de nuevos medicamentos, generando una medicina más eficiente. Aun así, los APAs son herramientas y como tal su uso posee limitaciones y puede implicar consecuencias graves en la salud de las personas.

2.2.3.2. APAs en la detección de enfermedades

En los últimos años, el uso del aprendizaje automático ha facilitado la detección de enfermedades a los médicos, especialmente en el ámbito de la radiología y en la patología. Su uso ha sido especialmente relevante para la mejora de la calidad de imágenes radiológicas, y la segmentación y detección de lesiones no visibles al ojo humano (Aguirre, et al., 2021).

Los APAs, en concreto aquellos que utilizan la técnica del *Deep learning*, han sido capaces de mejorar la calidad de imágenes radiológicas. Normalmente para la obtención de imágenes radiológicas es necesario suministrar una dosis de radiación al paciente. Sin embargo, debido a los efectos negativos que casusa una alta radiación, los médicos suministran una dosis mínima necesaria que les permita obtener una imagen radiológica interpretable (Moreno Berdón, 2021). El problema, es que conforme menor sea la radiación peor resolución presentará la imagen, lo cual dificulta su interpretación por parte del médico. Además, la imagen radiológica también se puede ver comprometida por la calidad del equipo utilizado para su obtención (Aguirre, et al., 2021).

Una de las maneras de mejorar la resolución de la imagen es la utilización de algoritmos de aprendizaje profundo para reducir el ruido de la imagen radiológica, el cual se refiere a la “aparición aleatoria de patrones o granulaciones en la imagen que no corresponden a la estructura anatómica real que se está visualizando” (Moreno Berdón, 2021). Para ello, se ha entrenado una red neuronal profunda con un conjunto de imágenes radiológicas que contienen ruido. Primero la red neuronal realiza una identificación de aquellos patrones y

características necesarias que le permita identificar el ruido de la imagen para su posterior separación. Una vez que la red ha sido entrenada, se puede utilizar para procesar imágenes radiológicas nuevas y eliminar el ruido de la imagen. Esta técnica permite potenciar los recursos técnicos y físicos disponibles, ya que no requiere cambiar los equipos de obtención de imágenes existentes por versiones más avanzadas. Además, también evita que los pacientes se sometan a una cantidad de radiación innecesaria y facilita la interpretación de las imágenes al facultativo médico (Aguirre, et al., 2021).

En la figura 1 se muestra un ejemplo una imagen mejorada a partir del aprendizaje profundo. Primero se realiza el *padding*, que es un ajuste de la imagen, para adaptarlo al tamaño requerido del modelo entrenado, el cual nos dará la imagen inicial con mayor calidad y en su tamaño original.

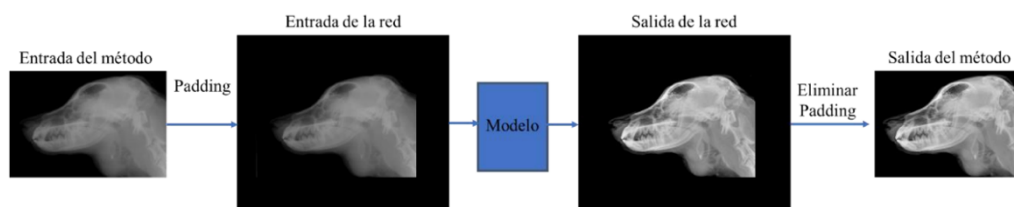


Figura 1

Fuente: Moreno Berdón, 2021

En relación con los algoritmos utilizados para la segmentación y detección de lesiones la Inteligencia Artificial ha permitido observar que son invisibles para el ojo humano y que los métodos convencionales no han sido capaces de detectar. En este sentido, destaca la utilización del software *Radiomics*, el cual ha sido especialmente eficaz para la segmentación de imágenes tumorales, gracias a su capacidad para detectar lesiones intraparenquimatosas ² (Aguirre et al., 2021). Para ello, *Radiomics* utiliza APAs para extraer la información numérica de las imágenes médicas, con el fin de obtener información cuantitativa y cualitativa de dichas imágenes. Esta técnica analiza la textura, forma y otros detalles en una

² Que está en el interior del parénquima (tejido esencial de determinados órganos) de un órgano

imagen médica, los cuales pueden contener información valiosa sobre la biología subyacente del tejido. *Radiomics* ha demostrado ser efectiva en la reducción de los falsos positivos en la detección de tumores en comparación con los métodos tradicionales, lo cual es especialmente beneficioso en el tratamiento de enfermedades graves como el cáncer (Aguirre et al., 2021).

En resumen, el uso de APAs ha demostrado ser efectivo en mejorar la calidad de las imágenes radiológicas y en la detección de lesiones que pueden ser difíciles de identificar según el uso de los métodos tradicionales. Estas aplicaciones han ayudado al facultativo médico a realizar diagnósticos más precisos y tempranos, lo que suele tener un impacto positivo en la salud y la calidad de vida de los pacientes.

2.2.3.3. APAs para la gestión hospitalaria

En este apartado se pretende analizar si el uso de los APAs permite una gestión hospitalaria más eficiente. Para ello se analizará su tasa de éxito para determinar la inasistencia de los pacientes y la mortalidad intrahospitalaria. Además, se evaluará las consecuencias de utilizar los APAs para el procesamiento del lenguaje natural de los registros electrónicos. (Aracena et al., 2022).

En España, se estima que alrededor del 14% de los pacientes no asisten a sus citas médicas, lo que puede llegar a suponer un gasto mayor de tres millones según el hospital (Mesa et al., 2017). Por ello, la existencia de algoritmos de aprendizaje automático para predecir la inasistencia de pacientes en los hospitales es muy valiosa. Actualmente, el mejor modelo es capaz de detectar el 82% de las inasistencias a las citas hospitalarias. Para lograr esto, el modelo ha considera variables como la cantidad de días entre la programación y la realización de la cita, las horas perdidas por el paciente en citas anteriores y el número de días desde la última cita. Además, utiliza una técnica de clasificación bayesiana que considera la probabilidad condicional (Aracena et al., 2022). Aunque el porcentaje de detección de

inasistencias aún no es del 100%, los resultados obtenidos por el modelo son prometedores y se espera su mejora (Aracena et al., 2022).

Además, cada vez son de mayor uso los APAs para determinar la inminencia de una muerte intrahospitalaria. Estos algoritmos operan a través de árboles de decisión entrenados a partir de la información existente de los pacientes. Esta información puede variar desde el sexo y edad, hasta las constantes vitales actuales del paciente, que al ser cambiantes implicaría una actualización continua del modelo (Aracena et al., 2022). En la actualidad estos modelos han tenido altas tasas de acierto, sin embargo, no son completamente fiables y debe seguir prevaleciendo la opinión del médico (Karcher, 2018).

Por último, la utilización de APAs que utilizan técnicas de *Deep learning* ha permitido el procesamiento de lenguaje natural en el texto clínico integrado de los registros médicos electrónicos. Gracias a ello se puede realizar una extracción y clasificación automática de la información relevante de los pacientes, lo cual facilita el diagnóstico a los facultativos médicos y les libera de carga profesional (Aguirre et al., 2021). También, ha facilitado que pueda compartirse información médica de manera más rápida y segura entre los diferentes profesionales y centros de salud, lo cual ha mejorado la coordinación y continuidad en la atención médica de los pacientes, evitando errores y redundancias en el proceso (Aracena et al., 2022).

En resumen, se ha comprobado que los APAs tienen un gran potencial para mejorar significativamente la gestión hospitalaria en términos de tiempo y dinero. Esto se debe a su capacidad para predecir la inasistencia de los pacientes a sus citas o la inminencia de su muerte dentro del hospital. Aunque estos modelos predictivos aún no tienen una tasa de acierto completa, sus resultados obtenidos son prometedores. Además, el uso de APAs para el procesamiento del lenguaje natural ha permitido la extracción automática de información relevante de cada paciente, lo que ha mejorado la comunicación entre la comunidad científica y ha aliviado la carga de trabajo del personal médico y administrativo del hospital.

Todos estos estudios y ejemplos indica la importancia cada vez mayor del uso APAs en el ámbito de la biomedicina. Sin embargo, los APAs son una herramienta que aún se encuentra en desarrollo y su aplicación suele presentar diversas problemáticas. Entre ellas, se destacan no sólo la posible imprecisión de los resultados de los algoritmos, sino también la aparición de problemas éticos y jurídicos (López Baroni, 2019), las cuales serán abordadas en el presente trabajo.

Para ello se comenzará realizando un análisis jurídico, en el cual se abordará la imputación de la responsabilidad jurídica por la utilización incorrecta de estos algoritmos en el ámbito de la salud. En este análisis se evaluarán los posibles sujetos a los que se les puede analizar la imputación y se destacará la solución propuesta por la UE. A continuación, se examinarán las principales consecuencias éticas detrás del uso de los APAs en el ámbito de la biomedicina. Estas consecuencias serán evaluadas bajo los principios de la modernidad: “libertad, igualdad y fraternidad” (Ortega, 2006) y abordará cuestiones como la falta de transparencia en el funcionamiento de los algoritmos, la presencia de sesgos en los resultados o la desigualdad en el acceso a los avances en inteligencia artificial. Además, se estudiarán las distintas propuestas para mitigar el impacto de las consecuencias éticas, donde volverá a destacar la legislación de la UE.

3. Desafío jurídico-Imputación de la responsabilidad

Una vez analizados algunos de los avances más relevantes de los APAs en el mundo de la biomedicina, se pasará a realizar un análisis jurídico sobre las implicaciones de su uso en este ámbito. Aunque el uso de los APAs en la biomedicina plantea amplias implicaciones jurídicas, este capítulo se focalizará en la imputación de la responsabilidad jurídica por la utilización incorrecta de estos algoritmos. Esto se debe a la imposibilidad de abordar todas las cuestiones legales en profundidad. Además, otras cuestiones jurídicas relevantes, como

el respeto a la privacidad de las personas, el tratamiento de la información personal o el derecho a la no discriminación, serán tratadas posteriormente en el análisis ético.

La regulación de la inteligencia artificial es una tarea compleja y en constante evolución. En este contexto, la Unión Europea ha asumido un papel pionero, destacando los reglamentos y directivas como elementos clave en la creación de un marco legal para esta tecnología en Europa (Leucona, 2020). Sin embargo, a pesar de sus esfuerzos legislativos, la UE aún no ha definido oficialmente quién debería ser considerado responsable jurídico por el uso inadecuado de la inteligencia artificial y, por lo tanto, de los algoritmos de aprendizaje automatizado. Esto se debe a que los mecanismos habituales usados por el derecho para atribuir personalidad y responsabilidad jurídica no pueden aplicarse de manera exacta para los algoritmos detrás de la inteligencia artificial y las máquinas que los utilizan (Lain Monayo, 2020).

En este sentido, el sistema tradicional de derecho diferencia entre la condición persona física y persona jurídica (Ruiz de Huidobro, 2015). Por un lado, todas las personas físicas poseen personalidad debido al derecho natural y además el Ordenamiento jurídico español realiza un reconocimiento de la personalidad jurídica al establecer su adquisición con el nacimiento (BOE, Art. 29, 1889). Sin embargo, para que una persona jurídica posea personalidad es necesario su otorgamiento, a parte de su reconocimiento, por parte de la ley. Esto implica la existencia de requisitos mayores, los cuales están íntimamente relacionados con el derecho societario. En otras palabras, las sociedades tendrán personalidad jurídica en función de aquello establecido por la ley, como establece el artículo 35 del Código Civil (BOE, Art. 35, 1889).

Está claro que para que la Inteligencia Artificial, tuviese personalidad jurídica sería por otorgamiento y no por reconocimiento. Esto se debe a la falta de derecho natural de la IA, al ser esta un ente sin consciencia y una herramienta creada por los humanos para realizar tareas

específicas (López Baroni, 2019). Hasta la fecha, ninguna ley ha otorgado personalidad jurídica a la IA. Por lo tanto, ¿a quién deberíamos imputar la responsabilidad jurídica por el uso incorrecto de los APAs?

Debido al interés de nuestro trabajo, esta pregunta será respondida para el caso de los APAs aplicados en el ámbito de la salud. Para ello, se analizarán las diferentes implicaciones de imputar la responsabilidad jurídica de manera individual a cada participante involucrado en la creación y aplicación de APAs en la biomedicina. Se evaluará la posible responsabilidad jurídica de los programadores, al médico que utiliza el APA, al hospital que obliga a su aplicación, al Estado y al paciente que opta por seguir el resultado dado por el uso de APA, en vez del diagnóstico médico tradicional. Posteriormente se examinará la solución propuesta por la Unión Europea, la cual implica una responsabilidad compartida y posee un mayor impacto en nuestra legislación (Andrés Segovia, 2021).

3.1. Imputación de la responsabilidad de manera individualizada

3.1.1. Creadores del algoritmo

Existen varios argumentos que defienden que la responsabilidad jurídica en la aplicación de APAs en el ámbito de la biomedicina, así como en otros ámbitos, debería recaer en los creadores de algoritmos o programadores (Grote & Berens, 2020).

Los creadores del algoritmo son quienes tienen el conocimiento y la experiencia para desarrollar un sistema seguro y éticamente correcto. Por ello, son los principales responsables de asegurarse de que los datos de entrenamiento del modelo sean precisos y representativos de la población, y de esta manera evitar que se produzcan resultados discriminatorios (Grote & Berens, 2020). Además, tienen la capacidad de monitorear y actualizar su sistema, lo que les permite corregir los errores y mejorar la precisión de los resultados (Lain Moyano, 2021).

Además, los algoritmos de aprendizaje automático son opacos. Parte de su opacidad, es permanente e incomprensible para cualquier ser humano. Sin embargo, otra parte de la opacidad es comprensible para aquellas personas que posean una base matemática-física suficiente, es decir los programadores (Grote & Berens, 2020). Ellos poseen un conocimiento mayor que los facultativos médicos sobre el funcionamiento de los algoritmos y por lo tanto deberían ser los responsables jurídicos ante cualquier problema que surja en la interpretación y aplicación de dichos algoritmos (Laín Moyano, 2021).

En resumen, existen argumentos por los que los programadores deberían ser los responsables jurídicos por el uso incorrecto de APAs. Estos argumentos se basan en su capacidad para desarrollar y monitorear sistemas seguros y éticos, y su conocimiento mayor sobre el funcionamiento de estos (Grote & Berens, 2020).

3.1.2. Médico que los utiliza

Uno de los motivos por lo que se considera que el médico debe ser el responsable jurídico de los resultados dados por los APAs es la importancia de su figura y el hecho de que a él le corresponde la decisión final sobre el tratamiento del paciente (Molnar-Gabor, 2020). Sin embargo, cuando se utiliza un APA para el diagnóstico de enfermedades, la responsabilidad del médico dependerá de la razón por la cual se haya producido un diagnóstico erróneo, lo cual puede ser debido a dos supuestos diferentes (Ramón Fernández, 2021).

El primero se produce cuando el algoritmo actúa de manera defectuosa. Según la Resolución del Parlamento Europeo, del 20 de octubre de 2020, en estos casos la responsabilidad no recaerá en el médico. En su lugar, habrá una responsabilidad solidaria o subsidiaria, debido a la existencia de vicios ocultos, que será compartida con los sujetos que comercializaron el algoritmo. Esta responsabilidad solidaria pretende asegurar que los desarrolladores y

proveedores de tecnología también asuman su parte de responsabilidad cuando se perjudica la salud de los pacientes. (Ramón Fernández, 2021).

El segundo supuesto es aquel producido por una incorrecta interpretación por parte del facultativo médico. En estos casos existen dos tipos de posturas. La primera de ellas es considerar que el médico es el responsable jurídicamente de los daños producidos por el algoritmo debido a su falta de diligencia (Ramón Fernández, 2021). La segunda, con más apoyo, considera que la responsabilidad solo recaerá en el médico si ha vulnerado el nivel de diligencia mínimo requerido. El motivo es la falta de base actual para la correcta interpretación de los algoritmos de los médicos. Por ello, hasta que no reciban la formación académica precisa, en la carrera o en cursos formativos posteriores, no se les puede exigir esta responsabilidad (Pucchio et al., 2020).

En resumen, la responsabilidad jurídica ante el diagnóstico erróneo producido por la aplicación de un APA varía en función del motivo del error. Por un lado, si el error es debido a un APA defectuoso habrá una responsabilidad compartida por los sujetos detrás de su comercialización (Ramón Fernández, 2021). Por otro, si el error se debe a una incorrecta interpretación del médico, deberá analizarse si ha actuado dentro de los mínimos de diligencia requeridos de la práctica médica para poder imputarle responsabilidad (Pucchio et al., 2020).

3.1.3. Hospital que los aplica

En este caso la asignación de la responsabilidad es más fácil, al menos en España, debido a que la Ley establece cuándo un hospital debe tener personalidad jurídica y su correspondiente responsabilidad jurídica (Ley 14/1986, de 25 de abril, General de Sanidad).

Considerando que el algoritmo está siendo utilizado por un centro clínico privado, se nos presentan dos situaciones. La primera es que el hospital haya contratado el uso del algoritmo con una empresa. En este caso la asignación de la responsabilidad será una cuestión

contractual, es decir, lo que establezca el contrato. La segunda situación es cuando el algoritmo está disponible para uso público y el hospital ha decidido implementarlo. En estos casos, la responsabilidad recaerá sobre el hospital, ya que la mayoría de los algoritmos de uso público suelen advertir que su utilización no siempre produce los resultados deseados, y eximen de responsabilidad a los desarrolladores del algoritmo (Ada Lovelace Institute, 2021).

Además, si el hospital ha establecido en su política sanitaria el uso obligatorio de APAs, en caso de error, la responsabilidad recaería en el hospital y no en el médico. Esto se debe a que la implementación de la política hospitalaria implica una relación contractual entre el hospital y los médicos y demás trabajadores del hospital donde los gerentes del hospital establecen instrucciones que deben ser seguidas por los empleados (Ley 14/1986, de 25 de abril, General de Sanidad).

Diferente asunto es que el algoritmo sea de obligada aplicación debido a un mandato del Estado. Hoy en día este supuesto es una cuestión ficticia, pero sirve para dar pie a la cuestión de si el Estado algún tipo de responsabilidad jurídica en la utilización de los APAs.

3.1.4. Estado

La imputación de la responsabilidad al Estado es una cuestión mucho más política que legal y depende del papel que queramos atribuir al Estado.

Por un lado, aquella parte de la población cuya ideología este más orientada a que el Estado debe ejercer una función paternalista, considerará correcto que asuma responsabilidad jurídica sobre las consecuencias que pudiera acarrear la utilización de estos algoritmos de aprendizaje automatizado en el mundo de la biomedicina. Además, este sector de la población

defenderá la existencia de una regulación muy específica que controle dichos avances (Ada Lovelace Institute, 2021).

Por otro lado, aquellos de ideología más liberal, abogaran por una menor intervención del Estado. En otras palabras, considerarán que la responsabilidad deberá imputarse a cualquier sujeto menos al Estado, es decir, a las empresas que crean los **APAs**, a los hospitales que han decidido utilizarlos, a los médicos que no han confiado en su criterio o a los pacientes que han preferido fiarse de la IA en vez de los médicos (Schwartz, 2021).

En el caso de España, no nos deberemos olvidar de su pertenencia a la Unión Europea, por lo que se verá influencia por el LIBRO BLANCO sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza_ elaborado en 2020 por la Comisión Europea. Según Andrés Segovia (2021) en este libro se establece la necesidad de que el “el sector público (a nivel nacional y de la EU) y el sector privado aúnen esfuerzos a fin de incrementar gradualmente las inversiones globales [...] y convertir a la IA en una [...] herramienta para el progreso público-privado y económico que afectará a prácticamente todos los ámbitos”(Andrés Segovia, 2021).

En resumen, la imputación de la responsabilidad al Estado en el uso de APAs en el ámbito de la biomedicina es una cuestión en principalmente ideológica en la que se defenderá una mayor responsabilidad jurídica del Estado cuando éste ejerza un papel más paternalista y viceversa. En España, destaca la influencia del LIBRO BLANCO que aboga por la intervención conjunta del sector público y privado para convertir la IA en una herramienta que permita el progreso económico y público-privado en diversos ámbitos, incluyendo la biomedicina.

3.1.5. Paciente

Para considerar el nivel de responsabilidad jurídica del paciente frente al uso inadecuado del APA, es necesario tener en cuenta su influencia en la decisión del tratamiento. Por ello, la determinación de su responsabilidad es una cuestión casual (Ramón Fernández, 2021).

Es relevante mencionar la evolución radical, especialmente en los últimos años, que ha habido en la relación médico-paciente. Anteriormente, la relación médico-paciente era de naturaleza vertical, en la cual el médico tomaba todas las decisiones sobre el tratamiento del paciente, sin necesidad de prestar el deber de información o de esperar el consentimiento del paciente. El médico ejercía un papel paternalista, influenciado por la sociedad del momento y por la falta de legislación de los derechos del paciente. (Lázaro & Gracia, 2006). Sin embargo, desde finales del siglo XIX y principios del siglo XX, se ha pasado a una relación de naturaleza horizontal, en la cual el médico ha dejado de tener un papel paternalista, y los derechos de los pacientes se han recogido legislativamente en gran parte de los países (Lázaro & Gracia, 2006). Ejemplo de ello es la Ley 41/2002, de 14 de noviembre, *básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica*. A partir de esta ley se reconoce el derecho del paciente a ser informado de una manera clara y concisa, pero también se le reconoce el derecho de rechazar el tratamiento en determinados casos, como establece en su artículo 2.4.

A pesar de que no exista una regulación oficial sobre ello, si el paciente decide seguir el criterio de un algoritmo en lugar del médico y lo hace de forma informada y libre de vicios, se le considerará responsable de los posibles resultados negativos que puedan ser causados por el algoritmo (Molnár-Gábor, F, 2020). Sin embargo, el paciente deberá tomar medidas para proteger sus derechos e intereses legítimos, de acuerdo con el artículo 22 del RGPD. Este artículo establece el derecho de las personas de no ser consideradas como objeto frente a una decisión automatizada. Por ello, antes de adaptarse a un tratamiento establecido por un APA el paciente debe tener al menos la capacidad de expresar su punto de vista y de impugnar la decisión del sistema automatizado, y también debe tener derecho a la intervención humana

por parte del responsable. En el contexto de un algoritmo utilizado para el diagnóstico de una enfermedad, el responsable puede ser el hospital, el médico, el programador o cualquier entidad que haya manejado la información médica y personal del paciente ((R (UE) n° 679/2016 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, art 4.7).

En resumen, desde finales del siglo XIX, la relación médico-paciente ha evolucionado de una naturaleza vertical hacia una naturaleza horizontal, lo que ha permitido al paciente tener mayor participación en su diagnóstico y la protección de sus derechos por ley. Aunque no existe una regulación oficial sobre ello, el paciente puede elegir seguir el criterio de un APA en lugar del médico. Aun así, su decisión no puede estar coaccionada y debe tomar las medidas necesarias para proteger sus derechos e intereses legítimos, ya que él será el responsable final de cualquier resultado negativo causado por este.

Es relevante destacar que la imputación de responsabilidad de manera individual a los sujetos que han intervenido en la creación y aplicación del APAs en el ámbito de la biomedicina es una solución deficiente. Esto se debe a que los algoritmos de aprendizaje automático son sistemas complejos que involucran múltiples actores y fuentes de datos, por lo que es difícil atribuir la responsabilidad a una solo de los actores. Además, los algoritmos pueden ser utilizados en una variedad de contextos y situaciones, lo que complica aún más la identificación de la fuente del problema y la asignación de responsabilidades (Ramón Fernández, 2021). Por ello la UE ha propuesto una solución basada en la responsabilidad compartida, que será analizada a continuación.

3.2. Responsabilidad compartida-Propuesta UE

Esta solución ha sido la preferida por la Unión Europea. Para analizar la opinión de la Unión Europea se ha considerado la Propuesta de Reglamento sobre inteligencia artificial de 2021, las Directrices Éticas para una Inteligencia Artificial Confiable de 2019 y el Reglamento General de Protección de Datos

La UE está tratando de evitar la situación de determinar quién es responsable de los resultados negativos que surgen del uso de la IA y los APAs detrás de ella. Para ello ha elaborado unas Directrices Éticas para una Inteligencia Artificial Confiable que pretenden ser una guía para mitigar las consecuencias negativas de la IA (Andrés Segovia, 2021). En ellas establece los principios y requisitos por los que debe regirse cualquier inteligencia artificial. Los cuatro principios son: respeto de la autonomía humana, prevención del daño, equidad, y explicabilidad. Existe una íntima relación entre ellos, al estar directamente vinculados a los derechos humanos, pero es relevante su diferenciación para que ninguno pierda su importancia. Por su parte los requisitos son siete: transparencia, inclusión, responsabilidad, robustez, privacidad, diversidad y bienestar social y medioambiental y son la concreción de los principios. Aun así, estas directrices se encuentran en el plano teórico y su aplicación no solo es imposible de realizar de manera completa, sino que también llevaría a contradicciones entre ellas (Villas & Camacho, 2022).

A pesar de los esfuerzos de la Unión Europea, es inevitable que el uso de APAs produzca en ocasiones resultados negativos. Por ello, la UE ha elaborado la Propuesta de Reglamento del Parlamento Europeo y del Consejo para establecer la imputación de responsabilidad a través de un análisis de rendición de cuentas (Reglamento de Inteligencia Artificial, 2021). Este análisis tiene como objetivo evaluar la cantidad de intervención humana en el funcionamiento de los algoritmos de aprendizaje automático, así como la capacidad de las personas para intervenir en ellos y comprender el método utilizado por los algoritmos (Villas y Camacho, 2022).

Como regla general, el análisis de rendición de cuentas servirá para establecer una responsabilidad compartida entre todas las partes involucradas en la creación, implementación y uso de sistemas de IA. Esto se debe a que, en la mayoría de los casos, el funcionamiento de un sistema de IA es el resultado de la colaboración de diferentes actores, como diseñadores, programadores, fabricantes, usuarios y reguladores, entre otros (Ramón

Fernández, 2021). Por lo tanto, el análisis de rendición de cuentas permitirá evaluar el grado de implicación y responsabilidad de cada uno de estos actores en el funcionamiento del sistema de IA en cuestión y “se podrán establecer medidas adecuadas para garantizar un uso seguro, ético y responsable de la IA, e implementar políticas de transparencia y supervisión adecuadas para prevenir o corregir los posibles riesgos y consecuencias negativas del uso de los sistemas de IA” (Villas y Camacho, 2022).

Independientemente de a quién se impute la responsabilidad jurídica, la UE ha destacado la importancia de reconocer el derecho de las personas de no ser consideradas como objeto frente a una decisión automatizada ((R (UE) nº 679/2016 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, art 22). Este derecho se encuentra recogido en el artículo 22 del Reglamento General de Protección de Datos (RGPD), donde se establece que las decisiones que afecten a los derechos y libertades de las personas no pueden ser tomadas exclusivamente por un algoritmo o sistema automatizado, sino que es necesaria la existencia de una intervención humana adecuada. La cantidad de intervención humana necesaria será aquella que permita la salvaguardia de los derechos e intereses legítimos de los interesados, es decir, de los individuos afectados por aplicación del sistema automatizado. Como estándares mínimos el RGPD ha establecido la posibilidad de que el interesado obtenga intervención humana por parte del responsable del tratamiento de sus datos personales, ya sea una entidad o persona física o jurídica. Además, el interesado debe poder expresar su punto de vista y tener la capacidad de impugnar la decisión proporcionada por el sistema automatizado. Es relevante destacar que estos estándares mínimos deben seguirse cumpliendo incluso cuando la decisión se basa en el consentimiento explícito del interesado, aunque en estos casos le corresponderá a él la adopción de las medidas adecuadas para salvaguardar sus derechos e intereses legítimos (Reglamento (UE) nº 679/2016 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, art. 7 y 22.4).

En resumen, la Unión Europea tiene un papel clave en la regulación y responsabilización del uso de los sistemas de IA. A través de su legislación, se han establecido los principios y

requisitos que deben reunir cualquier IA y se ha propuesto el análisis de rendición de cuentas como un medio para evaluar la implicación y responsabilidad de cada actor que ha intervenido en funcionamiento y aplicación de los sistemas de IA. Además, la UE ha destacado la importancia de proteger los derechos y libertades de las personas frente a las decisiones automatizadas, y ha establecido la necesidad de contar con una intervención humana adecuada en tales decisiones. Gracias a ello la UE ayuda a garantizar un uso seguro, ético y responsable de la IA ((R (UE) nº 679/2016 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, art 22).

En conclusión, el uso de algoritmos de aprendizaje automatizado en la biomedicina plantea una serie de implicaciones jurídicas que aún no han sido completamente resueltas. En este apartado se ha analizado la problemática de la imputación de la responsabilidad jurídica ante un uso incorrecto de los APAs en el ámbito de la salud. Actualmente no existe una respuesta oficial sobre quien debería ser considerado responsable jurídico y la falta de personalidad jurídica de la IA complica aún más la imputación.

Con el objetivo de presentar una respuesta a esta cuestión, se han analizado la posible imputación de responsabilidad a los actores que han intervenido en la creación y aplicación de los APAs, desde los programadores que los desarrollan, hasta el paciente cuya salud se ve afectado directamente por una decisión automatizada. Aun así, la atribución de responsabilidad de manera individual es una solución incompleta. Por ello, se ha evaluado la solución propuesta por la UE de un análisis de rendición de cuentas para establecer una responsabilidad compartida en función de la intervención de los actores. Esta solución, garantiza una mayor protección para los pacientes y una mayor claridad en cuanto a la imputación de la responsabilidad. Además, independientemente de a quien se le impute la responsabilidad, la UE ha establecido la imposibilidad de que debido a la utilización de un sistema automatizado las personas pierdan su condición de sujeto y pase a ser consideradas cómo objeto.

Una vez han sido analizadas las consecuencias jurídicas del uso de algoritmos de aprendizaje automatizado en la biomedicina, es relevante destacar que su aplicación también plantea cuestiones éticas relevantes que deben ser consideradas. Por lo tanto, en el siguiente apartado se abordará el análisis ético de la aplicación de los algoritmos de aprendizaje automatizado en el ámbito de la salud.

4. Parte ética

En este apartado, examinaremos las implicaciones éticas del uso de algoritmos de aprendizaje automático en la biomedicina. En primer lugar, explicaremos de manera general el funcionamiento de un algoritmo. A partir de esta base, abordaremos los principales problemas éticos que se presentan en su aplicación, como la vulneración de la privacidad de los usuarios en la gestión de su información personal, la falta de transparencia en su funcionamiento, la presencia de sesgos en los resultados y la inequidad en el acceso a los avances en inteligencia artificial.

Para una mejor organización del capítulo, abordaremos estas problemáticas éticas en función de los tres valores de la modernidad: la libertad, la igualdad y la fraternidad (Ortega, 2006). De esta manera, analizaremos cómo el uso de los algoritmos de aprendizaje automático en la biomedicina puede afectar a estos valores fundamentales y cómo se pueden abordar estas cuestiones éticas para garantizar una aplicación justa y responsable de la inteligencia artificial en el ámbito de la salud.

Actualmente la mayoría de los modelos de inteligencia artificial, y por ende los algoritmos de aprendizaje automático, se encuentran dentro de la categoría de IA específica. Los programadores configuran los algoritmos conforme a una serie de principios, normas o requisitos, fuera de las cuales la respuesta que proporcionan suele decepcionar. Esto se debe a que la inteligencia artificial “carece de capacidad para comprender el significado profundo de la información que maneja, el contexto social, político y cultural en que se produce” (López de Mántaras, 2020, p. 31). Por ello los creadores del algoritmo deben realizar una

programación ética, en la que puede haber excepciones a una regla general para evitar fallos y garantizar la equidad y justicia en su uso (Villas & Camacho, 2022).

Aun así, realizar esta tarea es especialmente complicada, y más aún cuando los programadores son los empleados de una empresa y deben guiarse por lo que les ha sido encomendado por sus superiores. En este sentido, el haber prestablecido la utilización de una determinada base de datos o de tener un tiempo limitado para la programación de la IA, puede afectar considerablemente el resultado (Karcher, 2018). En concreto, en el ámbito de la medicina la máquina no tiene las facultades subjetivas propias de las personas, ni tampoco tiene interiorizada la relación médico-paciente que siempre se ha caracterizado por la vocación de servir y ayudar a la persona enferma por parte del médico (Gracia, 1988). Sin embargo, los APAs no tienen esa misma vocación, ni la capacidad de interpretar la amplitud los principios éticos vinculados al ejercicio de la medicina (Karcher, 2018). Los algoritmos suelen caracterizarse por una filosofía puramente utilitarista, independientemente de si los resultados producidos resultan negativos para determinados sectores de la población. Esto no parece que vaya a cambiar debido a que la mayor parte de la investigación en IA se centra en construir sistemas eficientes para fines concretos y muy lucrativos, sin considerar las implicaciones éticas, como la falta de transparencia en su funcionamiento y la presencia de sesgos en los resultados (López de Mántaras, 2020).

En el caso de los APAs para la detección de enfermedades, ante el escenario en el que el médico tenga las mismas posibilidades de acertar que algoritmo, es preferible fiarse del diagnóstico dado por el facultativo médico. La mayoría de los algoritmos destinados al diagnóstico de enfermedades ofrece un resultado sobre-confiado, otorgando un porcentaje de éxito en su diagnóstico mayor al verdadero. Por su parte al médico le pasa lo contrario, y su confianza en su diagnóstico disminuye si el algoritmo produce una respuesta diferente a la suya. Por último, gracias a su lugar de trabajo, el médico tiende a reevaluar su diagnóstico con sus compañeros, lo cual hace que su diagnóstico sea más preciso (Mercier & Sperber, 2011).

En conclusión, debido a que los APAs son una herramienta en desarrollo, siguen careciendo de capacidad para comprender el significado profundo de la información que manejan y el contexto en que se produce. Por ello, es necesario programación una ética que garantice la equidad y justicia en su uso (Villas & Camacho, 2022). Sin embargo, actualmente la mayoría de los APAs han sido programados con fines utilitaristas, lo que puede implicar consecuencias éticas, que no sé producirían en la medicina tradicional donde, a diferencia de los APAs, existe una vocación de servir y ayudar por parte del médico. Además, ante el escenario en que el médico y el APA tengan la misma tasa de acierto en el diagnóstico de una enfermedad, es recomendable fiarse del médico ya que su resultado no estará afectado por la sobre confianza característica de la IA (Mercier & Sperber, 2011).

En este sentido, existe una serie de razones por las que no se debe confiar en exceso en los APAs, especialmente en el ámbito médico, debido a las implicaciones críticas que esto conlleva (Karcher, 2018). No obstante, ante el cada vez mayor uso de la IA, resulta esencial llevar a cabo un análisis sobre las consecuencias éticas de su uso. Para ello se utilizarán los tres valores de la modernidad y se evaluará cómo su aplicación afecta al derecho de libertad e igualdad de las personas y si contribuye a una mayor fraternidad en la sociedad.

4.1. Libertad

En el presente apartado se analizará el impacto que el creciente uso de los algoritmos de aprendizaje automatizado está teniendo en la libertad de las personas. Los APAs suelen requerir grandes cantidades de datos personales para su entrenamiento, y la recopilación de dichos datos puede afectar significativamente la privacidad de las personas. Además, la capacidad de los APAs para influir en las decisiones de las personas es cada vez mayor, Esto se debe a que la recopilación de información personal permite a los APAs dar resultados

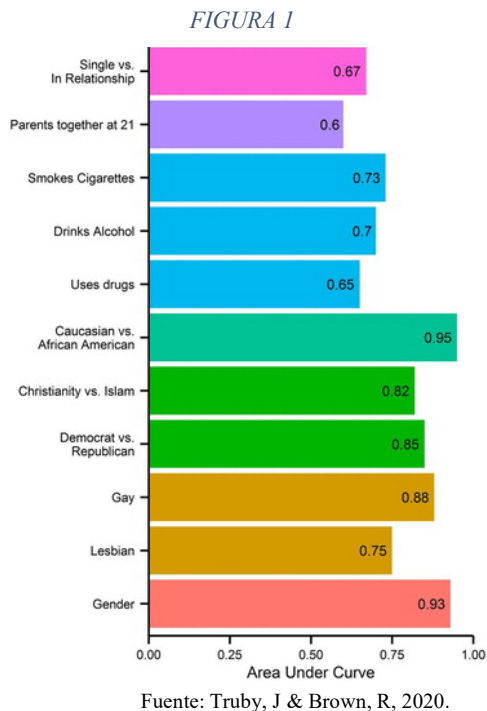
altamente personalizados, teniendo un mayor impacto en las decisiones que las personas toman y afectando contra su libertad individual (Villas & Camacho, 2022).

Esta cuestión ha sido abordada por numerosos sistemas legislativos. Ejemplo de ello es la consideración de la privacidad como derecho fundamental y su reconocimiento en el artículo 18 de la Constitución Española, que enfatiza la influencia de las tecnologías en la privacidad y establece que “la ley limitará el uso de la informática para garantizar el honor y la intimidad personal y familiar de los ciudadanos y el pleno ejercicio de sus derechos” (BOE, art 18.4, 1978). También posee relevancia la legislación de la UE, donde destaca el Reglamento General de Protección de Datos (RGPD). En él se pretende explicar y reforzar los mecanismos utilizados para la protección de datos personales” y cómo debe realizarse cualquier tratamiento posterior de los datos recolectados (Villas & Camacho, 2022). Su principal objetivo del reglamento es salvaguardar el artículo 8, sobre la Protección de datos de carácter personal, de la Carta de los Derechos Fundamentales de la UE (Europa, Parlamento y Europa, Consejo, 2016).

Aun así, numerosos expertos critican cómo la legislación en este ámbito precisa de una continua actualización debido a la velocidad la recopilación de datos y las nuevas y numerosas maneras de invadir nuestra privacidad (López de Mántaras, 2020). Ejemplo de ello, es la utilización de algoritmos de aprendizaje supervisado para la creación de clones digitales que consisten en una representación digital de un individuo mediante la combinación de datos de múltiples fuentes, incluyendo redes sociales, búsquedas en línea, transacciones financieras, historial médico y registros de ubicación (Wang et al, 2020). A pesar de que parezca una cuestión de ciencia ficción numerosos estudios ya han demostrado la existencia de este tipo de clones (Truby & Brown, 2020).

Los clones digitales son especialmente valiosos para las compañías, ya que permiten realizar una discriminación de precios y una publicidad engañosa mucho más efectiva. Por ello es

una gran manera de producir dinero, y se ha acuñado el término de la “economía de la privacidad” (Truby & Brown, 2020).



Gracias a los datos recopilados los **APAs** otorgan un cierto nivel de importancia mayor a cualidades específicas de las personas, como puede ser la raza, la orientación sexual, o su ideología política, para posteriormente realizar análisis predictivos sobre el comportamiento de las personas. De esta manera determinan la capacidad exacta que un usuario está dispuesto a pagar por un determinado producto. También sirve para determinar el tipo de publicidad (engañosa) sería necesaria para convencer a una persona de que necesita realizar la adquisición de un producto, que en realidad es inútil (Truby & Brown, 2020).

La figura 1 muestra un ejemplo de la relevancia que los algoritmos pueden otorgar a las cualidades de las personas para realizar un modelo predictivo de su comportamiento.

Estas estrategias son especialmente eficientes debido a que en la sociedad actual se ha confundido el concepto de felicidad por el de bienestar material, el cual es sinónimo de las posibilidades de consumo de las personas. Para la filósofa Adela Cortina (2014) esta confusión es una consecuencia del predominio de una cultura del consumo y el individualismo en la sociedad actual. Según su perspectiva, muchas personas tienden a buscar la felicidad a través de la acumulación de bienes y el consumo, lo cual genera una insatisfacción permanente y un alejamiento de la felicidad auténtica. Por su parte, la felicidad auténtica no se alcanza a través del consumo de bienes materiales, sino que requiere un cambio de perspectiva y de valores que nos permita desarrollar una conciencia más humana, solidaria y comprometida con el bienestar de todos los seres humanos (Cortina, 2014).

Una de las principales posibilidades planteadas para evitar que los a APAs y los clones digitales atenten contra nuestra libertad es poner el foco en el valor del empleo de los datos personales (Núñez et al., 2020). El objetivo es garantizar la fiabilidad de los intermediarios de datos, ya que en la actualidad los “precios de mercado y las acciones que los usuarios realizan para proteger la privacidad” no reflejan su valor. En este sentido, es necesario determinar un valor verdadero sobre la privacidad de cada individuo, que permita a la sociedad avanzar en la armonía del binomio regulación/innovación y garantice la privacidad de las personas (Núñez et al., 2020). De todas formas, es necesario que las personas estén atentos a sus acciones, especialmente las digitales. Existe una voluntad muy alta de intercambiar los datos personales, por un uso gratis y cómodo de la tecnología, lo que ha permitido a los *data miners* o mineros de datos comercializar la información de los usuarios y asegurarse de que ello no va a tener consecuencias legales. En palabras de la filósofa Zuboff hemos estado y estamos entrando en el “capitalismo de vigilancia”, por el cual, a través de la extracción y el procesamiento de información personal de los usuarios, las empresas u otras instituciones son capaces de cambiar la forma en que las personas interactúan con la tecnología y el mundo en general. (Truby & Brown, 2020).

En este sentido, es interesante la crítica realizada por Josep Monterde sobre la cultura de la protección de datos y la legislación que ésta posee por detrás. A través del concepto de “privacidad por diseño”, se señala como la supuesto protección y el anonimato de los datos no tendría que añadirse al final del producto o servicio para cumplir con una normativa. Por el contrario, debería estar integrado en todas sus etapas, y especialmente más tempranas, para garantizar que no se atenta contra la privacidad de los usuarios (Núñez et al, 2020).

Otra posibilidad para proteger nuestra privacidad y por ende nuestra libertad es la incorporación de dos figuras, el controlador de datos y el procesador de datos, realizada por la UE en su nueva propuesta del RGPD. El controlador de datos consiste en aquella persona

física o jurídica, autoridad, agencia que determina el procesamiento y el propósito de los datos personales. Por su parte el procesador de datos es aquel que se encarga de seguir las instrucciones del controlador y realizar un procesamiento de los datos responsable, que en numerosas ocasiones requieren la anonimización³ irreversible de ellos (Villas & Camacho, 2022).

Sin embargo, la anonimización irreversible de los datos, es complicada de realizar, y a veces engañosa. (Villas & Camacho, 2022). Las formas más habituales para anonimizar los datos la supresión, el enmascaramiento y la clasificación. La supresión consiste en eliminar completamente la información sensible de los datos. Por su parte, el enmascaramiento cambia determinados caracteres de la información personal por símbolos y la clasificación consiste en sustituir la información de una categoría por su etiqueta genérica.

A continuación, se presentará una tabla para mejor comprensión de las formas de anonimización.

Dato original	Supresión	Enmascaramiento	Clasificación
LARA RAMOS	' ' ' '	Lxxx Rxxxxx	[NOMBRE]
618385878	' ' ' '	6xx8xxxx8	[TELÉFONO]
lararamos@gmail.com	' ' ' '	xxxxxxxxx@gmail.com	[EMAIL]
29016	' ' ' '	2xxxxx	[COD_POSTAL]

Fuente: Adaptado de Villas & Camacho, 2022

³ También podría decirse anomatización

Esta anonimización de datos es fácil de realizar, ya existen multitud de bibliotecas accesibles al público para ello, lo difícil es la anonimización irreversible. El motivo es el hecho de que datos aparentemente genéricos, se pueden utilizar para identificar a una persona. En el año 2000 un estudio demostró como a partir de “la fecha de nacimiento, género y código postal se podía identificar a la mayoría de las personas de Estados Unidos” (Villas & Camacho, 2022). Otro ejemplo, fue realizado por Swenley (2002) el cual logró exitosamente la combinación de una base de datos médica con una lista de votantes para extraer el historial médico del gobernador de Massachusetts. Además, se ha demostrado que, a partir de los cuatro puntos espaciotemporales de cada persona, se puede identificar de manera al 95% de los individuos (de Montioye et al., 2013).

Además, cuando el conseguir realizar la anonimización irreversible de los datos puede afectar directamente al dilema privacidad-transparencia (Burrell, 2016). El dilema privacidad-transparencia se podría definir como aquellas “situaciones en las que, por preservar la privacidad, no se tiene acceso a suficiente información para tomar una decisión con precisión” (Villas & Camacho, 2022). En relación con las aplicaciones médicas de los algoritmos de aprendizaje automatizado, actualmente el acceso a los datos relativos a la salud está restringido, lo que ha generado en datos sesgados y poco representativos, que ha producido que el modelo produzca resultados discriminatorios (Starke et al, 2021).

Por último, es necesario mencionar el derecho al olvido, el cual es una vertiente del derecho de protección de datos digital que permite a los usuarios interesados el cifrado y borrado online de sus datos personales en aquellos casos que sean perjudiciales para sus derechos fundamentales, especialmente los derechos a la intimidad y al honor (Sancho, 2019, p.436). Se trata de un derecho autónomo, y a pesar de no estar en la Constitución ha sido reconocido por el Tribunal Constitucional en la STC 58/2018. Aun así, no se trata de un derecho absoluto y está limitado directamente el derecho a la información del artículo 20 de la Constitución Española (Claramount, 2020).

Por todo ello, podemos concluir que el uso de los algoritmos de aprendizaje automatizado tiene la capacidad de atentar contra la libertad de las personas al utilizar la información de los usuarios para influenciar en sus decisiones. Sin embargo, existe una legislativa creciente y diferentes estrategias para evitar un uso irresponsable de nuestros, y ayudar a la creación de inteligencia artificial que, en vez de coartar la libertad de las personas, la expanda.

En este sentido cabe destacar cómo el uso correcto de los algoritmos puede reducir significativamente la carga de ciertos trabajos, lo que permitiría a las personas no incurrir en actividades y poder tener más tiempo libre. Actualmente existen diferentes perspectivas sobre la posible influencia de la inteligencia artificial (IA) en el mercado laboral. Por un lado, algunos estudios sugieren que la IA podría reemplazar a gran parte de los trabajadores, generando una disrupción en el mercado laboral y afectando a la economía en general. Por otro lado, otros estudios consideran que la IA podría ser una herramienta complementaria que facilite el trabajo humano y permita una mayor eficiencia en las tareas realizadas (Benhamou, 2022). Entre los estudios que respaldan esta postura nos centraremos el informe realizado Benhamou (2022) sobre: *La transformación del trabajo y el empleo en la era de la inteligencia artificial Análisis, ejemplos e interrogantes realizado en 2022 por las Naciones Unidas*.

Es necesario clarificar cómo los estudios que defienden la disrupción del mercado laboral por parte de la IA no poseen bases científicas suficientes. En estos estudios no se hace una distinción entre los impactos que sustituyen el trabajo humano y los que lo complementan, lo que hace difícil obtener un consenso sobre las repercusiones de la IA en el empleo, los salarios o las categorías socio profesionales. Además, estos estudios se centran exclusivamente en el potencial técnico de supresión de puestos de trabajo, sin considerar que la IA también es capaz crear nuevos empleos y transformar el contenido de los trabajos. Por último, no hay una consideración de los costos y las dificultades que deben enfrentar las empresas para obtener datos que sirvan para entrenar los modelos de aprendizaje automatizado o la inversión necesaria para que los empleados adquieran competencias para

el correcto uso de estos modelos. Todo ello ralentiza la implementación y evita que la IA produzca una disrupción del mercado laboral (Benhamou, 2022).

Además, como se analizó en el apartado *Avances de los APAs aplicados al mundo de la biomedicina*, el uso de este tipo de algoritmos es capaz de reducir notablemente la carga laboral de los profesionales. En este sentido, se ha demostrado su potencial para la reducción de las tareas monótonas de los facultativos médicos (Aguirre et al., 2021) y para una gestión más efectiva del hospital (Aracena et al., 2022).

En conclusión, el uso de los algoritmos de aprendizaje automatizado tiene la capacidad de atender contra la libertad de las personas al utilizar la información de los usuarios para influenciar en sus decisiones. Sin embargo, para evitar un uso irresponsable de nuestros datos, existen diferentes estrategias y legislaciones en marcha que ayudan a garantizar que la IA expanda en lugar de coartar nuestra libertad. Entre las estrategias, destacan el uso responsable de los datos personales por parte de las personas y las empresas, la privacidad por diseño y la anonimización de datos. En la legislación, la UE ha desempeñado un papel relevante, gracias a la incorporación de nuevas figuras como el controlador de datos y el procesador de datos y defender la existencia del derecho al olvido de las personas. Además, es relevante considerar cómo el uso correcto de los algoritmos puede reducir significativamente la carga de ciertos trabajos, lo que permitiría a las personas tener más tiempo libre y un mayor bienestar.

Una vez que hemos evaluado cómo el uso de los algoritmos de aprendizaje automatizado puede afectar a la libertad, se procederá a analizar sus posibles consecuencias en otro valor fundamental de la modernidad: la igualdad.

4.2. Igualdad

En este apartado se analizará el potencial de los algoritmos *machine learning* aplicados al mundo de la biomedicina de producir resultados discriminatorios para las personas. En otras palabras, que se no traten igual a las personas, por razón de nacimiento, raza, sexo, religión, opinión o cualquier otra condición o circunstancia personal o social, sin causa justificada. (art 14, Constitución Española).

Para ello se analizará el problema de la opacidad de los algoritmos, que nos impide determinar si un algoritmo está actuando de manera discriminatoria o no, al no poder determinar en qué se han basado sus resultados. Igualmente, hay que abordar la cuestión de los efectos de estos algoritmos para la inclusión social y económica.

4.2.1. Algoritmos opacos

En este apartado profundizaremos en la opacidad de los algoritmos de machine learning, que ha sido mencionada a lo largo del trabajo. Para ello, se utilizará como referencia el trabajo de Grote & Berens (2020), específicamente su artículo "*On the ethics of algorithmic decision-making in healthcare*" (2020), en el que se distinguen tres clases de opacidad del algoritmo.

- La **opacidad como secreto corporativo** resultado del ocultamiento de información y la falta de transparencia en las prácticas empresariales. A día de hoy genera problemas éticos, especialmente relacionados con la privacidad. Aun así, posee solución cómo la creación de estándares regulatorios o el uso responsable de internet (Burrell, 2016).
- El segundo tipo se refiere a la **opacidad por falta de conocimiento técnico**. Esta opacidad es superada cuando el profesional que interpreta los resultados tiene un conocimiento suficiente de programación. Aun así, en los algoritmos de *machine learning* aplicados al ámbito biomédico, la responsabilidad de la aplicación final

recae en el médico, quien no está obligado a tener este tipo de conocimientos técnicos (Grote & Berens, 2020).

- El tercer tipo de opacidad es constante e insuperable. Se trata de la **opacidad derivada de la complejidad matemática de los algoritmos**. En este caso, la interpretación de la información no es posible para el conocimiento humano debido a la sofisticación de los cálculos implicados. Esta opacidad es especialmente relevante en las redes neuronales profundas, cuya complejidad hace que la interpretación de sus resultados sea extremadamente difícil o incluso imposible. Dado que esta opacidad es permanente suele generar mayores problemas éticos y de responsabilidad en el uso de algoritmos de inteligencia artificial (Grote & Berens, 2020).

El problema de la falta de transparencia en los algoritmos de aprendizaje automático está estrechamente relacionado con la rapidez con la que se desarrollan estos algoritmos y con la falta de supervisión en su desarrollo y aplicación. Esta falta de transparencia puede resultar perjudicial para grupos específicos y afectar a la inclusión social y económica (McCradden et al., 2020). En particular, se ha observado que estos algoritmos pueden afectar negativamente la inclusión social de mujeres y personas de color, así como de personas con discapacidad. En los próximos apartados, se evaluarán las consecuencias de la aplicación de los APAs en términos de la inclusión de mujeres y personas de color.

4.2.2 APAs e inclusión social

4.2.2.1 APAs y sesgo de género

En este apartado se aborda el impacto de los algoritmos de aprendizaje automatizado en la inclusión social, y se examina si su uso puede reducir el sesgo de género en el campo médico. Actualmente, ciertos campos de la biomedicina están afectados por el sesgo de género, lo cual puede implicar que los algoritmos de aprendizaje automatizado ofrezcan resultados discriminatorios hacia las mujeres. Esto se debe a que los algoritmos se entrenan con los

datos existentes, y al estar estos datos sesgados, los algoritmos darán resultados distorsionados (Leucona, 2020).

Ejemplo de ello es cómo el uso de APAs permite realizar una detección más eficaz de la posibilidad de sufrir un ataque cardíaco⁴ en hombres en comparación con las mujeres. Los síntomas que indican la posibilidad de sufrir un ataque cardíaco son diferentes según el género, pero la mayoría de los datos utilizados para entrenar los algoritmos corresponden a hombres. En este sentido, las mutaciones en el gen STK11 se asociaron con un riesgo más bajo de enfermedad de arteria coronaria en las mujeres, mientras que los cambios en el gen COL3A1 se asociaron con un riesgo más alto en los hombres. Además, es necesario tener en cuenta las diferencias en los niveles de hormonas y en la respuesta inflamatoria entre los géneros. Ello, evitaría que las tasas de supervivencia después de un ataque cardíaco sean significativamente más bajas en las mujeres que en los hombres (Dungan et al., 2022).

Otro ejemplo parecido es el expuesto por Chaki y sus colaboradores en 2020. Igual que en el caso anterior, los algoritmos de aprendizaje automático no han considerado que los patrones de la diabetes y sus síntomas son diferentes entre hombres y mujeres. La diabetes es una enfermedad crónica que se caracteriza por que el paciente tiene exceso de glucosa en sangre, lo cual le puede generar una serie de complicaciones que varían desde el cansancio a problemas oculares. Existen dos tipos principales de diabetes, diabetes tipo 1 y diabetes tipo 2. (Mayo Clinic, 2023) Las mujeres tienen una mayor probabilidad de presentar la diabetes tipo 2 "latente autoinmune", la cual es similar a la diabetes tipo 1, pero se desarrolla de manera más lenta y tiene síntomas diferentes. Esta consideración no ha sido teniendo en cuenta por numerosos de los algoritmos de *machine learning* utilizados para la detección de la diabetes, ya que suelen ser entrenados con datos que corresponden en su mayoría a los hombres, y por consecuencia las mujeres reciben una detección menos eficaz (Chaki et al., 2020).

⁴ El ataque cardíaco se produce debido a un estrechamiento de las arterias que suministran sangre al corazón, puede ser fatal o provocar una grave alteración del ritmo cardíaco

Por ello una solución para mejorar la precisión del diagnóstico en la detección de enfermedades cardiovasculares y diabetes tipo 2 en mujeres es entrenar los algoritmos con datos más representativos de la población. De esta manera, se puede lograr un diagnóstico igualmente eficaz para ambos géneros y reducir el sesgo de género en la detección de estas enfermedades y en la práctica médica (Leucona, 2020).

4.2.2.2. APAs y racismo en la biomedicina

El uso de los algoritmos de *machine learning* puede ayudar a disminuir el racismo en el mundo biomédico, ya que proporciona una mayor precisión y objetividad en la interpretación de los datos (McCradden et al., 2020). Sin embargo, la aplicación de estos algoritmos puede causar el efecto contrario, ya que los datos con los que se entrena el modelo pueden reforzar los estereotipos y las preconcepciones raciales existentes o habidas. Por ello, si los datos utilizados para entrenar estos modelos no están libres de sesgos y no reflejan adecuadamente la diversidad de la población, los resultados resultarán racistas (Leucona, 2020).

En este sentido, es relevante destacar el actual sesgo diagnóstico de esquizofrenia en Estados Unidos sobre la población negra. Esto se debe a que, ante unos mismos síntomas, la población negra es diagnosticada con diferentes trastornos, en teoría menos peligrosos para la sociedad, como la depresión o el trastorno bipolar. Por ello, si utilizamos un modelo de aprendizaje automático para la detección de la esquizofrenia y lo entrenamos con los datos de Estados Unidos, el modelo diagnosticaría la esquizofrenia en la población negra, lo que es un resultado racista (Starke et al, 2021).

Es relevante considerar, que el racismo de los modelos de *machine learning* no sólo se produce en el diagnóstico de enfermedades, sino también en los algoritmos utilizados en la medicina de precisión. El artículo realizado por Geneviève y sus colaboradores (2020) señala cómo la mayoría de los datos genéticos utilizados con fines de investigación en la medicina

personalizada proceden de participantes de ascendencia predominantemente europea. Esto se debe en parte a que la medicina de precisión requiere una mayor participación del paciente, la cual es menor en la población de color, debido a que poseen menor confianza y ganas de contribuir en los servicios sanitarios médicos. (Geneviève et al, 2020). Esta desconfianza se explica los numerosos casos de racismo que ha habido en la biomedicina, como el experimento Tuskegee o el caso de las células Henrietta Lacks. Además, en los casos en los que sí que existe confianza, es posible estas poblaciones no tengan el mismo acceso a la medicina y por lo tanto sus datos no se encuentran recopilados (Haenlein & Kaplan, 2019).

Por último, el racismo también puede estar presente en los algoritmos de aprendizaje automático que realizan la asignación de trasplante, más específicamente el trasplante de riñón. En este sentido, los algoritmos utilizados para medir la tasa de filtración glomerular (eGFR) pueden dar resultados más altos en pacientes negros, lo que lleva a un retraso en la asignación del trasplante y por lo tanto peores resultados para estos pacientes (Vyas et al, 2020).

4.2.3. Posibles soluciones para algoritmos menos discriminatorios

Una posible solución para que los algoritmos de *machine learning* no incurran en conductas discriminatorias sería aumentar su transparencia. Es cierto que siempre va a ver un nivel de opacidad correspondiente a la complejidad matemática de los algoritmos, pero podemos centrarnos diferentes estrategias que den mayor seguridad sobre que los resultados de los algoritmos no sean discriminatorios (Burrell, 2016).

Una de ellas es asegurarse de que el modelo no esté contaminado e influenciado por el efecto *Clever Hans*. Cuando esto pasa, las predicciones dadas por el modelo son aparentemente precisas y útiles, pero son falsas al no haber considerado el contexto adecuado y haber realizado correlaciones irrelevantes en los datos de entrenamiento (Lapuschkin et al, 2019). Este efecto es frecuente en los algoritmos de aprendizaje supervisado aplicados a la

biomedicina debido a la “complejidad de las correlaciones entre múltiples los factores biológicos y ambientales que pueden influir en un diagnóstico” (Starke et al, 2021).

Otra estrategia es asegurarse que los datos de entrenamiento se basan posean la suficiente diversidad para no incurrir en los prejuicios, existentes o habidos, de la práctica médica. En este sentido es necesario que los datos utilizados para entrenar los algoritmos de aprendizaje automatizado deben ser representativos de la población mundial, y no únicamente de la población con ascendencia europea. (Genieve et al, 2020).

Por su parte, la Declaración de Ginebra de la Asociación Médica Mundial sugirió evitar incluir la etnia, raza y género como variables de entrada en los algoritmos médicos para prevenir resultados discriminatorios. Sin embargo, esta medida puede resultar en una disminución de la precisión y calidad de los resultados obtenidos (Starke et al, 2021). Así, la exclusión de una categoría en los datos de entrenamiento puede resultar inefectiva ya que esa categoría se verá reflejada/inferida en otros datos aparentemente inocentes, como los códigos postales o el estatus socioeconómico (Lapuschkin et al, 2019). Además, ciertas situaciones precisan discriminación positiva, lo que justifica la inclusión de la etnicidad en los datos de entrenamiento. La discriminación positiva consiste “en la puesta en marcha de medidas que, aunque formalmente discriminatorias, están destinadas a eliminar o a reducir desigualdades fácticas” (Velasco Arroyo, 2007). En el ámbito de la biomedicina un ejemplo sería una mayor consideración de las poblaciones de color en los países occidentales sufran la enfermedad lupus eritematoso sistémico (LES). El no realizar una discriminación positiva y excluir de la categoría etnicidad en la entrada de los datos de entrenamiento, dificultaría la obtención un diagnóstico preciso para este sector de la población vulnerable, e implicaría que el algoritmo realice un resultado discriminatorio (Starke et al., 2021).

Por todo ello, no existe una solución genérica, por lo que será necesario encontrar un equilibrio entre la transparencia, la diversidad de los datos y la inclusión adecuada de ciertas categorías para prevenir la discriminación en los algoritmos de *machine learning* en la medicina. Además, siempre es necesario recordar que el aprendizaje automático refuerza los

estereotipos y las preconcepciones raciales que haya habido en sus datos de entrenamiento y es necesario un uso ético y responsable de ellos (Genieve et al, 2020).

Después de haber examinado el impacto de los algoritmos de aprendizaje automático en la igualdad, es relevante abordar sus posibles implicaciones en último valor fundamental de la modernidad: la fraternidad (Ortega, 2006).

4.3. Fraternidad

El presente apartado pretende abordar si la IA tiene la capacidad de crear un sistema más justo para el acceso de la medicina. Para ello, se analizará cómo el uso de APAs afecta a la inclusión económica en la atención médica, especialmente en relación con la asignación de recursos limitados y el acceso a servicios de salud para comunidades desfavorecidas.

La fraternidad se refiere a la idea de solidaridad y cooperación entre las personas, especialmente en la búsqueda del bien común. La fraternidad no solo implica actos filantrópicos, sino que está relacionada con la búsqueda de una justicia universal y la igualdad de oportunidades para todas las personas, lo que puede contribuir a la construcción de una sociedad más justa y equitativa (Ortega, 2006). En el contexto de la justicia, es relevante destacar que se trata de una noción de equidad y no de igualdad, tal como lo explica John Rawls. Para Rawls, la igualdad en sí misma no es suficiente para garantizar la justicia, ya que puede ser injusto tratar a todos de la misma manera si algunos tienen necesidades o circunstancias diferentes. En cambio, la justicia requiere que se tomen en cuenta las desigualdades y que se adopten medidas para corregirlas de manera equitativa (Swartz, 2012).

Por tanto, la capacidad de los APAs para proporcionar una atención médica más justa se relaciona con la idea de que todas las personas reciban un trato igual ante la ley, siempre y cuando no haya causas de discriminación positiva, y de que se logre una igualdad real y

efectiva en la práctica. En este sentido, se abordará si el uso de los APAs puede contribuir a la solidaridad mundial reduciendo la brecha social en el acceso de medicina o por el contrario puede perpetuar y ampliar las desigualdades existentes (Lain Moyano, 2021).

4.3.1. . APAs y aumento brecha social en el acceso a la medicina

Los APAs pueden aumentar la brecha social en el acceso a la medicina ya que la mayoría de ellos siguen una filosofía utilitarista (López-Mantarás, 2020). Esto significa que se centran en maximizar el resultado más favorable para la mayoría, en lugar de considerar la situación individual de cada paciente, lo que puede llevar a desigualdades en la atención médica. Estas desigualdades se pueden deber a no haber considerado factores socioeconómicos y demográficos o haberlos considerado de manera incorrecta (Genieve et al, 2020).

Un ejemplo de desigualdades en la atención médica producida por estos algoritmos es el evaluado en el artículo Dissecting racial bias in an algorithm used to manage the health of populations. En él se realice un análisis sobre el uso de algoritmos comerciales para tomar decisiones en el sistema de salud de Estados Unidos y se encontró que su utilización produce resultados injustos y racistas contra la población negra. Sin embargo, "racismo" no es totalmente preciso en este contexto, ya que la discriminación del algoritmo se debe en gran medida a factores socioeconómicos que, en este caso, también se ven relacionados con la raza. El algoritmo predice los costos de atención médica en lugar de la enfermedad. Debido a que ante un mismo nivel de necesidad el sistema de salud gasta menos en pacientes negros que en los blancos, el algoritmo concluye falsamente que los pacientes negros están más saludables y, por lo tanto, no necesitan la misma atención médica que los pacientes blancos. Remediar esta disparidad aumentaría el porcentaje de pacientes negros que reciben atención médica adicional del 17,7% al 46,5% (Obermeyer et al., 2019).

Además, uno de los mayores desafíos actuales es reducir los costos de implementación de este tipo de tecnologías. En este sentido Davenport & Kalakota (2019), demostraron que el costo de implementación del aprendizaje automático en hospitales puede variar significativamente y puede estar influenciado por factores como el tamaño y complejidad del hospital, el tipo de aplicaciones que se quieran desarrollar, la inversión en tecnología y hardware, y los costos de mantenimiento y actualización. Es por ello que aquellos hospitales con menores recursos físicos y financieros se beneficiaran menos de los avances de la IA producidos en el ámbito biomédico (Davenport & Kalakota, 2019).

Por último, relacionado con lo establecido en el apartado del racismo, la falta de representación equitativa en los datos utilizados para entrenar los algoritmos detrás de la medicina personalizada ha llevado a que esta tenga mejores resultados en países con mayor PIB (Genieve et al., 2020). Por lo tanto, la aplicación de la medicina personalizada puede tener sesgos incorporados y ofrecer una atención médica desigual y peor a los grupos bajos ingresos. Este problema también ocurre en los algoritmos aplicados para la detección de enfermedades (Starke et al., 2021).

En conclusión, los algoritmos de aprendizaje automático aplicados a la medicina pueden aumentar la brecha social en el acceso a la medicina debido a su enfoque utilitarista. Además, la falta de representación equitativa en los datos utilizados para entrenar los algoritmos y el alto costo de implementación también son barreras para lograr una atención médica más justa y equitativa. Sin embargo, el correcto uso de los APAs sí que puede favorecer la solidaridad mundial y permitir un acceso más equitativo y universal de la medicina, tal y cómo será evaluado a continuación.

4.3.2. APAs y disminución brecha social en el acceso a la medicina

El uso de los algoritmos de aprendizaje automatizado en la medicina puede disminuir la brecha social. Esto se debe a que su correcto uso ha permitido la reducción de carga profesional de los médicos, lo que les permite tener más tiempo para atender a más pacientes.

Además, como se mencionó en el capítulo uno, los algoritmos de aprendizaje automatizado tienen el potencial de realizar una gestión más efectiva en tiempo y en dinero del hospital, lo que puede mejorar el acceso a la atención médica y reducir la brecha social en la salud. Por otro lado, la telemedicina tiene el potencial de proporcionar atención médica a distancia a personas que viven en áreas remotas y tienen dificultad para conseguir una opinión médica (Aracena et al., 2022). Por último, el uso de la IA puede tener un gran potencial para reducir los costos en el desarrollo de medicamentos, lo que podría tener un impacto positivo en la atención médica y en la reducción de los costos a largo plazo (Jumper et al., 2021).

Respecto de las posibles soluciones para que los algoritmos contribuyan a la solidaridad universal y a la disminución de la brecha social en el acceso de la medicina, resultaría conveniente la aplicación de estrategias que solucionen el problema de la falta de representación en los datos que entrenan el modelo. Además, sería beneficioso que los Estados y las Organizaciones internacionales intervengan para reducir los costos de implementación requeridos para el uso de los algoritmos de aprendizaje automatizado en la medicina, a través de programas de financiación e inversión en investigación y desarrollo, tal y como se propone en el Libro Blanco sobre la inteligencia artificial - un enfoque europeo orientado a la excelencia y la confianza, elaborado en 2020 por la Comisión Europea (Andrés Segovia, 2021).

En definitiva, el uso adecuado de los APAs puede ser una herramienta para lograr una atención médica más justa y equitativa, y para promover la fraternidad en la sociedad. Aun así, será necesario que se tomen medidas convenientes para que su uso permita reducir la

brecha social en el acceso a la medicina y garantizar que todos los individuos tengan acceso a la atención médica necesaria.

5. Conclusiones

Al principio de este trabajo de investigación se definieron los tres objetivos principales que pretendían ser abordados. Estos objetivos han servido para estructurar el trabajo y en presente apartado se tratará de determinar el cumplimiento de cada uno de ellos.

El primero de los objetivos consiste en realizar una investigación sobre los principales avances que ha supuesto el uso de los APAs de la medicina. Gracias a este objetivo se ha evaluado la utilidad de los APAs para ofrecer una medicina más personalizada que permita disminuir los efectos adversos en tratamientos de determinadas enfermedades. Además, este tipo de algoritmos ha ayudado al facultativo médico en diagnóstico de enfermedades, ya que permite mejorar de la calidad de imágenes radiológicas, y realizar una segmentación y detección de lesiones invisibles al ojo humano. Por último, su uso puede llevar gestión hospitalaria más eficaz, al ayudar a predecir de manera anticipada la inasistencia hospitalaria y la muerte intrahospitalaria. Además, gracias al Procesamiento Natural del lenguaje ha permitido realizar una extracción más rápida de la información relevante de los pacientes contenida en sus registros electrónicos médicos y desarrollar una comunidad científica mejor comunicada. A pesar de los mencionados avances los APAs son una herramienta que aún se encuentra en desarrollo y su aplicación posee implicaciones jurídicas y éticas, las cuales han fundamentado el objetivo dos y tres de este trabajo.

El segundo objetivo de nuestro trabajo consiste en realizar un análisis de las problemáticas jurídicas y sus posibles soluciones que supone el uso de los APAs en el ámbito de la medicina. Para ello se ha analizado de una de las principales problemáticas actuales, la imputación de la responsabilidad jurídica por la utilización incorrecta los APAs. Debido al interés del trabajo se ha evaluado la imputación de la responsabilidad individual a los distintos sujetos

que han intervenido en la creación y aplicación de los APAs utilizado en el ámbito biomédico. En otras palabras, se ha evaluado la posible responsabilidad de los programadores, los médicos, el hospital, el estado y el paciente. Aun así, la atribución de responsabilidad de manera individual es una solución incompleta. Por ello, se ha evaluado la solución propuesta por la UE de un análisis de rendición de cuentas para establecer una responsabilidad compartida en función de la intervención de los actores. Esta solución, garantiza una mayor protección para los pacientes y una mayor claridad en cuanto a la imputación de la responsabilidad. Además, independientemente de a quien se le impute la responsabilidad, la UE ha establecido la imposibilidad de que debido a la utilización de un sistema automatizado las personas pierdan su condición de sujeto y pase a ser consideradas cómo objeto.

El tercer objetivo ha consistido en realizar un análisis de las problemáticas éticas y sus posibles soluciones que supone el uso de los algoritmos de aprendizaje automatizado en el ámbito de la medicina. Para ello se ha evaluado cómo afecta los APAs a los tres principios de la modernidad: libertad, igualdad y fraternidad. En este análisis se han planteado diferentes cuestiones cómo la falta de transparencia en el funcionamiento de los algoritmos, la presencia de sesgos en los resultados, la privacidad por diseño, el uso responsable de los datos personales, la desigualdad en el acceso a los avances en inteligencia artificial. Además, este análisis ha permitido evaluar la importancia de utilizar los APAs de manera diligente, ya que su creciente aplicación puede tener consecuencias tanto positivas como negativas para las personas y la sociedad, dependiendo del uso que se les dé. Por ello, se han estudiado las distintas propuestas para mitigar el impacto de las consecuencias éticas, donde ha vuelto a destacar la importancia de la legislación de la UE.

Todos estos objetivos permiten destacar el hecho de que la IA es una herramienta y, como tal, no deberíamos otorgarle una sobreconfianza excesiva, sino utilizarla de manera consciente y responsable. En este sentido, es necesario que cuente con una legislación adecuada que garantice su correcto uso y que tenga en cuenta las problemáticas jurídicas y éticas que su aplicación puede generar. Además, es fundamental seguir investigando y

evaluando los avances y problemáticas que surgen con la utilización de los APAs en la medicina, para poder mejorar su aplicación y evitar posibles consecuencias negativas. En definitiva, el uso adecuado de los APAs en la medicina puede suponer un gran avance para la sociedad, pero es necesario llevarlo a cabo de manera consciente, responsable y regulada.

Respecto de las problemáticas para realizar este trabajo se puede mencionar la dificultad de interpretar los estudios redactados por médicos e ingenieros debido a la formación en ciencias sociales de la autora del trabajo. Esto puede haber limitado el alcance de la investigación y la comprensión de ciertos aspectos técnicos relacionados con los avances en el uso de APAs en la medicina. Por otro lado, la amplia variedad de temas relacionados con la IA y su interconexión ha supuesto un reto en cuanto a la organización y estructuración del trabajo. Esto podría ayudar a obtener una comprensión más completa de cómo diferentes actores ven la cuestión de la responsabilidad en el contexto de los algoritmos de aprendizaje automatizado en la medicina y podría haber proporcionado información valiosa para futuras discusiones sobre la legislación y regulación de los APAs en el ámbito biomédico.

En cuanto a futuras líneas de investigación podría ser interesante examinar el papel de la educación en el uso responsable de los APAs en la medicina, tanto desde la perspectiva de los profesionales médicos como de los pacientes y usuarios finales. Además, se podrían evaluar las políticas y regulaciones en torno a los APAs en la medicina en diferentes países y su efectividad en la protección de los derechos y la seguridad de los pacientes. En el ámbito jurídico hubiera sido llamativo recabar información sobre la opinión de las personas en cuanto a la imputación de responsabilidad por el uso inadecuado de los algoritmos, y en particular, si esta opinión varía en función de su ejercicio profesional.

6. Bibliografía

- Ada Lovelace Institute,. (2021, Agosto). *Algorithmic Accountability for the Public Sector*. Recuperado de: <https://www.opengovpartnership.org/documents/algorithmic-accountability-public-sector/>
- Aguirre, F et al. (2021, Agosto). INTELIGENCIA ARTIFICIAL APLICADA A LA IMAGEN MÉDICA. *Revista de Imagenología*, 24 (2), 09 - 20. Recuperado de: <<https://sriuy.org.uy/ojs/index.php/Rdi/article/view/94>>.
- Álcami, J et al. (2009). *BIOLOGIA. 2 BACHILLERATO*. Ediciones SM
- Andrés Segovia, B. (2021). El reinicio tecnológico de la inteligencia artificial en el servicio público de salud. *Ius et Scientia*, 7 (1), 327-356. Recuperado de: <https://doi.org/10.12795/IETSCIENTIA.2021.i01.17>
- Aracena, C., Villena, F., Arias, F., & Dunstan, J. (2022). Aplicaciones de aprendizaje automático en salud. *Revista Médica Clínica Las Condes*, 33(6), 568-575. Recueperado de: <https://doi.org/10.1016/j.rmclc.2022.10.001>
- Benhamou, S (2022, julio). La transformación del trabajo y el empleo en la era de la inteligencia artificial Análisis, ejemplos e interrogantes. *CEPAL*. Recuperado de: https://repositorio.cepal.org/bitstream/handle/11362/47985/S2200188_es.pdf?sequence=1&isAllowed=y
- Berrar, D. (2018). Bayes' theorem and naive Bayes classifier. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 403, 412. Recuperado de: <https://doi.org/10.1016/B978-0-12-809633-8.20473-1>
- BOE (14 de noviembre de 2002). Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica. «BOE» núm. 274, de 15 de noviembre de 2002, páginas 40126 a 40132.
- BOE (25 de abril de 1986). Ley General de Sanidad. Ministerio de Sanidad y Consumo. Madrid, 1986.
- BOE, Art. 18.4 (29 de diciembre de 1978) Constitución española. «BOE» núm. 311, de 29 de diciembre de 1978. *Cortes Generales*

- BOE, (16 de Agosto de 1889). Código Civil. Real Decreto de 24 de julio de 1889 por el que se publica el Código Civil. Legislación Consolidada. Madrid, España.
- Burrell, J. (2016). How the machine ‘thinks’: Understanding opacity in machine learning algorithms. *Big data & society*, 3(1), Recuperado de: <https://doi.org/10.1177/20539517156225>
- Chaki, J., Thillai Ganesh, S., Cidham, S.K., Ananda Theertan, S, A. (2020, Julio). Machine learning and artificial intelligence-based Diabetes Mellitus detection and self-management: A systematic review. *ScienceDirect*, 34(6). Recuperado de <https://www.sciencedirect.com/science/article/pii/S1319157820304134>
- Claramunt, J. C. (2020). La gestión de la información en el paradigma algorítmico: inteligencia artificial y protección de datos. *MEI: Métodos de Información*, 11(21), 42-58. Recuperado de: <https://dx.doi.org/10.5557/IIMEI11-N21-059082>
- Cortina, A. (2021). *Justicia y felicidad en ¿Para qué sirve realmente la ética?*. Planeta.
- Davenport T, Kalakota R. (2019). The potential for artificial intelligence in healthcare. *Future Healthc J*;6(2):94-98. Recuperado de <https://doi.org/10.7861/futurehosp.6-2-94>
- de Montjoye, YA., Hidalgo, C., Verleysen, M. *et al.* Unique in the Crowd: The privacy bounds of human mobility. *Sci Rep*, 3, 1376 (2013). Recuperado de: <https://doi.org/10.1038/srep01376>
- Dungan, J. R., Qin, X., Gregory, S. G., Cooper-Dehoff, R., Duarte, J. D., Qin, H., ... & Kraus, W. E. (2022). Sex-dimorphic gene effects on survival outcomes in people with coronary artery disease. *American heart journal plus: cardiology research and practice*, 17, 100152. Recuperado de: [10.1016/j.ahjo.2022.100152](https://doi.org/10.1016/j.ahjo.2022.100152)
- Geneviève, L.D., Martani, A., Shaw, D. et al (2020). Structural racism in precision medicine: leaving no one behind. *BMC Med Ethics*, 21(17). Recuperado de: <https://doi.org/10.1186/s12910-020-0457-8>
- Gracia, D. (1988). Ética médica. *En F. Rozman, Medicina Interna, 13a Ed. Barcelona: Doyma.*
- Grote, T, Berens, P. (2020). On the ethics of algorithmic decision-making in healthcare. *Journal of Medical Ethics*, 46, 205-211. Recuperado de: <http://orcid.org/0000-0002-9832-6046>
- Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, 61(4), 5-14.

- Hajirasouliha, I., & Elemento, O. (2020). Precision medicine and artificial intelligence: overview and relevance to reproductive medicine. *Fertility and Sterility*, 114(5), 908-913. Recuperado de: <https://doi.org/10.1016/j.fertnstert.2020.09.156>
- Jumper, J. et al. Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583-589. Recuperado de: <https://doi.org/10.1038/s41586-021-03819-2>
- Karches, K. E. (2018). Against the iDoctor: why artificial intelligence should not replace physician judgment. *Theoretical Medicine and Bioethics*, 39(2), 91-110. Recuperado de: <https://doi.org/10.1007/s11017-018-9442-3>
- Laín Moyano, G. (2021). Responsabilidad en inteligencia artificial: Señoría, mi cliente robot se declara inocente. *AIS: Ars Iuris Salmanticensis*, 9(1), 197-232. <https://doi.org/10.14201/AIS202191197232>
- Lapuschkin, S., Wäldchen, S., Binder, A. et al. (2019). Unmasking Clever Hans predictors and assessing what machines really learn. *Nat Commun* 10, 1096 Recuperado de: <https://doi.org/10.1038/s41467-019-08987-4>
- Lázaro, J., & Gracia, D. (2006). La relación médico-enfermo a través de la historia. In *Anales del sistema sanitario de Navarra*, 29, 7-17. Gobierno de Navarra. Departamento de Salud. Recuperado de: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1137-66272006000600002&lng=es&nrm=iso
- Leucona, I (2020, octubre). Aspectos ético, legales y sociales del uso de la IA y el big data en salud, en un contexto de pandemia. *Revista Internacional de pensamiento político*, 15 (1), 139-166. Recuperado de: <https://doi.org/10.46661/revintpensampolit.5599>
- Liang, G., Fan, W., Luo, H., & Zhu, X. (2020). The emerging roles of artificial intelligence in cancer drug development and precision therapy. *Biomedicine & Pharmacotherapy*, 128, 110255. Recuperado de: <https://doi.org/10.1016/j.biopha.2020.110255>
- López Baroni, M. J. (2019). Las narrativas de la inteligencia artificial. *Revista de Bioética y Derecho*, 46, 5-28. Recuperado de: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1886-58872019000200002&lng=es&tlng=es

- López de Mántaras, R. (2020). El traje nuevo de la inteligencia artificial. *Investigación y Ciencia* 526, 50- 59. . Recuperado de: <http://hdl.handle.net/10261/235756>
- Mayo Clinic. (2023, Enero, 20). *Diabetes - Síntomas y causas*. Recuperado de: <https://www.mayoclinic.org/es-es/diseases-conditions/diabetes/symptoms-causes/syc-20371444>
- McCadden, M. D., Joshi, S., Mazwi, M., & Anderson, J. A. (2020). Ethical limitations of algorithmic fairness solutions in health care machine learning. *The Lancet Digital Health*, 2(5) Recuperado de: [https://doi.org/10.1016/S2589-7500\(20\)30065-0](https://doi.org/10.1016/S2589-7500(20)30065-0)
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and brain sciences*, 34(2), 57-74. Recuperado de: <https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/why-do-humans-reason-arguments-for-an-argumentative-theory/53E3F3180014E80E8BE9FB7A2DD44049>
- Mesa, M. J., Asencio, J. M., Ruiz, F. R., & González, M. P. (2017). Análisis del coste económico del absentismo de pacientes en consultas externas. *Revista de Calidad Asistencial*, 32(4). Recuperado de: <https://doi.org/10.1016/j.cali.2017.01.004>
- Molnár-Gábor, F. (2020). Artificial intelligence in healthcare: doctors, patients and liabilities. *Regulating Artificial Intelligence*, 337-360. Recuperado de: https://doi.org/10.1007/978-3-030-32361-5_15
- Moreno Berdón, P. (2021). *Mejora para la imagen de Rayos X mediante el uso de Deep Learning*. Universitat Oberta de Catalunya [Trabajo de Fin de Grado]. Recuperado de: <http://hdl.handle.net/10609/132606>
- Müller, V. C., & Bostrom, N. (2016). Future progress in artificial intelligence: A survey of expert opinion. *Fundamental issues of artificial intelligence*, 376, 555-572. Recuperado de: https://doi.org/10.1007/978-3-319-26485-1_33
- Núñez, D. B., Mascaró, J. C., Gutiérrez, L. Q., & Gol-Montserrat, J (2020). INTELIGENCIA ARTIFICIAL Y DECISIONES CLÍNICAS: Cómo está cambiando el comportamiento del médico.

- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453. Recuperado de: <https://www.science.org/doi/10.1126/science.aax2342>
- Ordiales, H., & Barrera, G. M. (2019). Transferencia de estilo entre audios mediante redes neuronales. *Revista Digital DIIT*, 4,(1). Recuperado de: <http://repositoriocyt.unlam.edu.ar/handle/123456789/425>
- Ortega, R. V. M. (2006). los ideales de la modernidad V. *Valores e historia en la Europa del siglo XXI*, 1, 297.
- Porcelli, A. M. (2020). La inteligencia artificial y la robótica: sus dilemas sociales, éticos y jurídicos. *Derecho global. Estudios sobre derecho y justicia*, 6(16), 49-105.
- Pucchio, A., Eisenhauer, E. A., & Moraes, F. Y. (2021). Medical students need artificial intelligence and machine learning training. *Nature Biotechnology*, 39(3), 388-389. Recuperado de: <https://doi.org/10.1038/s41587-021-00846-2>
- Ramón Fernández, F. (2021). Inteligencia artificial en la relación médico-paciente: Algunas cuestiones y propuestas de mejora. *Revista chilena de derecho y tecnología*, 10(1), 329-351. Recuperado de: <http://dx.doi.org/10.5354/0719-2584.2021.60931>
- Ruiz de Huidobro de Carlos, J.M. (2015). *MANUAL DE DERECHO CIVIL. PARTE GENERAL* (4ta ed.). Dykinson.
- Salehi, H., & Burgueño, R. (2018). Emerging artificial intelligence methods in structural engineering. *Engineering structures*, 171, 170-189. Recuperado de: <https://doi.org/10.1016/j.engstruct.2018.05.084>
- Schwartz, S. P. (2012). *A brief history of analytic philosophy: From Russell to Rawls*. John Wiley & Sons.
- Starke, G., De Clercq, E. & Elger, B.S. Towards a pragmatist dealing with algorithmic bias in medical machine learning. *Med Health Care and Philos* 24, 341–349 (2021). Recuperado de: <https://doi.org/10.1007/s11019-021-10008-5>
- Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05), 557-570. Recuperado de: <https://doi.org/10.1142/S0218488502001648>

- Truby, J., Brown, R. (2020). Human digital thought clones: the *Holy Grail* of artificial intelligence for big data. *Information & Communications Technology Law*, 30(2), 140-168. Recuperado de: [10.1080/13600834.2020.1850174](https://doi.org/10.1080/13600834.2020.1850174)
- Unión Europea, Parlamento y Consejo. (2016). 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 (LCEur 2016\605). *Diario Oficial de la Unión Europea*, 4-5-2016, L 119, 1-88. Recuperado de: <https://bit.ly/2FIPg7O>
- Unión Europea: Comisión Europea, Dirección General de Redes de Comunicación, Contenido y Tecnologías, (2019). *Directrices éticas para una IA fiable*, Oficina de Publicaciones. Recuperado de: <https://data.europa.eu/doi/10.2759/14078>
- Unión Europea: Comisión Europea (2018). *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*, 25 de abril 2018, COM (2018). Recuperado de: <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=COM%3A2018%3A237%3AFIN>
- Unión Europea: Parlamento Europeo y Consejo (2021). Propuesta de Reglamento de Inteligencia Artificial: el Consejo pide que se promueva una IA segura que respete los derechos fundamentales. Bruselas, 21.4.2021, COM(2021) 206 final. Recuperado de: <https://www.consilium.europa.eu/es/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/>
- Velasco Arroyo, J. C. (2007). DISCRIMINACIÓN POSITIVA, DIVERSIDAD CULTURAL Y JUSTICIA. *Daimon Revista Internacional de Filosofía*, (41), 141–156. Recuperado a partir de <https://revistas.um.es/daimon/article/view/20821>
- Villas, M., & Camacho, J. (2022). *Manual de ética aplicada en Inteligencia Artificial* (2da ed.). Anaya.
- Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in plain sight—reconsidering the use of race correction in clinical algorithms. *New England Journal of Medicine*, 383(9), 874-882. Recuperado de: <https://www.nejm.org/doi/full/10.1056/NEJMms2004740>

- Wang, Y., Liu, D., & Hu, Y. (2020). Personalized digital twins: A new paradigm for personalized services. *IEEE Consumer Electronics Magazine*, 9(1), 36-44. Recuperado de: [10.3390/jpm11080745](https://doi.org/10.3390/jpm11080745)
- Xu, J., Yang, P., Xue, S., Sharma, B., Sanchez-Martin, M., Wang, F., ... & Parikh, B. (2019). Translating cancer genomics into precision medicine with artificial intelligence: applications, challenges and future perspectives. *Human genetics*, 138(2), 109-124. Recuperado de: <https://doi.org/10.1007/s00439-019-01970-5>
- Zhang, L., Tan, J., Han, D., & Zhu, H. (2017). From machine learning to deep learning: progress in machine intelligence for rational drug discovery. *Drug discovery today*, 22(11), 1680-1685. Recuperado de: <https://doi.org/10.1016/j.drudis.2017.08.010>