



COMILLAS
UNIVERSIDAD PONTIFICIA

ICAI

ICADE

CIHS

Facultad de Ciencias Económicas y Empresariales (ICADE)

Doble Grado en ADE y Relaciones Internacionales

TRABAJO DE FIN DE GRADO

***MACHINE LEARNING EN SU
BÚSQUEDA DE LA
RESPONSABILIDAD ÉTICA***

Autora: Coleta Múgica Lecuona

Director: José Luis Fernández Fernández

Madrid, abril 2021

MACHINE LEARNING EN SU BÚSQUEDA DE LA RESPONSABILIDAD ÉTICA

Resumen

Machine Learning, como parte de la Inteligencia Artificial, es una realidad ya innegable en la sociedad. De repente, el mundo se ha visto en medio de una revolución digital sin precedentes con un desconocimiento sobre qué efectos puede tener, ni sus implicaciones en un futuro. A pesar de los esfuerzos de algunas instituciones y entidades por regular su presencia en la multitud de industrias y entornos en los que habita, aún sigue sin haber claridad cuando intentamos explicar sus funciones o su alcance. El debate ético que despertó hace unos años, sigue hoy en día desarrollándose y completándose a medida que la opacidad de los datos y de la tecnología se hace más evidente. Uno de los desafíos éticos centrales de la Inteligencia Artificial es si se le puede conceder responsabilidad ética o si ésta recae en los desarrolladores o los usuarios. De cualquier manera, el concepto de respuesta y responsabilidad trae consigo la cuestión de la transparencia, la interpretabilidad y la comprensión de esta materia.

Palabras clave: Aprendizaje Automático, Inteligencia Artificial, Algoritmos, Big Data, Ética, Marco Legal, Utilitarismo, Responsabilidad, Opacidad, Consciencia, Moralidad

Abstract

Machine Learning, as bran of Artificial intelligence, is already an undeniable reality of society. In the blink of an eye, the world finds itself in the middle of an unprecedented digital revolution, lacking knowledge of its effects or its implications in the future. Despite the efforts of some institutions to regulate it in the multitude of environments it appears in, there is still no clarity when we try to explain its scope or functions. The ethical debate that has been growing for some years, continues to unfold and stretch as the opacity of data and technology becomes more evident. One of the main ethical challenges of Artificial Intelligence is whether it can have responsibility or whether this lies with the developers or users. Either way, this latest concept brings the questions of transparency, interpretability and the comprehension of this matter.

Key words: Machine Learning, Artificial Intelligence, Algorithms, Big Data, Ethics, Legal Framework, Utilitarianism, Responsibility, Opacity, Consciousness, Morality

Índice

<u>1. Introducción</u>	4
1.1. Objetivos.....	4
1.2. Justificación del tema.....	5
1.3. Metodología y estructura.....	6
<u>2. Estado de la cuestión</u>	8
2.1. ¿Qué es la Inteligencia Artificial?.....	8
2.2. ¿Qué es <i>Machine Learning</i> dentro de la IA?.....	13
2.3. Marco legal hasta el momento.....	16
2.4. Debate ético más actual.....	18
<u>3. El Aprendizaje Automático y la Ética</u>	20
3.1. Principales teorías éticas de la Inteligencia Artificial.....	20
3.2. Dimensión ética en <i>Machine Learning</i>	23
3.2.1. El papel del profesional.....	24
3.2.2. El papel de los algoritmos.....	24
3.2.3. El papel de los datos.....	26
3.3. Vacío de un marco legal común para <i>Machine Learning</i>	29
3.4. <i>The responsibility gap</i>	30
<u>4. Casos que apoyan la capacidad de <i>Machine Learning</i> para actuar éticamente</u>	33
4.1. Caso 1. <i>Machine Learning</i> como testigo en juicios de violencia y asesinato...33	
4.2. Caso 2. <i>Machine Learning</i> en la Medicina.....	35
<u>5. Casos que socavan la capacidad de <i>Machine Learning</i> para actuar éticamente</u>	38
5.1. Caso 1. <i>Machine Learning</i> en el espacio laboral.....	38
5.2. Caso 2. La huella de <i>Machine Learning</i> en la salud mental.....	39
<u>6. Conclusiones</u>	41
<u>7. Bibliografía</u>	44

1. Introducción

“Los algoritmos, silenciosamente, estructuran nuestras vidas” (Martin, 2019). La tecnología lleva formando parte del ser humano desde el neolítico, cuando se diseñan las primeras técnicas agrícolas. Sin embargo, es en el siglo XX cuando se comienza a producir lo que conocemos como “Revolución Digital”, considerada a su vez la “Tercera Revolución Industrial” que trae consigo nuevos conceptos de la tecnología y un desarrollo radical. La constante innovación y evolución digital de los recientes años ha llevado al rápido surgimiento de la Inteligencia Artificial que, a su vez, ha supuesto la entrada a una “Cuarta Revolución Industrial” (Schwab, 2016).

A pesar de su corta vida, la Inteligencia Artificial -o IA- ya es una realidad que nos rodea y podríamos decir que prácticamente forma parte “inherente” del ser humano. Se ha observado que su poder es ilimitado e inabarcable, lo que nos lleva a pensar que su gestión tiene y tendrá un gran impacto tanto al hacer un uso éticamente correcto de las herramientas, como al hacer un uso éticamente incorrecto. Es precisamente por ello, por lo que la Inteligencia Artificial -y sus ramas, como *Machine Learning*- deben someterse a análisis, ya que un desarrollo tecnológico debe ir enmarcado dentro de un desarrollo teórico-ético. De no ser así, ¿no sería la “Cuarta Revolución Industrial” un paso precipitado en la historia de la innovación y la tecnología?

1.1. Objetivos del trabajo

En base a lo citado previamente, el objetivo principal de esta investigación es analizar más a fondo el funcionamiento de *Machine Learning* -o Aprendizaje Automatizado- como rama de la Inteligencia Artificial, así como conocer el alcance del impacto negativo que estos sistemas pueden tener con un vacío legal y ético, sin un marco teórico.

Además, a través de este trabajo se pretende llegar a un cuestionamiento de la implementación de la ética en cada paso de la cadena de aprendizaje de los algoritmos de estas máquinas. Junto a ello, otro objetivo es estudiar las distintas teorías éticas más relevantes y que más pueden explicar la revolución tecnológica que estamos viviendo.

Los programas automatizados con el conocimiento aplicado tienen la capacidad de aprender de forma autónoma junto al uso de un gran volumen de datos sin intervención humana. Sin embargo, ¿Son los algoritmos capaces de aprender dentro de ciertos límites éticos? ¿Pueden las máquinas considerar y evitar ciertas situaciones de forma autónoma respetando una serie de marcos éticos? ¿Es conveniente establecer un rígido marco ético

y legal incluyendo un control más férreo de la evolución de las máquinas y designando responsabilidades humanas? El objetivo último de este ensayo será responder también a las cuestiones aquí planteadas, cuya investigación nos llevará a responderlas favorable o contrariamente.

1.2. Justificación del tema

La Inteligencia Artificial, como decíamos, se ha convertido en parte esencial del día a día en la mayoría de los países del mundo. No hay nadie que no haya oído hablar sobre la nueva era tecnológica que nace con ella. Sin embargo, cuando se profundiza sobre la Inteligencia Artificial, surge una cierta desconfianza y recelo por dos cuestiones en particular: por el lado de la parte más subjetiva, tenemos el hecho de que, aunque conozcamos por nombre la tecnología, no todos somos capaces de entender el funcionamiento de los mecanismos y algoritmos que se utilizan tras la Inteligencia Artificial; por otro lado, desde el aspecto más objetivo, aparecen los riesgos importantes asociados a la Inteligencia Artificial.

Aunque somos conscientes de que nuestra racionalidad no alcanza algunos límites, el ser humano aun así se preocupa tanto por lo que no ve, como por lo que no entiende. Esta “consabida ceguera” es la que nos hace poner en duda la Inteligencia Artificial.

Normalmente, los riesgos que se asocian a una tecnología van en proporción con la magnitud de los beneficios (Floridi et al, 2018). En consecuencia, viendo la precisión y el poder que está adquiriendo la Inteligencia Artificial, podemos esperar que conlleve altos riesgos al mismo nivel en numerosos ámbitos entre los que destacan los derechos humanos, la privacidad de datos y la “brecha de responsabilidad” –“*responsibility gap*” (Matthias, 2004) como desarrollaremos más adelante.

La opacidad que rodea a los datos y a los algoritmos utilizados por la Inteligencia Artificial es una de las mayores preocupaciones en este contexto. Por ello, se extiende también a sus ramas, en las que se encuentra el Aprendizaje Automatizado o *Machine Learning*, que trabaja de forma autónoma en base a una información que percibimos como poco transparente, de la que no sabemos nada y que un uso puramente utilitarista podría derivar en un uso poco moral de la herramienta.

Por los riesgos descritos, es importante abordar el tema desde una perspectiva ética puesto que no podemos continuar con lo que llamábamos una “consabida ceguera”. Es un elemento del que ya no nos podemos despegar y es difícil mirar hacia atrás e

imaginarse un mundo sin utilizar Inteligencia Artificial para ámbitos económicos, sociales o comerciales. Tampoco podremos imaginarnos una sociedad sin *Machine Learning* debido a su rápida proliferación. No obstante, es precisamente por eso por lo que hay que tomar conciencia. Sea como fuere, la tecnología nos ha introducido en su mundo y debemos ser responsables y conscientes de que los límites debemos ponérselos nosotros.

Debido a esta reflexión, la actualidad, la relevancia y la urgencia de este asunto, la Inteligencia Artificial y particularmente su rama de Aprendizaje Automatizado y Autónomo, son merecedoras de la siguiente investigación y narración.

1.3. Metodología y estructura

Con el fin de responder a todas las cuestiones y llegar efectivamente al objetivo, al mismo tiempo que realizamos un análisis lo más exhaustivo y claro posible, hay que destacar que se seguirán y analizarán numerosas investigaciones en torno a esta tecnología.

En primer lugar, se detallará en qué consiste la Inteligencia Artificial y qué papel juegan los algoritmos. A continuación, se hará una introducción concreta de lo que es *Machine Learning* o el Aprendizaje Automatizado en el marco de la Inteligencia Artificial. Al estar directamente relacionado, continuaremos con una aclaración de la función de *Deep Learning* o Aprendizaje Profundo. Para concluir este apartado, se evaluará objetivamente la dimensión legal y de la ética hasta el momento. La información obtenida se encuentra en ensayos escritos por profesionales que luego se publican en webs académicas y también de una sesión de la asignatura de Seminarios de Cuestiones de Actualidad de ICADE impartida por Ricardo Guerrero Gómez-Olmedo en la que nos presentó la realidad de la Inteligencia Artificial.

En segundo lugar, una vez asentados los cimientos y con los conceptos explicados para la familiarización con los mismos, se estudiará específicamente el *Machine Learning* y su dimensión ética: la historia que une o no une a los dos, el vacío legal y ético que predomina en la relación entre ambos y, para finalizar, el marco teórico de la ética, con especial enfoque hacia las tecnologías. Todo ello se estudiará a través de artículos y libros en formato web de contenido ético y tecnológico.

En tercer y último lugar, se estudiarán casos específicos relacionados con el Aprendizaje Automatizado o Automático y su ética. Teniendo en cuenta las diversas conclusiones que estos ofrecen, se compararán para extraer una nueva conclusión sobre

la viabilidad ética de las máquinas automatizadas y para comprobar si se pueden aplicar valores éticos en el proceso de aprendizaje de algoritmos. De esta forma conoceremos la posible urgencia o necesidad del establecimiento de límites rígidos para *Machine Learning*. Nuevamente, los casos se obtendrán de sitios web académicos y de las noticias más destacadas en relación con la cuestión planteada.

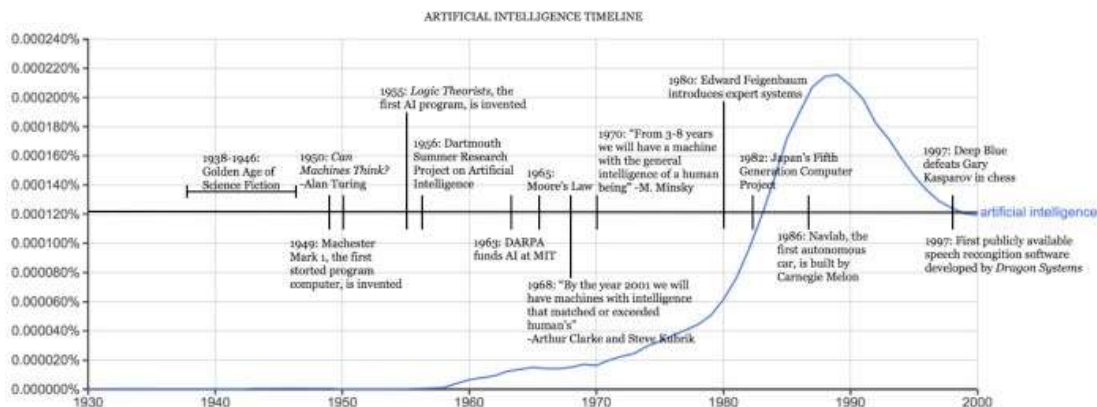
2. Estado de la cuestión

2.1. ¿Qué es la Inteligencia Artificial?

Historia

Como decíamos, la Cuarta Revolución Industrial viene de la mano de la Inteligencia Artificial -*Artificial Intelligence* o AI- cuya idea surge mucho antes de su materialización: con la idea de los robots en películas o novelas -por ejemplo, el Mago de Oz- como máquinas inteligentes. Sin embargo, la fecha en la que propiamente se acuña el término es en 1956. Seis años antes, Alan Turing, un joven británico publicó su ensayo *Maquinaria Computacional e Inteligencia* donde explicaba cómo construir máquinas inteligentes y cómo examinar esa inteligencia a partir de un funcionamiento similar al de los humanos: utilizando información disponible para solucionar problemas y tomar decisiones. (Anyoha, 2017). Sin embargo, el alto precio de los ordenadores en la época retrasó unos años su implementación en el momento.

En la conferencia de Darmouth, DSRPAI -*Dartmouth Summer Research Project on Artificial Intelligence*- en 1956, el anfitrión John McCarthy reunió a los mejores investigadores para discutir sobre la inteligencia artificial a partir del ensayo de Turing, concluyendo todos que era posible llegar a la Inteligencia artificial. Así pues, los siguiente veinte años vieron el florecimiento del desarrollo tecnológico en el ámbito que llegó a su pico en los 80 con la expansión de los algoritmos y, especialmente, de inversiones en el campo. A continuación, tenemos en un gráfico la evolución de la materia:



Fuente: (Anyoha R. , 2017)

Podemos comprobar como para inicios del siglo XXI se habían conseguido prácticamente todos los objetivos, dando cabida a muchas más innovaciones rápidas en

un futuro muy próximo, convirtiéndose el presente en el ayer y el mañana en el hoy. Para el futuro más próximo se espera que el lenguaje artificial -como las llamadas realizadas por una máquina o traducciones simultáneas- sean la realidad de todas las empresas y personas. En definitiva, que las máquinas sean más inteligentes que nosotros mismos.

Definición

La Inteligencia Artificial se ha definido como matemáticas, ingeniería de software, una nueva lengua, psicología, etc., y muchos entendidos parecen coincidir en que no hay una sola vertiente en la que ésta encaje, sino que es tan extensa que tiene muchas respuestas y muchas caras. (Schank, 1987). La enciclopedia británica ha definido este ámbito como «la habilidad de un ordenador digital o un robot controlado desde un ordenador para realizar tareas comúnmente asociadas con seres inteligentes» (Copeland, s.f.); mientras, la Real Academia Española lo define como la «disciplina científica que se ocupa de crear programas informáticos que ejecutan operaciones comparables a las que realiza la mente humana como el aprendizaje o el razonamiento lógico». (RAE, s.f.)

Más allá de las definiciones puramente teóricas, tenemos las de los investigadores pioneros como McCarthy que en 1988 visionó la IA como «un conjunto de métodos para alcanzar objetivos con información disponible, que puedan relacionarse al problema presentado y similares a una combinación de solucionadores como humanos y programas de ordenador». En la ambigüedad del comienzo, ya se planteada como una inteligencia con la propiedad de tener la «habilidad para resolver problemas complicados» (Minsky, 1985) al mismo nivel que el ser humano. Pero algo que está en continua evolución y va cambiando de forma, va modelándose para ser también el «estudio del diseño del espacio de las múltiples inteligencias» (Davis, 1996) y la «integración y sinergias de componentes en un sistema» (Branchman, 2006).

Ninguna de las definiciones es incorrecta ni del todo completa. De forma general, sí pueden servirnos para hacernos una idea del campo en el que se encuentra: en el de la tecnología; así como para saber que se pretende emular con ello acciones propias del razonamiento de los humanos. Sin embargo, el campo es extenso y es difícil dar una sola definición válida puesto que cada día su dimensión es mucho más amplia. De hecho, se han identificado cientos de definiciones para la inteligencia y la IA. (Wang, 2019).

En la dificultad por comunicar una definición completa también radica la dificultad de abordar la IA desde distintos ámbitos. A pesar de ello, la premisa común desde el inicio de la IA es su motivo de creación: una o unas herramientas que sirvan para

crear una máquina inteligente -de pensamiento y actuación racional- que lleve a alcanzar objetivos en base a tomas de decisiones y conseguir, además, encontrar y profundizar sobre la naturaleza de la inteligencia. Esta será la idea sobre la que reflexionaremos y trabajaremos más adelante desde la perspectiva de la ética.

El funcionamiento de la IA

El funcionamiento de la Inteligencia Artificial se inspira en el funcionamiento racional del ser humano. Se tienen en cuenta las distintas abstracciones de la inteligencia humana, de la capacidad mental de nuestros cerebros. Por ello, se hace una simulación del cerebro humano teniendo en cuenta sus “redes neuronales biológicas” (Lake et al, 2017). Esto es lo que se conoce en el idioma anglosajón como *biologically inspired computing*

Para conseguir replicar estas redes, llamadas perceptrones, se introducen en los sistemas informáticos mediante la programación de los mismos, una serie de fórmulas y comandos binarios -que toman valores 0 o 1- que desestiman o aceptan una opción u otra, según el peso de las variables introducidas que toman esos valores binarios. Estas fórmulas que hacen los cálculos se llaman algoritmos, que son la base del funcionamiento de las máquinas inteligentes y que, a su vez, como veremos más adelante, son el elemento básico que mueve al *Machine Learning*. (Agrawal, Gans, & Goldfarb, 2017). Sin ellos, que son el fundamento del lenguaje programado, no tendría sentido hablar de la IA y podríamos considerar que ésta es, potencialmente, el estudio algorítmico del proceso en todos los campos planteables. (Edwards, 2017). Básicamente, se basa en alimentar a la máquina para su funcionamiento y la forma de hacerlo es con los algoritmos y los procesos.

De todas formas, lo que concierne es que sabemos de la existencia de algoritmos, pero no llegamos a conocer su contenido no sólo por nuestros propios conocimientos limitados, sino, recordamos, por la opacidad que envuelve a la tecnología en este aspecto.

Alcance de la IA

Una de las intenciones respecto a la Inteligencia Artificial es que pueda ser explicable su uso, esto es que sea comprensible e interpretable (Edwards, 2017), pero mientras la IA alcanza cada rincón de nuestra realidad, su entendimiento cada vez se percibe como más difuso. (Doran, 2017).

Ahora mismo, pocos ámbitos económicos, sociales o de la salud no utilizan la nueva tecnología. Debido a que con la IA se pueden cumplir objetivos tomando

decisiones en numerosos ámbitos desde la medicina, hasta el marketing, pasando por el diseño y cada vez muchos más, resulta impensable plantearse otra forma distinta de ejercer una actividad profesional -y muchas veces personal- a pesar de no haber recibido formación ni información completa sobre lo que se encuentra detrás.

Las industrias que ha alcanzado la Inteligencia Artificial con mayor éxito y notoriedad son las siguientes: (Forbes, 2019)

- En el sector de la moda: uno de los que más ha aprovechado la herramienta. Se espera que el gasto en la tecnología aumente más de 7 billones de dólares en los próximos dos años por parte de grandes empresas textiles. Este gasto estará enfocado especialmente en la publicidad y en hacer del mundo del *retail* una experiencia pre-venta.
- En el sector de la manufactura: con un potencial ilimitado, la IA va a transformar la industria automatizando gran parte del trabajo humano con menor error y mayor calidad.
- En el sector sanitario: la trayectoria de la IA aún es corta, pero ya se están viendo ciertos avances en su aplicación. En especial se espera una aplicación intensiva en la lectura de rayos X y seguridad en la prescripción de medicamentos.
- En el sector de la construcción: la Inteligencia Artificial con foco en el Aprendizaje Automatizado, el proceso de construcción será más seguro, rápido y menos costoso utilizando el Big Data -los Macrodatos-.
- En el sector del cuidado de la tercera edad: en este caso, a pesar de que la Inteligencia Artificial no es capaz de comprender las necesidades culturales o físicas de las personas, sí puede proporcionar actualizaciones antes recursos obsoletos.
- En la asistencia personal virtual: los famosos *chatbots* se están extendiendo cada vez más.
- En ciberseguridad: al traer consigo un razonamiento complejo y autodidáctico en una escala masiva, puede aumentar la ciber-resiliencia para numerosas empresas que reciben ataques cibernéticos a diario.
- En el diagnóstico de la Salud Mental y Tratamiento: la IA puede servir de herramienta para identificar a personas -especialmente jóvenes- con riesgo de padecer algún tipo de desajuste emocional y recomendar terapia antes de

derivar en una enfermedad de mayor envergadura. Casualmente, se enfoca a los numerosos casos de adicción a aparatos electrónicos.

- En el sector de la educación: creando con ayuda de la IA un camino de aprendizaje personalizado y dinámico en el que todo el mundo pueda encajar.
- En el sector medioambiental: aumento en el uso de drones que son capaces de crear ecosistemas y serán la herramienta más precisa para detectar fugas e impactos adversos.

Sus aplicaciones, como aquí demostramos, son muchas. Se da una gran importancia a minimizar el error de la acción humana en multitud de tareas y se ve como un principal beneficio, lo que para muchos es otra razón para desconfiar de los sistemas, como mencionábamos en la introducción. A pesar de ello, la utilidad sigue primando y para llegar a unos objetivos óptimos, las empresas exprimen y exploran la Inteligencia Artificial y sus instrumentos.

Entre los dispositivos o plataformas que utilizan la Inteligencia Artificial para fines provechosos, encontramos: drones utilizados en misiones de defensa o en empresas de agricultura o energía, las aplicaciones móviles como Uber, las redes sociales - especialmente Facebook, la que comenzó toda la revolución social-, los navegadores de internet en el que destaca Google por sus predicciones utilizando datos de localización, vuelos comerciales que utilizan un autopiloto que integra IA, cuentas de correo que categorizan los emails, sistemas de plagio, lectores-robot, sistemas de toma de decisiones financieras, así como en webs de compras (EMERG, s.f.).

El elemento fundamental para poner en marcha a la Inteligencia Artificial, además de los mencionados algoritmos, son los Macrodatos o Big Data: información de trillones de movimientos en el mundo a través de la tecnología a gran escala, esta información se agrupa y se utiliza la necesaria para cada caso. Es esta información la que también se recopila para hacer de los algoritmos que también se utilizan en *Machine Learning* algo más acertado y para que estos se vayan moldeando de forma automática. Antes de entrar en el fondo del Aprendizaje Automatizado, debemos tener en cuenta que todo *Machine Learning* es Inteligencia Artificial, pero no toda la Inteligencia Artificial es *Machine Learning*. (EMERG, s.f.)

2.2. ¿Qué es Machine Learning dentro de la IA?

Definición

Machine Learning o el Aprendizaje Automatizado, al igual que ocurre con la Inteligencia Artificial, no es fácil aterrizarlo en una sola definición consensuada para todos los públicos. De todas formas, resumiendo numerosas aproximaciones de expertos, podemos decir que es “la ciencia que lleva a ordenadores y computadoras a aprender y actuar como los seres humanos, mejorando su aprendizaje de forma autónoma con el tiempo, alimentándoles con datos e información en forma de observaciones e interacciones del mundo real” (EMERJ, s.f.)

En el entorno de *Machine Learning* los protagonistas son de nuevo los algoritmos: la base para que las máquinas puedan hacer predicciones o determinar un comportamiento. Sin necesidad de una programación explícita en cada eslabón del proceso de la toma de decisiones, los ordenadores o sistemas informáticos actúan -casi podríamos decir que piensan- como los humanos (Ng, s.f.). Mediante la generalización de ejemplos, el algoritmo es capaz de hacer que se desempeñen una serie de tareas y así la máquina en cuestión aprende de la experiencia y de la repetición de esos comandos.

Funcionamiento de Machine Learning

Como decíamos, la Inteligencia Artificial se construye con inspiración en el cerebro humano, ya que se pretende que el funcionamiento sea igual para la máquina. En este caso, las redes neuronales artificiales se construyen como nodos conectados por enlaces. La “capa de entrada” está formada por nodos con valores para variables independientes. Estos valores se escalan para que la información de varios nodos sea comparable y pueda utilizarse para calcular una media ponderada. (Kaplan, 2018)

Los valores de entrada de los nodos de la capa de entrada se conectan a un segundo conjunto de nodos en la capa oculta. Los nodos que siguen a las variables de entrada se llaman “neuronas” porque procesan la información de entrada. Estas neuronas constan de un operador de suma que coteja la información -como una media ponderada-, y la pasa a una función de activación -normalmente no lineal-, para generar un valor a partir de los valores de entrada. Este valor se transmite a otras neuronas de las capas ocultas siguientes.

Un proceso relacionado, la “propagación hacia atrás”, se emplea para revisar los pesos utilizados en el operador de suma a medida que la red aprende de sus errores. (Kaplan, 2018)

El problema es que puede haber múltiples capas ocultas con nodos vinculados y el investigador debe determinar la estructura de la red. Sin embargo, el investigador lo hace cuando el Aprendizaje Automatizado es supervisado, pero no cuando no lo es, por ello es fundamental distinguir los tres tipos de *Machine Learning*:

- **Aprendizaje supervisado**

Utiliza datos de entrenamiento etiquetados para guiar al programa de ML hacia una mayor precisión de previsión. Por ejemplo, sirve para identificar a los manipuladores de ganancias: un programa informático podría utilizarse para identificar patrones que identifiquen a los manipuladores en otro conjunto de datos. Las tareas típicas del aprendizaje supervisado son la clasificación y la regresión. Los algoritmos pueden diseñarse para la clasificación binaria, como clasificar las empresas como probables o no probables, o la clasificación multi-categoría como sería una clase de calificación para los bonos.

- **Aprendizaje no supervisado**

En este tipo de aprendizaje al programa de Machine Learning no recibe datos de entrenamiento calificados -o datos de práctica-; en su lugar, se proporcionan entradas, como pueden ser las características. Sin embargo, no se aplica ninguna conclusión sobre ellas. En ausencia de cualquier variable objetivo, el programa busca la estructura o las interrelaciones en los datos. Es decir, la agrupación de los datos.

- **Deep Learning o Reinforced Learning – el Aprendizaje Profundo y Reforzado**

Los algoritmos de *Deep Learning* se utilizan para tareas complejas como el reconocimiento de imágenes, el procesamiento del lenguaje natural, etc. Los programas que aprenden de sus propios errores de predicción se denominan algoritmos de aprendizaje reforzado. Ambos tipos de algoritmos se basan en las redes neuronales, un grupo de algoritmos de *Machine Learning* aplicados a problemas con importantes no linealidades. (Kaplan, 2018)

Para medir si un modelo es capaz de generalizar y aprender para llevar a cabo futuras tomas de decisiones de forma óptima, los analistas de datos crean tres conjuntos de datos que no se solapan: la muestra de entrenamiento -utilizada para desarrollar el modelo-; la

muestra de validación -utilizada para afinar el modelo-; y, la muestra de prueba -utilizada para evaluar el modelo utilizando nuevos datos-. Los errores de predicción dentro de la muestra se producen con la muestra de entrenamiento, mientras que los errores de predicción en las muestras de validación y de prueba se conocen como “errores fuera de la muestra”. Los científicos de datos descomponen estos errores en los siguientes: (Kaplan, 2018)

- Error de sesgo. Es el error dentro de la muestra que resulta de los modelos con un mal ajuste.
- Error de varianza. Es el error fuera de la muestra que resulta de los modelos sobreajustados que no se generalizan bien.
- Error de base. Son los errores residuales debidos al ruido aleatorio.

Una curva de aprendizaje traza la tasa de precisión (es decir, 1 - tasa de error) en la muestra de validación o prueba frente al tamaño de la muestra de entrenamiento. Un modelo robusto y bien generalizado mostrará una tasa de precisión mejorada a medida que aumenta el tamaño de la muestra. (Kaplan, 2018)

Esta explicación técnica es una de las formas de interpretar el Aprendizaje Automatizado desde un punto de vista funcional, puesto que, con cada nuevo avance, el campo amplía su abanico de posibilidades y de funcionamiento. (Mengnan, 2019)

Alcance de Machine Learning

Al igual que decíamos con la Inteligencia Artificial, *Machine Learning* al ser una parte fundamental en esta tecnología digital, llega a numerosos ámbitos de nuestra vida y a multitud de industrias en las que se espera que prolifere en un corto periodo de tiempo.

Tanto para predecir el rendimiento académico de los estudiantes a través de la elaboración de trabajos previos (Kotsiantis, 2011), como para la predicción del tiempo y los recursos que consumen las aplicaciones móviles utilizando asimismo Big Data (Matsunaga, 2010), también para predecir defectos de software en ingeniería informática (Shepperd, s.f.), e incluso para el descubrimiento de nuevos medicamentos, prediciendo su aplicación o variaciones en el consumo, el uso de distintas técnicas como la validación de procesos o una aplicación intensiva de las redes neuronales artificiales -ANN: Artificial Neural Networks- (Lipton, 2016).

Se puede decir que *Machine Learning* es la máxima expresión de la Inteligencia Artificial, es la rama que más aplicaciones está teniendo y la que mejor se espera que funcione. Es por ello, que se invierte en su investigación de forma constante y podemos encontrarlo específicamente en Internet -en los buscadores y navegadores web para encontrar resultados o para el correo electrónico-, en la medicina -con diagnósticos asistidos para identificación de enfermedades-, para detectar virus tecnológicos, nuevamente para el uso de drones, detección de fraudes, selección de personal o de un segmento de clientes, etc. Estos son por el momento los usos más extendidos o que más llegamos a ver en nuestro día a día, pero el alcance es mucho mayor y sigue avanzando. Esto por esto, que cada vez más personas entienden la necesidad de dar una responsabilidad a las máquinas tanto legal como ética y algunas instituciones ya están trabajando de forma activa en ello.

2.3. Marco legal hasta el momento

Decíamos que la rápida implementación de la Inteligencia Artificial y los cambios que ha generado y generará en multitud de ámbitos, lleva también a la evaluación de sus riesgos en los que resalta la opacidad en la toma de decisiones, los filtros y distinciones basados en el género y otro tipo de discriminaciones e invasión en nuestra privacidad. (Comisión Europea, 2020).

Por razones obvias, la Unión Europea ha visto la necesidad de aproximarse a la cuestión motivado por la Estrategia Europea para la IA de abril de 2018. Basándose en sus valores y la protección de sus ciudadanos, así como en sus principios de cooperación, se contempla una mejora y modificación en el marco legislativo para evitar ciertos peligros. Se tienen en cuenta los siguientes aspectos:

- La efectiva aplicación y fortalecimiento de las existentes legislaciones de la UE y sus países miembros.
- Los límites en el alcance de la legislación de la UE.
- Las funcionalidades cambiantes de los sistemas de Inteligencia Artificial.
- Las actualizaciones en el concepto de seguridad.

Por primera vez, los países miembros de la Unión Europea piden un debate de una manera más acentuada y urgente, recalcando la importancia de poner límites legales y éticos a la revolución tecnológica y del dato. Sin dejar de impulsar la innovación y su

aplicación en las beneficiosas industrias europeas, se pretende construir un marco que garantice la seguridad y garantías del uso de las tecnologías pero, sobre todo, un marco que no fragmente la Unión de Europa. Esta nueva problemática nace a raíz de países como Alemania, Malta o Dinamarca que desde sus respectivas comisiones de datos éticos han puesto en evidencia la falta de amparo legal y los riesgos más acuciantes de la Inteligencia Artificial.

Es cierto que, previamente, ya se contemplaba en el marco legal la responsabilidad humana sobre las máquinas, pero a medida que la IA evoluciona, también deben hacerlo las leyes. Hasta 2020, existían seis leyes específicas sobre la robótica que el propio Parlamento Europeo había promovido:

1. Obligatoriedad de contener un interruptor para emergencias en caso de darse una situación de peligro
2. Prohibición de damnificar a los humanos
3. Prohibición de generarse emociones o relaciones con las computadoras
4. Obligatoriedad de contratar un seguro para aquellas herramientas o máquinas que puedan tener un mayor impacto en la sociedad y en la economía. Los dueños de las máquinas serán los responsables de abonar el coste ante cualquier daño de objetos materiales.
5. Al igual que para las personas, sus obligaciones al igual que sus derechos estarán sujetos a la ley
6. Obligatoriedad de tributación a la Seguridad social. Al suponer un riesgo sobre la mano de obra humana y muchos trabajos, las máquinas inteligentes -sus responsables- deberán aportar a la seguridad social del país para subvencionar el desempleo que originarán. (Iberdrola, s.f.)

A pesar del gran salto que supusieron estas directrices, como decíamos aún hay países insatisfechos con estas medidas e incluso la propia Comisión Europea es consciente de que su papel como responsable de establecer una normativa común ha sido escaso hasta el momento. Es por ello, que podríamos esperar este año o el siguiente nuevos avances en las legislaciones europeas en torno a la Inteligencia Artificial, maximizando así su eficiencia y pudiendo explorar aún más parte del universo que aún no conocemos, pero sin faltar a la integridad o dignidad de las personas.

Además de los esfuerzos europeos por restringir los riesgos tecnológicos de la actualidad, no hay más normativas en papel. Si bien es cierto que encontramos numerosas

propuestas por parte de expertos del derecho y de la ética, aún no hay una solución global ni común (Edrélyi, 2018). De hecho, grandes empresas tecnológicas como Facebook o Google no quieren una regulación para la IA como analizaremos más adelante.

2.4. Debate ético más actual

Los reglamentos y las propuestas de creación de un marco legal para la Inteligencia Artificial surgen a raíz de un planteamiento proveniente de la ética. Al igual que hablamos de la ética profesional, la tecnología digital, al realizar tareas profesionales y al rodear directamente a las personas, también debe cumplir una serie de requisitos y principios que garanticen la calidad y la ética de su trabajo, como ocurre con la colegiación del profesional. (Sáenz, 2018).

Desde el cambio de siglo, cuando la tecnología tomó un giro acelerado, existe una preocupación generalizada sobre la capacidad ética y de valoración de ciertas situaciones específicas y excepcionales que puedan tener las máquinas a las que se les aplica la Inteligencia Artificial y *Machine Learning*. La preocupación hasta el momento ha girado más en torno a al uso que se le dé a la tecnología, no a la tecnología como tal. Es por ello, que se evalúa la ética del profesional de nuevo, de la actuación que el propio humano en la gestión y funcionamiento de las máquinas. (Cheung & Tomlin, 2020)

La cuestión es que cada vez más robots que aprenden de forma autónoma llegan a superar el entendimiento de los humanos. Esto hace imposible su monitorización y se pierde el sentido del control. Es cierto que, hasta el momento, cuando un proceso tecnológico llega a este punto se ha podido parar su evolución. Sin embargo, como cada vez evolucionan más rápido, el proceso puede desarrollarse solo y que escape del alcance de la monitorización. Por ello, en estos momentos se le está dando mucha importancia a profundizar en la ética del profesional encargado de la monitorización de las máquinas y en la posibilidad de crear máquinas no sólo eficaces, sino con consciencia. (Cheung & Tomlin, 2020).

Una de las personas más relevantes al hablar de la ética de la IA es el cosmólogo y físico Max Tegmark, que se refiere a la IA como “tanto lo mejor o lo peor que haya podido experimentar el mundo” (Benefit, 2021), ya que se ve que es un arma de doble filo: lo mismo puede servir para conseguir la mejor calidad de vida que para acabar con el mundo. Su posición, favorable a los esfuerzos europeos por regular la Inteligencia Artificial y la protección de datos al mismo tiempo que hacer un análisis ético, refuerza la idea de preparar la sociedad desde los valores para el uso de la Inteligencia Artificial y

para el conocimiento universal de la misma. Por ello, ha creado un programa, “Through the Wormhole”, en el que intenta acercar la Inteligencia Artificial a la realidad de todos, para que conozcamos sus implicaciones, riesgos y sus beneficios con un buen uso. (Cheung & Tomlin, 2020).

3. El Aprendizaje Automático y la Ética

La Inteligencia Artificial poco a poco está sustituyendo la presencia humana en ciertas tareas. Se habla de la posibilidad de que las máquinas con esta tecnología generen sus propios deseos y estos no tienen por qué estar en consonancia con los deseos humanos. Aquí es cuando observamos uno de los riesgos que puede traer consigo la IA. Otro de ellos es el uso que una persona o entidad dé a la tecnología. Todo ello, debe plantearse desde el punto de vista de la ética y la legalidad como haremos a continuación.

3.1. Principales teorías éticas de la Inteligencia Artificial

Al enfocar la tecnología desde una perspectiva ética, nos encontramos con distintas teorías o ramas que nos pueden ayudar a aproximarnos a la Inteligencia Artificial desde la moralidad.

El debate ético que concierne a la tecnología ha estado presente desde 1940, con las primeras grandes evoluciones en este campo. Ya se empieza a hablar de los “agentes” que se ven envueltos al hablar de la Inteligencia Artificial. Entre ellos, a propósito de este trabajo, destacan los “agentes morales” cuando hablamos de un marco ético. Se intuye que los agentes deben ser interactivos, autónomos y adaptativos. Estas tres características deben ser comunes para todos los agentes de forma que pueda funcionar el proceso de automatización. (Floridi, 2004).

Partimos de la idea que llevamos resaltando desde un principio: la Inteligencia Artificial y los algoritmos del Aprendizaje Automatizado pretenden emular el funcionamiento del cerebro humano. Por ello, tiene aún más sentido someter a ambos a un análisis ético teniendo en cuenta que la ética se basa en valorar el comportamiento humano y su toma de decisiones desde el bien y el mal, desde los valores y la moralidad. Al igual que al profesional de cualquier ámbito, incluido el tecnológico, a las máquinas y sistemas computacionales que realizan labores profesionales se les debe evaluar también a partir de los dos enfoques de la teoría ética: el deontológico -basado en valorar el carácter del acto en sí- y el teleológico -fundamentado en valorar el acto en función de la consecuencia-. (Berzai, 2017)

Cuando Turing visionó esta tecnología, la planteó como un instrumento para solucionar problemas y tomar decisiones. Se diseñó pensando únicamente en cumplir con este ambicioso objetivo, haciendo uso de cualquier dato o mecanismo para hacerlo funcionar y obtener una herramienta mucho más eficaz para llegar a una meta. Por lo tanto, la aproximación más acertada en el caso de la Inteligencia Artificial es la

teleológica o, concretamente, el enfoque utilitarista. Es precisamente la escasa presencia de la deontología, en el inicio, la que se intenta desarrollar a través de propuestas legales y tomando conciencia de la realidad tecnológica en la que nos vemos, intentando complementar e incluso compensar la sola presencia utilitarista que envuelve a la Inteligencia Artificial.

Según el utilitarismo, han de elegirse las acciones que maximicen el beneficio y minimicen el fallo. Tanto el beneficio como el fallo se definen por una función de utilidad. El problema es que, para llegar a maximizar ese beneficio, cualquier medio está justificado. No importa la calidad ética de la acción mientras la consecuencia derivada de dicha acción sea la más positiva y ventajosa. La intencionalidad tras el acto -también conocida como el comportamiento- no es el elemento que se valora, aunque sí se puede expresar esta intencionalidad a través de las interacciones de un agente en ciertas situaciones pasadas. No obstante, de nuevo, se juzga el nivel en el que ha sido de utilidad esa intención para llegar al objetivo previsto. (Hibbard, 2015).

El utilitarismo propio de la Inteligencia Artificial propone la asignación numérica para todo tipo de situaciones, rigiendo sus acciones por valores concretos que distan de la realidad cambiante del ser humano. A pesar de ser valores definidos por situaciones y experiencias, siempre se pueden dar casos no introducidos en el sistema bajo una estimación. Es ahí cuando nos damos cuenta de que la dimensión ética de la Inteligencia Artificial debería ir más allá de la teleología. (Hibbard, 2015).

La pérdida de adaptabilidad al individuo y de captación de particularidades de cada ser humano que propone esta teoría, supone una toma de decisiones por parte de la tecnología que no siempre es acertada. Para empezar, hay una falta de entendimiento de valores subjetivos como son el dolor o la satisfacción que varían enormemente de una persona a otra. Las emociones y sentimientos quedan fuera del plano utilitario en el que emergió la Inteligencia Artificial. (McKibben, 2003)

Del mismo modo, al igual que la tecnología no es capaz de entender esas características particulares e intrínsecas de cada individuo, no todos los seres humanos somos capaces de entender el alcance de la IA ni sus riesgos. Esta reflexión nos lleva a pensar que la teoría utilitarista, aun siendo la más adecuada para hablar de esta cuestión, tiene un límite y se encuentra en la frontera que separa los beneficios y los peligros de la IA que no conocemos. Pese a que su misión es tomar decisiones y ejecutarlas de la forma más efectiva posible, no sabemos si la decisión que ha tomado ahora beneficiosa supondrá

un fallo en un futuro ya que eso no se valora. Al buscarse únicamente la eficacia instantánea, se deja de lado su viabilidad a futuro y su impacto en la vida de las personas.

La perspectiva consecuencialista que venimos detallando, hace que surja la necesidad de construir un marco ético más completo, es importante que el foco ahora centrado en los agentes de las propias herramientas tecnológicas, se trasladen a la realidad de la sociedad. Se deben considerar las siguientes cuestiones éticas esenciales para la creación de un marco comprensible: agentes morales, la consciencia y experiencias, el antropocentrismo, la autonomía, la incertidumbre y complejidad moral, la propia existencia y presencia de los numerosos agentes involucrados, los derechos, los valores, las virtudes, la responsabilidad, la transparencia, la utilidad, la confianza y, por último, los efectos de gran carga moral. (Dameski, 2020). La ética de la Inteligencia Artificial debe, por lo tanto, trasladarse hacia una teoría de análisis más cercana a la Ética del Discurso. (Braidbach et al, 2019)

La estructura ética debe estar fundamentada en la moralidad y tiene que ser coherente, fundacional, híbrida, multidisciplinaria, unificada, bien contextualizada y debe ser aplicable a toda la Inteligencia Artificial y a su interacción con el entorno. (Floridi, 2004). Al final, es la economía, la política y, en definitiva, la sociedad la que convive con esta transformación tan rápida como incierta, que ya tiene un gran impacto en nuestras vidas y que quizás nos encuentre en unos años en una situación atropellada que no habíamos podido predecir al no pararnos a analizar las implicaciones morales.

La urgencia por elaborar un contexto ético que acompañe a la Inteligencia Artificial es sin duda cada vez mayor. Se quiere conseguir que esta sea realmente beneficiosa, pero al escalar tan rápidamente, los esfuerzos tienen que ser aún mayores e inmediatos. Para ello, lo fundamental es conocer primero los beneficios y especialmente los peligros relevantes que puede traer consigo la innovación actual. Esta tarea requiere de una gran investigación y profundización en las bases de la tecnología, nos obliga a predecir y prevenir ciertas situaciones comprometedoras. (Coeckelbergh, 2021).

Fomentar este debate es lo que conseguirá crear unas directrices y una información relevante, transparente y común para que todos seamos conscientes de a qué nos estamos enfrentando y cómo debemos utilizar esta herramienta para efectivamente ser eficaces y moralmente correctos. Dando más importancia a la ética del dialogo, a la deontología y a la figura del ser humano como parte esencial de la robótica y máquinas inteligentes, se puede obtener una herramienta tecnológica justa y fiable.

3.2. Dimensión ética en *Machine Learning*

Como parte de la Inteligencia Artificial, *Machine Learning* está también configurado como un medio que lleva a cabo sus propios procesos de aprendizaje para llegar a una finalidad que es supuestamente beneficiosa para la mayoría.

El desafío que se plantea cuando centramos la Inteligencia Artificial en el Aprendizaje Automático, es intentar explicarlo a través de la creación de una ética de los modelos computacionales. Al ser muy ambiguo, primero se intenta explicar la sociología de la supervisión de este aprendizaje, es decir, se habla primero de la ética del profesional que ha desarrollado la inteligencia artificial de la máquina y luego de como la sociedad acepta y percibe estos avances. (Greene, 2019).

Comentábamos que, en *Machine Learning*, las redes neuronales son las responsables de aprender de la información dada en forma algorítmica para desarrollar reglas que discriminan o aceptan patrones en distintos ámbitos. Al existir muchos conjuntos de datos para entrenar a esas redes de forma gratuita -como la cesión que hacen inconscientemente los usuarios de redes sociales como Facebook- se alimenta a ML sin ningún tipo de coste y sin filtro alguno. Esto provoca que sea la sociedad misma la que se ve directamente envuelta en el Aprendizaje Automatizado sin ser consciente de ello.

Además de la importancia que tiene la sociedad en la dimensión ética del Aprendizaje Automatizado al ser un sujeto importante en su evolución y también ser parte desarrolladora del mismo -ya sea como informático o como cedente de datos de forma activo o pasiva-, debemos tener en cuenta, asimismo, las tres características de la misma: el talento, la potencia y los datos. Estos tres pilares conforman la base para crear modelos en *Machine Learning* que pueden actuar en base a reglas establecidas con anterioridad -aprendizaje supervisado- o a través de normas que van estableciendo las propias máquinas a medida que avanza su proceso -aprendizaje no supervisado-. (Greene, 2019)

Las tres características que acabamos de mencionar tienen implicaciones distintas en cuanto a la ética y se relacionan con tres sujetos: el profesional que crea el sistema, los algoritmos que son creados por el profesional y la información que alimenta a la máquina y es “digerida” por los algoritmos.

3.2.1. El talento (el papel del profesional)

En primer lugar, el talento se puede relacionar a la habilidad del profesional que programa y supervisa, o no, el Aprendizaje Automatizado. (Hibbard, 2015). Su papel ha sido primordialmente regido por el utilitarismo propio del ámbito, pero se le da cada vez más importancia a su papel deontológico y moral. Se apela a su responsabilidad como informador y mitigador de los daños que puedan causar estas máquinas. Asimismo, como ocurre en el caso europeo, aunque sea escasa la regulación de la IA, esta ya manifiesta especialmente la competencia del profesional al controlar los efectos del Aprendizaje Automático, como el verdadero responsable de su buena o mala conducta.

Es muy difícil que un instrumento informático al que aún no se le ha definido una personalidad jurídica ni moral, tenga también responsabilidad jurídica y pueda ser sancionado, como profundizaremos más adelante. Por eso, hasta el momento, debemos tener claro que seguimos siendo los humanos a los que se les puede pedir un compromiso legal y ético, incluso cuando multitud de profesionales no estén de acuerdo. Aunque haya un desarrollo en la robótica creando incluso figuras humanas y se les quiera otorgar una responsabilidad, como es el caso del robot Sophie, su personalidad jurídica sigue estando sujeta a debate, sus propios desarrolladores quieren desprenderse de su responsabilidad. (Lean, 2019). Aun así, todavía intentando solventar el problema de adjudicar una personalidad legal para las máquinas en vez de únicamente para sus controladores humanos, sí podemos hablar de las implicaciones éticas más allá del papel del hombre.

De todas formas, lo que acabamos de mencionar debemos hacerlo volviendo a la idea de que también es el ser humano quien ha desarrollado conceptos como los algoritmos o los que hemos creado los datos para llegar a un fin. Con ello, queremos recordar que el ser humano en este ámbito ha sido conscientemente utilitarista. Por consiguiente, no podemos despegarnos del todo de la responsabilidad del profesional y de su moralidad al hacer uso de los datos y de los procesos informáticos.

3.2.2. La potencia (el papel de los algoritmos)

Una vez matizado este aspecto, nos paramos en la segunda característica de *Machine Learning* que es el poder y la potencia que tiene esta gracias a la existencia y el alcance de los algoritmos. Estos son capaces de predecir la raza, lo que leeremos, el tiempo, etc. Estos algoritmos se definen como neutros por las empresas o entidades que los usan y no hacen uso de datos sesgados o inadecuados. (Martin, 2019). Sin embargo, en ocasiones

se ha discriminado a candidatos por su raza o la máquina les ha identificado como animales sin tener en cuenta sus particularidades o habilidades.

Lo que se plantea es si se pueden incluir algoritmos basados en valores morales - en el aprendizaje supervisado-, o si se puede configurar una máquina para que los desarrolle ella misma -en el aprendizaje no supervisado.

En el primer caso, esto podría ser posible, y en muchos casos ya lo es, con un control continuo. No obstante, ya carece de sentido hablar de una serie de algoritmos no sesgados y neutros, puesto que la intencionalidad del desarrollador estaría presente y la responsabilidad ética de este sería inseparable. (Desai & Kroll, 2017). La ventaja en este caso es que, de crear un marco legal más completo enfocado al profesional informático, el algoritmo desarrollado no va a poder distanciarse de las premisas pautadas dentro de un contexto legal y ético de su desarrollador.

En el segundo caso, es más ambigua la relación entre la creación de valores morales y el aprendizaje totalmente autónomo. Partimos de la implementación por parte del tecnólogo de unos algoritmos iniciales que asumen la forma que tomará el dato y la situación basada en experiencias anteriores. El algoritmo va luego creando él solo sus modelos de forma estadística que, por ende, no contemplan una forma ética, sino numérica y utilitaria (Akrich, 1993).

El cuadro moral en el que se desea introducir a los algoritmos no es otro que el marcado por la transparencia, la justicia y la capacidad explicativa para su propia naturaleza en la toma de decisiones. Esta necesidad de enmarcar a los algoritmos vuelve a recaer de nuevo en el ser humano: en los grupos de investigación, políticos e incluso la población. (Lepri et al, 2018). En este contexto tenemos la OPAL (“Open Algorithms”), un grupo que surge por la unión de expertos en la materia de las universidades e instituciones más prestigiosas del mundo como son el MIT, Orange, el Imperial College de Londres o la alianza Data-Pop. Su organización sin ánimo de lucro pretende presentar la revolución digital y del dato de forma clara, evidenciando la opacidad del sector privado en el uso de estos datos y sus repercusiones positivas si fueran de dominio público. Se dedican a elaborar una imagen más cercana a la realidad humana para proporcionársela a los grupos influyentes que pueden proponer unos estándares éticos. (OPAL, s.f.).

La misión de la OPAL es ser un garante de información verídica y cristalina frente a la cerradura de la explicación que han puesto los más interesados económicamente sobre el funcionamiento de sus algoritmos. Igualmente, quieren construir un sistema

tecnológico-institucional con el foco puesto en la ética y ser el impulsor de una nueva estructura de instituciones que den un apoyo ya no tecnológico, sino humanístico al uso de la Inteligencia Artificial. (OPAL, s.f.).

Antes de pasar al siguiente factor, es importante que recapitemos y tengamos claros los dilemas éticos que se extraen de estos elementos: la cuestión de la transparencia y la opacidad, la complejidad de su entendimiento, los sesgos en su configuración y la posibilidad de cargar a los algoritmos con un factor moral. (Mittlestadt et al, 2016).

3.2.3. Big Data (El papel de los datos)

Carece de sentido hablar de los algoritmos sin hablar de su motor fundamental, que es a su vez el cimiento de *Machine Learning* y, por supuesto, de la Inteligencia Artificial. Una cosa es la evolución del algoritmo y su refinamiento, y otra es el tipo de información o datos que un sistema o un software recibe que serán luego sujetos a análisis a partir de los algoritmos y las funciones incluidas.

Los datos son la fuerza que mueve a la máquina a formarse, a tomar decisiones y, en definitiva, a ofrecer predicciones en base a modelos de repetición que se obtienen de esa información. El tráfico masivo de datos, universalmente conocido como “Big Data”, da sentido a la Inteligencia Artificial y especialmente al Aprendizaje Automatizado.

Actualmente, con un 103% de conexiones móviles sobre el total de la población, que se traduce en 1,53 conexiones móviles por persona (Shum, 2020), el tráfico de datos que nos encontramos es de tal magnitud que resulta inabarcable para las personas. Es por ello, que tiene sentido la aplicación de las máquinas inteligentes con su función clasificatoria de la información a través de los algoritmos. El denominado “terremoto de datos” (Alpaydın, 2020) vienen a través de este uso intensivo de la tecnología y el internet de las cosas, destacando especialmente la presencia de la población mundial en las redes sociales: la mayor fuente de datos que existe.

La recogida de multitud de datos no sólo es rápida cuando se tiene una red social, sino que además es gratuita y muy beneficiosa económicamente hablando, ya que los interesados en estos datos son todo el tipo de empresas e incluso los gobiernos y políticos. Gracias a este flujo de información los algoritmos llevan a los ordenadores a ser muy potentes, de una forma que no es fácil de comprender, así que de nuevo tenemos un problema de transparencia.

Venimos hablando con lo mencionado de la *ciencia de datos*, que no sólo implica un análisis de los mismos. El proceso incluye la recolección y preparación de la información para su posterior análisis e interpretación, que nos llevará a los resultados. Este proceso puede ser fácilmente procesado por los sistemas informáticos. Sin embargo, el reto se encuentra en la comprensión humana del proceso. El conocimiento humano de los profesionales especializados debe estar formado para entender la lógica del proceso que muchas veces uno mismo había diseñado en un principio, pero a medida que la máquina aprende, este proceso es más complejo. (Coeckelbergh, 2021).

En el apartado anterior, hablando de los algoritmos, decíamos que hay una pretensión de que estas fórmulas que crean modelos sean neutras y esto es lo que se nos intenta transmitir. La realidad, a pesar de ello, es que sí están sesgados y esta cuestión se da así también porque al obtener los datos, se refina el método para extraerlos de la forma más efectiva. (Kelleher & Tierney, 2018). Añadimos que nuestra realidad además se ve distorsionada en las redes informáticas y al estar representada de forma abstracta. Por ello, una máquina por el momento no puede ser capaz de tomar decisiones realmente acertadas para situaciones de la vida real, que es lo que se pretende. Esta observación da pie a cuestionar con mayor ojo crítico las elecciones que toma el Aprendizaje Automático. (Coeckelbergh, 2021).

Con los datos aparece un nuevo dilema ético, cuyo debate cada vez es más sonado, pudiendo concienciar cada vez más a toda la población que cede sus datos día a día sin reparar en su destino. Estamos hablando de la protección de datos y la privacidad del individuo. En la actualidad de *Machine Learning*, esta es la cuestión más polémica y que más forma va tomando en la mente de los usuarios de internet. Es el uso de la información personal sin consentimiento y sin conocimiento del individuo lo que se ha estado ocultando y explotando económicamente. El peligro es cuando esta información compromete al individuo y se utiliza para controlar a la población más vulnerable ante la Inteligencia Artificial.

El trato que se le da a la información recogida depende en última instancia del ser humano, que, aunque en el Aprendizaje Autónomo no tenga implicación directa en cada eslabón que compone el proceso cognitivo, es el que ha programado a la máquina con unas características y no con otras, para obtener de ella un objetivo específico. Volvemos, con ello, a exigir la responsabilidad jurídica y moral de las personas desarrolladoras y de los analistas de datos. Estos actores, muchas veces han manipulado nuestras decisiones como consumidores y, sin nosotros conocerlo, también han explotado nuestro tiempo

invertido en las redes llegando a viciar nuestro uso para que utilicemos aún más las redes sociales y páginas web de forma que incrementemos nuestra actividad proporcionando datos. (Marcuse, 1991).

En definitiva, el Big Data se ha convertido en el petróleo de la nueva era: un recurso ilimitado, muy lucrativo y con un entorno muy opaco. Sin embargo, cuando su uso se centra en el objetivo de maximizar el beneficio económico respaldado por una ética utilitarista, se puede desatar una nueva Guerra del Golfo.

Debemos tener presente que solución para garantizar una correcta aproximación moral al Aprendizaje Automático sigue sin ser universal ni consensuada, y todavía se discute quién debe ser el que decida los objetivos y principios de esta materia. ¿Debe ser una decisión tomada por todos los usuarios de la Inteligencia Artificial? ¿Por los que más perjudicados pueden salir? ¿Por los interesados en su explotación económica?

Recientemente, sí que parece que hay una idea común al profundizar en la Inteligencia Artificial basada en que los valores de ésta estén también alineados con los valores humanos. (Russell, 2019). Sin embargo, este tipo de iniciativas no llegan tanto por las empresas privadas como por grupos de interés o fóruns creados para presionar a gobiernos. Este es el caso del fórum AI4People que, con una misión similar a la OPAL, pretende moldear el impacto social de las aplicaciones de la Inteligencia Artificial.

Este es un movimiento europeo que nace en 2018 y un año más tarde la Unión Europea define también sus objetivos éticos en torno a la nueva era tecnológica. A pesar de haber recomendaciones sin tomar una decisión o definición concreta, se consideran como desafíos que deben ser abordados la protección de la privacidad del dato, la transparencia, la necesidad de clarificación de los conceptos. Como hablamos del contexto político europeo, es también central la coordinación entre todos los países de la Unión como actores fundamentales para obtener una decisión consensuada del marco moral y quizás legal en un futuro.

Para culminar, la Unión Europea como institución política, internacional y de derecho, hace un llamamiento a las potencias mundiales como China o Estados Unidos y a organizaciones internacionales para llegar a un acuerdo que posicione a la Inteligencia Artificial en el perímetro de una ética más centrada en el ser humano que en la eficacia de la máquina, sin dejar atrás la propia responsabilidad del profesional ni las oportunidades que ofrecen las ramas como el Aprendizaje Automático. (European Parliament, 2019).

3.3. Vacío de un marco legal común para *Machine Learning*

Uno de los sentidos de la ética es crear un debate que justifique y derive en la elaboración de una serie de normas de comportamiento moral comunes. Decíamos que la Unión Europea ya ha hecho algún esfuerzo en este aspecto con sus seis leyes de robótica. A pesar de este avance, si hay algo que nos queda claro tras un análisis exhaustivo de la Inteligencia Artificial, es que esta es muy amplia y comprende muchos elementos, muchas ramas y muchos potenciales problemas.

En algunos países con un mayor desarrollo digital, como es el caso de Dinamarca, el reglamento se centra en la propiedad de la tecnología alegando que la posesión de un derecho de explotación -patentada- incluye también la obligación de responder legalmente y moralmente ante ello si se es un agente externo a la entidad que tiene dicho derecho que ha extraído información no acordada. El gobierno corporativo es un actor clave para la garantía del uso sensato de la tecnología y especialmente de los datos dentro de la organización, pero su capacidad de respuesta legal ante las decisiones que toman sus sistemas informáticos no es evidente. (Minseen, 2020).

Recordamos que el debate más urgente de los últimos años era el de la privacidad de los datos y es ese el componente de la Inteligencia Artificial y de *Machine Learning* al que más importancia se le da en el derecho de los países que han ido tomando conciencia. El obstáculo es que el marco legal de los distintos gobiernos tampoco llega a ser del todo conciso sobre el contenido concreto de la información que ha de ser bien utilizada. Se habla de los distintos factores como independientes, sin clarificar las funciones del Aprendizaje Automatizado y no exponiendo las sanciones ni los riesgos concretos menos obvios que se pueden encontrar.

Ante un contexto marcado por la incertidumbre legal y la escasa concreción de pautas para un correcto uso de *Machine Learning*, se requiere la colaboración de los investigadores informáticos para obtener unas políticas más razonables y basadas en resultados empíricos. Por un lado, si estos trabajasen junto a abogados y responsables políticos, se podría tener un mayor entendimiento sobre los ataques y las defensas específicas que difieren de aquellas de sistemas tradicionales (Kumar, 2018). Por otro lado, los sistemas de *Machine Learning* podrían construirse con un sistema de identificación de ataques y de peligros regulados. Por último, en este grupo de perfiles distintos se podría elevar la importancia de la protección de los derechos humanos, la libertad de expresión, etc. para luego aplicarlos a los sistemas digitales. (Chen, 2017).

Una unión recurrente entre estos profesionales para llegar a un reglamento internacional y común podría facilitar la resolución de sentencias como la del caso Google España y la cuestión del “derecho al olvido”. El “derecho al olvido” consiste en el derecho que pueden ejercer las personas cuando desean que, en un buscador como es Google, no aparezcan enlaces web a datos individuales como forma de proteger la privacidad.

Google no reconoció el “derecho al olvido” impuesto por España y, además, el concepto no estaba del todo desarrollado por la Agencia Española de Protección de Datos, presentando todavía ciertas lagunas. El conflicto se llevó al Tribunal de Justicia de la Unión Europea, donde se aplicó la Ley Europea de Protección de Datos a favor del “derecho al olvido” y poniendo como responsable a Google, al ser el último decisor en el tratamiento y determinación de esos datos. Sin embargo, como Google es una empresa americana, el “derecho al olvido” no se le aplicó para todo su dominio, sino únicamente fue reconocido en el ámbito europeo. Ahora bien, con esta sentencia, España introdujo de forma concreta este derecho en su reglamento y supuso un giro a favor de la necesidad de regular el dato y la tecnología digital. (Álvarez, 2014).

Este acontecimiento nos puede servir como referencia para llegar a la reflexión de que debe haber un organismo regulador común para la protección de la privacidad y de los valores del ser humano al utilizar una máquina. Aunque, también nos hace pensar en la cantidad de información que se ha aprovechado previa imposición de este derecho. Adicionalmente, resulta también inquietante el hecho de que, por no tener una regulación rígida, ni clara, ni una información completa, puede haber muchos casos similares que aún no han salido a la luz y sobre los que no se ha discutido.

Es por este último motivo por el que debemos reforzar la idea de la importancia de crear grupos de investigación de distintos perfiles. Sólo así cada evolución tecnológica podrá estar sujeta a una buena práctica y la responsabilidad de cada actor estará bien definida, circunstancia que ahora mismo está difusa como explicaremos a continuación.

3.4. The responsibility gap

A lo largo de este ensayo, hemos podido observar en numerosas veces la presencia de la responsabilidad al referirnos a las consecuencias éticas y legales. Esta es una cualidad propia de las personas y es uno de los principales valores éticos.

La responsabilidad implica el cumplimiento de una serie de pautas, así como el cuidado de los demás y de uno mismo en los distintos ámbitos de la vida. Asimismo, ser responsable está conectado con la idea de responder ante las consecuencias de los actos de uno mismo. Para asumir esto, es importante ser consciente de los actos y las máquinas a las que se les aplica Inteligencia Artificial que, decíamos que funcionan de forma automática sin reparar en los actos, sólo ofreciendo un resultado óptimo.

¿En quién recae, por tanto, la moralidad de las acciones? Para poder considerar sujeto moral a un elemento este debe poder «formar intenciones, tener emociones, consciencia y realizar acciones voluntarias sobre las que luego reflexionar» (Coeckelbergh, 2021). Por ello, en el sistema legal español, tenemos la figura de la tutela en ámbitos domésticos, otorgando la responsabilidad de un animal o un menor de edad a su responsable.

Sin embargo, parece mucho más fácil y menos comprometedor decir que una máquina está programada para ser consciente y para aprender a forjar sus propios valores, de manera que la única responsable de sus actos sea la propia máquina. Así sus desarrolladores no se ven envueltos en debates judiciales y no hay responsables a los que pedir una indemnización por daños.

Esta situación es la que se conoce como “*responsibility gap*” o brecha de responsabilidad. Al igual que hay lagunas legales, hay un vacío cuando se trata de encontrar el agente que debe asumir las consecuencias de sus resultados no siempre éticos ni beneficioso.

Pongámonos un ejemplo un tanto extremo: cuando a una máquina se le encomienda que un grupo de 5 personas infectadas por el Coronavirus no contagie al otro grupo de 5 personas que se encuentran en la misma estancia, la máquina lo hará de la forma más rápida y eficaz posible. Entre sus dos opciones está el proporcionar mascarillas y desinfectantes y asegurarse que todo el mundo cumple con las medidas sanitarias, o matar a las 5 personas infectadas sin ningún tipo de coste ni espera de tiempo. Un dispositivo cuyo objetivo es realizar la tarea mucho antes que cualquier otro agente, ¿sería capaz de ver más allá de la eficiencia? Si recordamos, las máquinas no forman intenciones ni acciones voluntarias, así que lo más probable es que no. ¿Quién sería, entonces, el responsable de la muerte de 5 personas?

Nosotros sí podemos llegar a este razonamiento, pero no podemos afirmar que las máquinas lleguen más allá y, desde luego, no son conscientes de haber violado nuestras

premisas éticas, sobre todo cuando están respaldadas por la concepción utilitarista (Champagne et al, 2015).

En el dilema que presenta Sparrow, se añade un nuevo factor a la definición de un agente que puede ser moralmente responsable: debe tener también la capacidad de ser premiado o castigado a través de una remuneración o sanción proporcionada. Es verdad que una máquina puede llegar a estar programada para su autodestrucción como una habilidad frente a los mayores riesgos, pero ¿en qué situaciones concretas se daría el caso?

Otro dilema que se plantea es el gran número de agentes que se pueden ver envueltos en el uso de esta tecnología. Para empezar, tenemos al desarrollador y programador de la máquina y luego tenemos a los usuarios, cuya cadena puede ser muy larga. Si, por ejemplo, en una misión militar se hiciese uso de una de esta tecnología por un oficial y el resultado fuese devastador, se plantearía el dilema de colocar la responsabilidad a la máquina, al desarrollador o al militar. Como no sabemos si el problema ha sido de información sobre el uso o la programación o toma de decisiones de la propia máquina, es complicado saber a qué agente elegir la responsabilidad.

Una de las propuestas más aceptadas para intentar solventar la brecha de responsabilidad, es el consentimiento informado en cada parte de la cadena que involucre la Inteligencia Artificial, *Machine Learning* y el uso de datos. (Krishnan, 2009). Con el consentimiento del uso, se aceptaría también la responsabilidad del uso y esta recaería en el ser humano, que es el que tiene la capacidad de ser juzgado. De hecho, se podría pensar en la idea de proponer el consentimiento de responsabilidad como condición exclusiva para el uso de la IA y sus ramas, especialmente en determinados casos que ahora analizaremos.

4. Casos que apoyan la capacidad de *Machine Learning* para actuar éticamente

Desde un principio, veníamos hablando de la dualidad que caracteriza a la Inteligencia Artificial y a *Machine Learning*: son herramientas que traen consigo grandes beneficios, pero, precisamente por ello, conllevan también muchos riesgos al mismo nivel.

Sin duda, lo primero que resuena son sus beneficios. Gracias a las bondades del uso de esta tecnología digital y su potencial alcance, se han invertido muchos recursos en su investigación para poder abarcar incluso más ámbitos que mejoren la calidad de vida de la sociedad. *Machine Learning* es en algunos casos un sinónimo de justicia, de salud y de bienestar. Específicamente, nos centraremos en el análisis de dos casos: el primero relacionado con su papel en una resolución judicial y, el segundo, vinculado a los avances y predicciones médicas.

4.1. Caso 1. *Machine Learning* como testigo en juicios de violencia y asesinato

Los algoritmos de *Machine Learning* más presentes en nuestro día a día doméstico, son los aplicados a los recientes asistentes virtuales. Estos, mientras estén conectados o tengan batería, están en funcionamiento constante. Es una de las fuentes de datos e información más sustanciosa ahora mismo, que puede hacer que las empresas conozcan cada paso que damos, lo cual genera un patrón para cada individuo y se puede llegar a él fácilmente. Con toda la información que hemos dado anteriormente, inmediatamente salta la alarma de la protección de datos, pero como en el siguiente caso, estos se pueden utilizar para apoyar ciertas situaciones críticas.

En 2017, una mujer americana recibió a su pareja abusiva en casa e iniciaron una pelea en la que el hombre comenzó a pegarle a la mujer, amenazándole a punta de pistola. En un instante, el hombre le gritó de forma violenta a la mujer a ver si había llamado a la policía. Dio la casualidad de que en su domicilio había un asistente Alexa -de Amazon-, un dispositivo conectado constantemente a la red que tiene muchos micrófonos y funciona por medio de comandos por voz y términos clave. Al escuchar “llama” y “policía”, el dispositivo conectó directamente con la oficina de policía de la zona que entendieron la pelea y tomaron medidas inmediatamente, presentando las evidencias auditivas más tarde en el juicio por violencia de género. (EFE, 2017)

Un caso similar, en el mismo país, pone de manifiesto el papel de testigo de Alexia. También en el mismo año, un hombre mata a dos mujeres a cuchillazos. En la escena del crimen no hay huellas, lo único que se encuentra es el asistente virtual de Amazon. El juzgado pide entonces al diversificado *e-commerce* que revise las grabaciones o los dispositivos que pudieran estar vinculados a su asistente. Ante esta circunstancia, Amazon como protector de datos, requiere del juzgado una demanda legal y obligatoria para la adquisición de estos datos (ABCNews, 2017).

Evidentemente, vemos que este patrón trae también consigo el dilema de la privacidad. Además, podemos intuir el sentido de la responsabilidad que tienen ciertas empresas. Sin embargo, podemos ver el primer caso como un ejemplo de excepción respecto a la protección de la información y como una forma de compensar la mala práctica en relación a la escucha constante, con la bondad moral y ética de haber salvado probablemente una vida y de haber hecho justicia. Asimismo, el segundo caso puede transmitir una imagen de empresas como Amazon al mismo tiempo leal como opaca.

A raíz de este suceso, Alexa y otros asistentes virtuales como Siri, se han transformado en armas para combatir la violencia doméstica y de género. Salvando así muchas vidas. Esta propuesta está impulsada también por la ONU en los Derechos Humanos de las Mujeres. Se trata, al final, de dar la vuelta a el tratamiento de datos, destacando la importancia de los mismos para identificar ciertas conductas que pueden acabar en graves injusticias.

Machine Learning aplicada a estos asistentes se ha convertido en el instrumento más adecuado para predecir la violencia doméstica. Ya no es sólo cuestión de distinguir ciertas ofensas o conflictos que pueden darse en el domicilio o en lugares de trabajo. Ahora *Machine Learning* es capaz de distinguir, mediante conductas cotidianas, a las personas que pueden llegar a ser maltratadoras basándose en patrones pasados. A través de una serie de variables, incluyendo más de 30.000 casos, la máquina ya empezó a funcionar para dar estimaciones sobre arrestos asociados a violencia física, los no asociados a ella y los no arrestos con presencia de violencia doméstica.

Con esos datos también se podía hacer una predicción de la futura peligrosidad del agresor una vez cumplida su condena y puesto en libertad. Los resultados y la investigación llevada a cabo durante un periodo de tiempo relativamente largo evidencian que de usar *Machine Learning* para predecir, el fallo por absolución de la

condena sería menor en la mitad. La máquina había sido capaz de predecir qué perfiles volverían a cometer actos violentos y volverían a ser arrestados. Se estimó que cerca de 2.000 nuevos arrestos posteriores a haber cumplido condena se podrían haber ahorrado en ese periodo. Es decir, se podría haber evitado que 2.000 personas hubiesen vivido una situación de violencia doméstica. (Berk et al, 2016).

Por último, otro ejemplo que tenemos respecto a la bondad del uso de algoritmos en el Aprendizaje Automático en situaciones de violencia física o verbal es el de Facebook. Esta red social, muy criticada por su poca transparencia en el uso de datos y modelos tecnológicos, utiliza una serie de filtros para detectar mensajes de odio en su red. En constante examen, se pretende conseguir una herramienta verdaderamente eficaz que garantice un lenguaje correcto y un uso ético del usuario final, contemplando la opción de sancionar. La intención bajo esta premisa es promover la moralidad y el respeto en nube entre individuos y grupos, pero cuando el impulsor es Facebook -la empresa más enriquecida gracias a la falta de entendimiento de la sociedad sobre los datos-, la propuesta no se percibe como algo tan positivo. (Vigna et al, 2017).

4.2. Caso 2. Machine Learning en la Medicina

El ámbito de la sanidad y la medicina es tan importante como sensible. La carga ética que el ejercicio de la medicina conlleva probablemente sea de las más altas de entre todas las esferas de la sociedad. Al ser una profesión tan enfocada en las personas y en algo tan valioso como su salud y sus vidas, no es de extrañar que la cuestión requiera una delicadeza especial. *Machine Learning* aquí aparece como complemento del ejercicio médico, más que como un sustituto. Y aunque ya se empieza a hablar de la precisión robótica en cirugías normalmente realizadas por humanos, analizaremos los avances más controlados hasta el momento.

En primer lugar, tenemos la presencia de la tecnología digital en el tratamiento y prevención del cáncer. Al día mueren cerca de 300 personas únicamente en España, que más allá del coste económico que asciende a unos 19 millones de euros, están las implicaciones emocionales.

Machine Learning, sin necesidad de utilizar técnicas invasivas y mucho más costosas como la endoscopia, es capaz de desarrollar un modelo predictivo para el diagnóstico del cáncer gástrico. Como pasa con todos los tumores, lo mejor es

descubrirlos y tratarlos a tiempo y los algoritmos médicos introducidos en este sistema ayudan a detectarlos incluso antes de su total formación. El cáncer gástrico es cada vez más común y su menor o mayor impacto depende de que se encuentre rápido. Teniendo en cuenta variables de la sangre, historial médico familiar y personal y el modo de vida, entre otras variables, se consigue una eficiencia y precisión en la predicción del 83%. (Zhu et al, 2020).

En cuanto a la extensión de los carcinomas -que se inician en tejidos internos o externos- especialmente en el hígado, se ha empezado a investigar sobre los beneficios de utilizar el Aprendizaje Profundo, o Deep Learning para la detección y clasificación de este tipo de cáncer. Al igual que el cáncer gástrico, esta es una enfermedad también muy común en el mundo. Durante ocho años, se estuvo investigando sobre su aplicación, adaptando los métodos y técnicas para construir unos algoritmos adaptativos más útiles. Nuevamente, la precisión ha resultado ser consistentemente alta. (Bousabarah et al, 2021). En todo caso, este desarrollo no queda aquí en lo que al cáncer se refiere. También puede predecir, a través de sistemas más sofisticados con ultra-sonidos, los depósitos tumorales antes de las operaciones e incluso antes de las terapias, para poder ajustar luego, a criterio del profesional médico, la mejor terapia y dosis de medicamentos para el paciente.

El objetivo de *Machine Learning* en la medicina es, en definitiva, aumentar la habilidad del profesional al identificar las regiones con tumores para poder valorar de forma más óptima la necesidad de una intervención y ser lo menos invasivos posibles. El potencial que tiene la tecnología va a más por el gran abanico de métodos que ofrece para acumular y adaptar el conocimiento médico. Eso sí, aunque la claridad no sea predominante en las características de Machine Learning, en el entorno de la medicina el agente moral no deja de ser el profesional. Decíamos que esta es sólo una herramienta que, mientras quizás podría sustituir en determinadas tareas al profesional, no se prevé que las autoridades éticas médicas lo vayan a aceptar en ningún momento cercano.

La medicina ya tiene un marco ético muy potente. Empezamos con el juramento hipocrático que se basa en resaltar la función del profesional como una persona que tiene por objetivos priorizar el bien y la salud del enfermo. Esta concepción de persona a persona es la que refuerza el papel de *Machine Learning* como una herramienta, pero no como un actor activo de las decisiones médicas. En este caso, al localizar la responsabilidad en las personas, se exigirá a los programadores una información mucho

más completa y se le dará una importancia mayor al entendimiento de estas técnicas y a la formación del profesional médico en la dimensión tecnológica. (Magoulas, 1999)

Estos casos, nos pueden servir de ejemplo para configurar la Inteligencia Artificial de una forma mucho más humana en otros campos. La medicina es un claro modelo de un equilibrio entre revolución digital y ética tradicional. El foco sigue puesto en el paciente y el desarrollo es lento, con muchas investigaciones y pruebas que no imponen de forma abrupta un método desconocido que busca la eficacia, sin tener la capacidad de pensar en el bienestar del individuo. Este proceso en vez de dejar las consecuencias al criterio de la máquina, el objetivo y potencial resultado de su proceso cognitivo siempre va a pasar bajo el ojo crítico del profesional, porque él es el responsable de su trabajo y del trabajo de la máquina que está a su cargo.

5. Casos que socavan la capacidad de *Machine Learning* para actuar éticamente

Si llevamos tiempo hablando de la existencia de un dilema ético en el entorno de la revolución digital, es porque al igual que hay muchas ventajas, ya se han percibido muchos de los riesgos que ponen al ser humano en situaciones precarias. La tecnología supone una amenaza para el modo de vida tal y como lo conocemos, especialmente en el ámbito laboral en muchas industrias. Además, aunque las aplicaciones médicas de la Inteligencia Artificial incluyan una buena precisión al tratar la salud mental, gran parte de los trastornos psicológicos han aparecido justamente a raíz de esta ciencia. Como hemos hecho en el apartado anterior, utilizaremos dos circunstancias para observar la otra cara de la IA: el primer caso se centrará en la cuestión de los trabajadores y su futuro, el segundo caso abordará la cuestión psicológica del uso de la tecnología.

5.1. Caso 1. *Machine Learning* en el espacio laboral

La preocupación que se distingue al pensar en la efectividad de las tareas que realizan las herramientas inteligentes, es que lo puedan hacer de forma más rápida, más eficiente y más barata que los humanos. Al hacer una evaluación de esta tecnología, parece que se puede hablar más de una recolocación de los empleados, que de una sustitución parcial a una casi-completa de la mano de obra humana.

Ahora bien, este mensaje tranquilizador no termina de dar solución a las personas más adultas formadas en materias en las que se les está formando a las máquinas ahora. Es verdad que puede crear nuevas oportunidades laborales, pero para los que no consiguen llevar el ritmo del desarrollo tecnológico ¿qué les depara el futuro? (Ilsoe, 2018). La inversión en la tecnología debería ir entonces de la mano de la inversión en la formación de estas personas. No sería éticamente correcto ni moral, dejar sin trabajo a las personas sin darles una solución.

Probablemente, los debates que critican la posición de aquellos que visualizan el desempleo como una de las consecuencias de la Inteligencia Artificial, no llegan a comprender que tampoco se puede hacer una recolocación automática ni fácil. Cuando se ha pertenecido muchos años a una empresa y se es fiel a ella, supone una gran decepción comprobar cómo se es sustituido sin miramientos. Por eso, la aproximación debería ser mucho más gradual que la actual. Ahora bien, sí se podría hacer hincapié en las generaciones venideras de la importancia de una formación tecnológica más allá de su propia experiencia, que se inspire también en la ética de la misma. De todas formas, ¿es

justo pedir a los seres humanos que se enfoquen en materias que quizás no sean su vocación? Aquí de nuevo tenemos una nueva disyuntiva ética.

También en relación con el trabajo, tenemos la polémica del reconocimiento facial y los sesgos discriminatorios a los que se ven sometidos muchas veces los candidatos a un puesto de trabajo por los métodos de elección que proponen los algoritmos. El sistema se basa en el reconocimiento facial, de la formación y la experiencia y clasifica a los candidatos en distintos grupos. Este método es injusto en muchas ocasiones, esto es, que no es equilibrado para los distintos grupos faciales. Los programadores, entonces, se encuentran aquí con una duda: ¿deberían ignorar los valores algorítmicos de la raza y si se produce una desigualdad que sea por mero accidente, o deberían considerar estadísticas específicas que mitigasen el sesgo que hace la propia máquina? Si el deseo es promover justicia, lo adecuado sería lo segundo, aunque podría generar el riesgo de suprimir la meritocracia. Sin embargo, si se sigue manteniendo la idea de adjudicar a la máquina la responsabilidad de sus actos, daría igual una opción u otra desde la perspectiva del programador ausentado de la ética. (Benthall & Haynes, 2018)

5.2. Caso 2. La huella de *Machine Learning* en la salud mental

La dependencia a la tecnología es un problema que afecta especialmente a los jóvenes y adolescentes que se iniciaron pronto en la revolución digital o que, directamente han nacido en su pleno apogeo y no conocen más pantallas que las de los nuevos smartphones.

El creciente flujo de datos o las constantes interacciones sociales a través de las redes disminuyen la concentración, distraen la función psicomotora y llegan a perturbar las horas y la calidad del sueño. (Jacobson & Bentley, 2019). Así como la IA es capaz de predecir emociones humanas y potenciales enfermedades psicológicas, especialmente con los datos que se recogen en las redes sociales, también es la causa de la aparición de estos mismos trastornos.

Facebook es una de las redes que usa algoritmos específicos para ayudar a predecir la depresión. A través de ellos, se llegó a la conclusión de que el uso de webs sociales como esta o Instagram e incluso Twitter se asociaban a un incremento en el riesgo de acabar desarrollando depresión. *Machine Learning* se utiliza aquí para detectar un problema, que de cierta forma ello mismo está generando.

Lo que podemos concluir de este caso es que por mucho que con *Machine Learning* se llegue a la conclusión de que su uso de datos y sus impactos son la causa de un problema, estos algoritmos luego no son capaces de minimizar su efecto. No llegan a entender ni a argumentar que es su propia actuación la que ha desencadenado ese problema y está en sus propios medios revertirla o cambiarla. Este hecho acentúa la falta de responsabilidad que se le podría aplicar a *Machine Learning*, ya que esa asociación entre culpabilidad y capacidad de cambio la podemos hacer nosotros, pero no un sistema informático.

6. Conclusiones

Una explicación profundizada en lo que es la Inteligencia Artificial, *Machine Learning* y sus distintos tipos o la ética, a pesar de lo exhaustiva que pueda ser, sigue sin abarcar cada milímetro de esta ciencia infinita e incierta que se extiende cada día. En el “desgranaje” de cada agente o de cada elemento que ayuda a dar vida a la Inteligencia Artificial encontramos un nuevo reto moral, legal y tecnológico sobre el que tenemos que pararnos para analizar su impacto. De todas maneras, sí podemos recoger cinco ideas principales sobre las que hacer una última reflexión desde el punto de vista ético, humano y de la innovación: la consciencia, la responsabilidad, la transparencia, el consentimiento y la legalidad. De todas formas, al mismo tiempo que nos adentramos en esta meditación, debemos recuperar las preguntas propuestas al principio del ensayo.

En primer lugar, cuando nos preguntamos si los algoritmos por sí sólo serían capaces de aprender dentro de unos límites éticos, debemos tener en cuenta la consciencia. Una máquina que aprende sola, por mucho que quiera emular la mente humana, no es consciente de sí misma porque no deja de ser una máquina. Por eso, *Machine Learning* de forma no supervisada, que es el máximo ejemplo de una máquina autónoma, no sería capaz de conocer por ella misma si está actuando moral o inmoralmemente. Esto es como un niño a quien, si no regañas ni le premias, no aprenderá realmente sobre la ética de sus acciones y actuará en un limbo moral pues no tiene referencias ni del bien ni el mal y simplemente buscará su bienestar físico. (Wachter et al, 2017). Al final, la conciencia es algo que se va adquiriendo a través del entorno y de normas o recomendaciones cívicas que se añaden a la mente de cada uno, pero las máquinas no tienen esta facultad.

Ahora bien, por mucho que esto no sea posible, no podemos menospreciar el alcance que pueda llegar a tener esta tecnología. Como dice Max Tegmark, uno de los expertos de la materia ya mencionados: «debemos aceptar el hecho de que pueda haber algo más inteligente que el propio ser humano». Esta perspectiva, a su vez, nos hace cuestionar el enfoque antropocentrista que podría o no tener la tecnología.

En casos como la medicina, el antropocentrismo es sin duda clave ya que es una materia que se centra en el paciente y en el profesional humano que le atiende. Por lo tanto, tiene sentido no dar más protagonismo a una máquina. Ahora bien, para otras cuestiones, valorar solo el beneficio humano puede dejar de lado cuestiones como el medioambiente o el cambio climático. (Coeckelbergh, 2021). Para que esto no pase, pero

se siga pudiendo dar al ser humano el protagonismo que necesita, el enfoque antropocéntrico no sólo debe contemplar un «estatus moral de primacía único e inalienable en los campos de lo civil, lo político, lo económico y lo social», también tiene que ser el protagonista de lo personal y lo ético. (EuropeanComission, 2018).

Esta concepción antropocéntrica no significa que debamos tomar una posición egoísta. Si nosotros vamos a ser el centro de todo lo que ocurre, siendo las máquinas nuestra herramienta, también vamos a ser nosotros los responsables de lo que suceda. Que algo se desenvuelva a nuestro favor ocurrirá porque nosotros mismos lo hemos querido así, por lo tanto, somos responsables de ello y debemos responder ante las consecuencias.

Podemos entonces decir que los algoritmos no tienen por qué aprender de valores éticos, así que los humanos, que solemos tomar una perspectiva antropocéntrica, tenemos que seguir haciéndolo sin absolvernos de la responsabilidad que ello implica.

En segundo lugar, como decíamos que no está claro que los algoritmos aprendan valores éticos, tampoco está claro entonces que vayan a poder evitar situaciones de forma autónoma dentro de límites éticos. Con la poca transparencia que rodea al flujo de datos y al no consentir su uso, *Machine Learning* y la Inteligencia Artificial ya hacen sus tareas de forma poco ética. No obstante, como no hay un rígido y común marco legal, no resulta sencilla la tarea de exigir responsabilidades. La opacidad que caracteriza a la IA ya presenta un dilema ético, luego a medida que se van conociendo los entresijos en muchas industrias, se descubren cuestiones también preocupantes. Por el momento, no se trata de que la máquina actúe en los límites, sino de que se le permita hacerlo siendo el propio programador o usuario el que sea moralmente correcto. Las máquinas, decíamos, no dejan de ser utilitarias y aunque esta imagen podría ir transformándose para que sea más humana, en ese proceso somos nosotras, las personas, quienes tienen que aportar la parte más humanística y ética hasta que se consiga diseñar este modelo de principios y valores para las máquinas.

Machine Learning llegará a la opción más rápida y que más beneficios económicos o menos gastos vaya a reportar al humano. Sin embargo, en última instancia, la decisión de la mejor opción -equilibrio entre beneficios económicos y humanos-, ahora mismo sólo la puede tomar el ser humano, que es el único que puede evitar ciertas situaciones de peligro y, de hecho, ya lo ha hecho en ocasiones apagando sistemas inteligentes.

En tercer lugar, es evidente la necesidad urgente de establecer un marco ético rígido que luego sirva de boceto para un marco legal unificado e internacional. Al parecer, una de las pocas instituciones que está llevando a cabo esta iniciativa es la Unión Europea. De todas formas, aún tiene un largo recorrido que hacer entre poner de acuerdo a todos los países de la Unión y transmitir este marco a un nivel global. Muchas empresas tecnológicas con oficinas o actividad en Europa son americanas o asiáticas y al final es difícil pedirles respuesta cuando estas se rigen por su propia normativa. Esto solo ralentiza el proceso judicial y emborrona aún más el alcance ético de la Inteligencia Artificial.

Ahora bien, incluir un control muy férreo de las máquinas puede desincentivar la inversión en la innovación en las mismas. Quizás, lo importante sería crear un proceso de producción e investigación con grupos formados por expertos de multitud de ámbitos que vayan cuestionando, desde su perspectiva profesional, el peso de los sistemas en sus propios ámbitos laborales.

Los últimos apuntes que se podrían hacer para concluir con el ensayo es que la responsabilidad tiene que ser no un agente del que los involucrados en la Inteligencia Artificial se quieran librar, sino un impulsor del progreso de la tecnología digital. Las máquinas no son, ni tienen por qué ser malas. Si su uso y su intención por parte del ser humano es buena, entonces serán beneficiosas. Tenemos el gran ejemplo de la medicina, que reconoce que la responsabilidad no recae en nadie más que en el profesional y *Machine Learning* no es la finalidad, sino un medio.

Pidiéndose la responsabilidad de los humanos hasta que la máquina no sea perfectamente independiente y capaz de hacer juicios de valor, puede impulsar mucho más su desarrollo e investigación. Además, cada avance se daría dentro de un cuadro ético y legal innegable en el que las limitaciones éticas se descubrirían antes de ser lanzadas al mundo y habrá una mayor transparencia. Asimismo, si se toman los sistemas informáticos como parte del bien común para el ser humano antropocéntrico, como con la tecnología que detecta tumores, muchas más personas querrán involucrarse y también esto haría que la información llegase a la superficie, sin rodearle tanta niebla.

Para finalizar, *Machine Learning* siendo la rama más cercana y comparable al cerebro humano, evidencia como a veces pasa inadvertida la presencia de algo tan propio del hombre como es su instinto, y su capacidad adaptativa y de razonamiento para distinguir entre lo moral y lo inmoral.

7. Bibliografía

- ABCNews. (2017). *Judge orders Amazon to hand over Echo recordings in double murder case*. Obtenido de KCRG: <https://www.kcrg.com/content/news/Judge-orders-Amazon-to-hand-over-Echo-recordings-in-double-murder-case-500214082.html>
- Agrawal, A., Gans, J., & Goldfarb, A. (2017). What to Expect From Artificial Intelligence. *MIT Sloan Management Review*.
- Akrich, M. (1993). The description of technological objects. En *Shaping technology/building society: Studies in sociotechnical change* (págs. (pp. 205–224)). Cambridge, MA: MIT Press.: In W. Bijker & J. Law (Eds.).
- Alpaydin, E. (2020). *Introduction to Machine Learning, Fourth Edition*. Obtenido de MIT Press: <https://mitpress.mit.edu/books/introduction-machine-learning-fourth-edition>
- Álvarez, C. (2014). Sentencia Google Spain y el Derecho al Olvido. *Actualidad Jurídica Uría Menéndez*, págs. 110-118. Obtenido de Actualidad Jurídica Uría Menéndez.
- Anyoha, R. (2017). *The History of Artificial Intelligence*. Obtenido de Harvard Science in the News: <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>
- Anyoha, R. (2017). *The History of Artificial Intelligence*. Obtenido de Harvard University: <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>
- Benefit, J. (2021). *Max Tegmark: “La Inteligencia Artificial puede ser lo mejor que haya sucedido nunca, o lo peor”*. Obtenido de <https://theconversation.com/max-tegmark-la-inteligencia-artificial-puede-ser-lo-mejor-que-haya-sucedido-nunca-o-lo-peor-154759>
- Benthall, S., & Haynes, B. (2018). *Racial categories in machine learning**. Obtenido de New York University: <https://arxiv.org/pdf/1811.11668.pdf>
- Berk, R., Sorenson, S., & Barnes, G. (2016). *Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions*. Obtenido de Wiley Online Library: *Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions*
- Berzai, L. (2017). *How Ethical Theories Apply to IT Professionals*. Obtenido de CompTIA: <https://www.comptia.org/blog/how-ethical-theories-apply-to-it-professionals#:~:text=These%20two%20theories%2C%20the%20teleological,may%20act%20morally%20regarding%20technology.>
- Bousabarah, K. (2021). *Automated detection and delineation of hepatocellular carcinoma on multiphase contrast-enhanced MRI using deep learning*. Obtenido de Public Medicine: <https://pubmed.ncbi.nlm.nih.gov/32500237/>
- Braidbach, C., Asadi, I., & Davern, M. (2019). *Ethical Issues in Big Data Analytics: A Stakeholder Perspective*. 44(34):718-747.
- Branchman, R. (2006). More than the sum of its parts. *AI Magazine*, 27(4):19-34.
- Champagne, M., Tonkens, & R. (2015). Bridging the Responsibility Gap in Automated Warfare. *Philosophy of Technology*, págs. 28:125–137.
- Chen, X., Liu, C., Li, B., Lu, K., & Sond, D. (2017). *Targeted backdoor attacks on deep learning*. Obtenido de CoRR, abs/1712.05526, 2017. U: <http://arxiv.org/abs/1712.05526>.
- Cheung, J., & Tomlin, N. (2020). Ep. 8 The Ethics of AI. Spotify.

- Coeckelbergh, M. (2021). *Ética de la inteligencia artificial*. Obtenido de Cátedra Teorema: https://www.catedra.com/primer_capitulo/etica-de-la-inteligencia-artificial.pdf
- Comission, E. (2020). *WHITE PAPER On AI-A A European approach to excellence and trust*. Bruselas.
- Copeland, B. (s.f.). *Artificial Intelligence*. Obtenido de Britannica: <https://www.britannica.com/technology/artificial-intelligence/Alan-Turing-and-the-beginning-of-AI>
- Dameski, A. (2020). *A Comprehensive Ethical Framework for AI Entities*. Obtenido de Research Gate: https://www.researchgate.net/profile/Andrej-Dameski/publication/327792821_A_Comprehensive_Ethical_Framework_for_AI_Entities_Foundations/links/5ba4c576a6fdccd3cb67d1fc/A-Comprehensive-Ethical-Framework-for-AI-Entities-Foundations.pdf
- Davis, R. (1996). What are intelligence? And why? . *AI Magazine*, 19(1):91-111.
- Desai, D., & Kroll, J. (2017). *Trust but verify: A guide to algorithms and the law*. Obtenido de Harvard Journal of Law and Technology.: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2959472.
- Doran, D. (2017). *What Does Explainable AI Really Mean?* Dayton, Ohio, EE.UU: Wright State University.
- Edrélyi, O. (2018). Regulating Artificial Intelligence: Proposal for a Global Solution. *ACM Digital Library*, 95-101. Obtenido de ACM Digital Library.
- Edwards, L. (2017). Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review*, págs. 40-46. Obtenido de https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855
- EFE. (2017). *Detienen a un hombre por violencia doméstica gracias a llamada de Amazon Alexa*. Obtenido de El Comercio: <https://www.elcomercio.com/guaifai/amazon-alexa-asistente-inteligenciaartificial-policia.html>
- EMERG. (s.f.). *Everyday Examples of Artificial Intelligence and Machine Learning*. Obtenido de EMERJ. The AI Research and Advisory Company: <https://emerj.com/ai-sector-overviews/everyday-examples-of-ai/>
- EMERJ. (s.f.). *What is Machine Learning?* Obtenido de EMERJ. The AI Research and Advisory Company: <https://emerj.com/ai-glossary-terms/what-is-machine-learning/>
- EuropeanComission. (2018). *Draft Ethics Huidelines for Trustworthy AI: Working Document for Stakeholders*. Bruselas: European Comission.
- Floridi, L. (2004). On the Morality of Artificial Agents. *Minds and Machines*.
- Forbes. (2019). *Industries Soon To Be Revolutionized By Artificial Intelligence*. Obtenido de Forbes Technology Council: <https://www.forbes.com/sites/forbestechcouncil/2019/01/16/13-industries-soon-to-be-revolutionized-by-artificial-intelligence/?sh=3c3640953dc1>
- Greene, D. (2019). Making Ethics in Machine Learning (Ethics of AI in Context). Spotify Podcast.
- Hibbard, B. (2015). *Ethical Artificial Intelligence*. Berkeley, CA: Machine Intelligence Research Institute.
- Iberdrola. (s.f.). Obtenido de Iberdrola.
- Ilsoe, A. (2018). *Progressing the voluntarist approach*. Obtenido de University of Copenhagen:

- https://curis.ku.dk/portal/files/198709895/Work_in_the_Digital_Age.pdf#page=143
- Jacobson, N., & Bentley, K. (2019). *Ethical dilemmas posed by mobile health and machine learning in psychiatry research*. Obtenido de World Health Organization: <https://www.who.int/bulletin/volumes/98/4/19-237107/en/>
- Kaplan. (2018). *CFA 2018 Level II. Schweser Notes*. Volume 1: Kaplan Schweser.
- Kelleher, J., & Tierney, B. (2018). *Data Science*. Obtenido de MIT Press: <https://books.google.es/books?hl=es&lr=&id=UlpVDwAAQBAJ&oi=fnd&pg=PP7&dq=kelleher+and+tierney&ots=vVj-Uhea4H&sig=Oe60NZ0NdWVLkvgjLtHRXTy5U7Y#v=onepage&q=kelleher%20and%20tierney&f=false>
- Kotsiantis, S. (2011). Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades. *Artificial Intelligence Review volume*, 37, 331–344.
- Krishnan, A. (2009). *Killer robots: legality and ethicality of autonomous weapons*. . Farnham: Ashgate.
- Kumar, R., O'Brien, D., Vijloen, S., & Albert, K. (2018). *Law and Adversarial Machine Learning*. Obtenido de Berkman Klein Center for Internet and Society: <https://arxiv.org/pdf/1810.10731.pdf>
- Lake, B., Ullman, T. D., Tenenbaum, J. B., & and Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40:E253.
- Lean, P. (2019). *The Extension of Legal Personhood in Artificial Intelligence*. Obtenido de Revista de Bioética y Derecho: http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1886-58872019000200004
- Lepri, B., Oliver, N., Letouze, E. F., Pentland, A. P., & Vinck, P. (2018). *Fair, Transparent, and Accountable Algorithmic Decision-making Processes*. Obtenido de MIT: <https://dspace.mit.edu/handle/1721.1/122933>
- Lipton, Z. (2016). *The myths of model interpretability*. Obtenido de Workshop on Human Interpretability in Machine Learning.
- Luciano Floridi, J. C. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. págs. pages689–707.
- Magoulas, G. (1999). *Machine Learning in Medical Applications*. Obtenido de Research Gate: https://www.researchgate.net/profile/George-Magoulas/publication/225171947_Machine_Learning_in_Medical_Applications/links/55f2ac0108ae51c380bffd1a/Machine-Learning-in-Medical-Applications.pdf
- Marcuse, H. (1991). *One-dimensional man*. Boston: Beacon. .
- Martin, K. (12 de 01 de 2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, págs. 835-850.
- Matsunaga, A. (2010). On the Use of Machine Learning to Predict the Time and Resources Consumed by Applications. *10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, Melbourne*, 495-504, doi: 10.1109/CCGRID.2010.98.
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology* , págs. 175–183.
- McKibben, B. (2003). *Enough: Staying Human in and Engineered Age*. Times Books.

- Mengnan, D. (2019). *Techniques for Interpretable Machine Learning*. Texas A&M University: Department of Computer Science and Engineering.
- Minseen, T. (2020). *AI, Machine Learning & Big Data*. Obtenido de Global Legal Insights: <https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/denmark>
- Minsky, M. (1985). *The Society of Mind*. Nueva York: Simon and Schuster.
- Mittlestadt, B., Allo, P., & Taddeo, M. (2016). *The ethics of algorithms: Mapping the debate*. Obtenido de SAGE journals: <https://journals.sagepub.com/doi/full/10.1177/2053951716679679>
- Ng, A. (s.f.). *Machine Learning*. Obtenido de Stanford: <https://www.coursera.org/learn/machine-learning>
- OMCE. (2021). *Cada día mueren en España 300 personas por cáncer, 110.000 al año*. Obtenido de Organización Médica Colegial de España: <http://www.medicosypacientes.com/articulo/cada-dia-mueren-en-espana-300-personas-por-cancer-110000-al-ano#:~:text=Cada%20d%C3%ADa%20mueren%20en%20Espa%C3%B1a%2000%20personas%20por%20c%C3%A1ncer%2C%20110.000%20al%20a%C3%B1o,-Cada%20d%C3%ADa%20mueren>
- OPAL. (s.f.). *About OPAL*. Obtenido de Opal Project: <https://www.opalproject.org/about-opal>
- Parliament, E. (2019). *EU guidelines on ethics*. Obtenido de EuroParl: [https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI\(2019\)640163_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/640163/EPRS_BRI(2019)640163_EN.pdf)
- RAE. (s.f.). *Inteligencia Artificial*. Obtenido de Real Academia Española: Real Academia Española
- Russell, S. (2019). *Human Compatible: AI and the Problem of Control*. Bristol: Allen Lane.
- Sáenz, M. (2018). *Colegiarse, ¿para qué?: El valor de los Colegios Oficiales profesionales*. Obtenido de Observatorio: [https://www.observatoriorh.com/opinion/colegiarse-para-que-el-valor-de-los-colegios-oficiales-profesionales.html#:~:text=Mediante%20la%20colegiaci%C3%B3n%20se%20da,exigentes%20y%20que%20est%C3%A1n%20contrastados.&text=Adem%C3%A1s%2C%20las%20ventajas%20no%](https://www.observatoriorh.com/opinion/colegiarse-para-que-el-valor-de-los-colegios-oficiales-profesionales.html#:~:text=Mediante%20la%20colegiaci%C3%B3n%20se%20da,exigentes%20y%20que%20est%C3%A1n%20contrastados.&text=Adem%C3%A1s%2C%20las%20ventajas%20no%20)
- Schank, R. (1987). What is AI, Anyway? *AI Magazine*, Volume 9 Number 4.
- Schwab, K. (2016). *The Fourth Industrial Revolution: what it means, how to respond*. Obtenido de World Economic Forum: <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>
- Shepperd, M. (s.f.). *Researcher Bias: The Use of Machine Learning*. Obtenido de Bura Brunel: <https://bura.brunel.ac.uk/bitstream/2438/8784/2/Fulltext.pdf>
- Shum, Y. (2020). *Situación Global Mobile 2020*. Obtenido de Yi Min Shum: <https://yiminshum.com/mobile-movil-app-2020/#:~:text=Lo%20m%C3%A1s%20interesante%20de%20esto,de%20telefon%C3%ADa%20similar%20o%20diferente>
- Vigna, F., Cimino, A., Petrocchi, M., & al, e. (2017). *Hate me, hate me not: Hate Speech Detection of Facebook*. Obtenido de Istituto di Informatica e Telematica, CNR, Pisa, Italy: <http://ceur-ws.org/Vol-1816/paper-09.pdf>
- Wachter, S., Floridi, L., & Mittelstadt, B. (2017). *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection*

- Regulation*. Obtenido de Oxford Academic:
<https://academic.oup.com/idpl/article/7/2/76/3860948>
- Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence*, 20-24.
- Zhu, S.-L., Dong, J., Zhang, C., Huang, Y.-B., & Pan, W. (2020). *Application of machine learning in the diagnosis of gastric cancer based on noninvasive characteristics*. Obtenido de Plos One:
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0244869>