



Facultad de Ciencias Humanas y Sociales  
Grado en Relaciones Internacionales

Trabajo Fin de Grado

# La Ética de la Inteligencia Artificial

Estudiante: José Fernández Piera

Director: Javier Fuertes Pérez

Madrid, Junio 2020

**Resumen:** En un mundo ya totalmente dependiente de los ordenadores y en el que los avances tecnológicos y computacionales aumentan de manera exponencial a través, de entre otros avances, de la Inteligencia Artificial, asalta la duda y la preocupación de dónde están los límites de este nuevo tipo de tecnología. La creación de máquinas capaces de aprender y tomar decisiones por sí mismas y la implantación de nuevas tecnologías en nuestro día a día, acarrearán numerosos dilemas éticos.

A lo largo de esta última década, hemos podido ver muchos ámbitos en los que se ha empezado a implementar la IA, mostrando las potenciales ventajas que esta implementación podría traer, pero también nos ha mostrado el lado más oscuro de esta nueva tecnología. Así, hemos visto como la IA trae consigo infinitas posibilidades, pudiendo traducirse en una eficiencia económica o un desarrollo en la investigación científica sin precedentes, gracias a la capacidad de procesamiento de grandes conjuntos de datos y de la capacidad de ir “aprendiendo” de la experiencia propia, pero también puede traer consigo un incremento en las desigualdades sociales, un mayor control social, entre otros, haciendo que las perspectivas de futuro no sean tan esperanzadoras. La IA puede desarrollarse para potenciar tanto los aspectos positivos como los negativos del ser humano, por lo que está en nuestra mano decidir qué camino tomará.

Este trabajo surge de la necesidad de empezar a plantearse cuáles son los aspectos éticos que debe de seguir el desarrollo de esta tecnología, ya que, teniendo en cuenta la velocidad a la que progresa, puede que en menos tiempo del que prevemos, ya sea demasiado tarde.

**Palabras clave:** Inteligencia Artificial, Ordenador, Ética, Consciencia, Economía, Principios, Democracia, Sociedad, Relaciones Internacionales.

**Abstract:** In a world that is already totally dependent on computers and in which technological and computational advances are increasing exponentially through, among other advances, Artificial Intelligence, doubt and concern arise as to where the limits of this new type of technology are. The creation of machines capable of learning and making decisions on their own and the implementation of new technologies in our daily lives entail numerous ethical dilemmas.

Throughout this last decade, we have been able to see many areas in which AI has begun to be implemented, showing the potential advantages that this implementation could bring, but it has also shown us the darker side of this new technology. Thus, we have seen how AI brings with it infinite possibilities, which can be translated into unprecedented economic efficiency or development in scientific research, thanks to the capacity to process large data sets and the ability to "learn" from one's own experience, but it can also bring with it an increase in social inequalities, greater social control, among others, making future perspectives less hopeful. AI can be developed to enhance both the positive and negative aspects of the human being, so it is in our hands to decide which path to take.

This work arises from the need to start thinking about the ethical aspects that the development of this technology should follow, since, taking into account the speed at which it progresses, it may be that in less time than we foresee, it may be too late.

**Keywords:** Artificial Intelligence, Computer, Ethics, Conscience, Economy, Principles, Democracy, Society, International Relations.

## ÍNDICE

<b>1. INTRODUCCIÓN.....</b>	<b>5</b>
<b>1.1. OBJETIVOS.....</b>	<b>5</b>
<b>1.2. METODOLOGÍA .....</b>	<b>5</b>
<b>2. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL? CARACTERÍSTICAS Y SUS DELIMITACIONES. ¿HABLAMOS DE SERES CONSCIENTES? .....</b>	<b>6</b>
<b>3. EL AVANCE Y DESARROLLO DE LA IA Y SUS CONSECUENCIAS .....</b>	<b>15</b>
<b>a. Una nueva globalización .....</b>	<b>15</b>
<b>b. Una nueva democracia.....</b>	<b>23</b>
<b>c. ¿Un nuevo futuro?.....</b>	<b>29</b>
<b>4. ÉTICA DE LA INTELIGENCIA ARTIFICIAL .....</b>	<b>40</b>
<b>a. La roboética .....</b>	<b>41</b>
<b>b. Principios éticos para una digitalización humanista.....</b>	<b>48</b>
<b>5. CONCLUSIONES.....</b>	<b>54</b>
<b>6. BIBLIOGRAFÍA.....</b>	<b>58</b>

*“Hasta la fecha, no se ha diseñado un ordenador que sea consciente de lo que está haciendo; pero, la mayor parte del tiempo, nosotros tampoco lo somos” – Marvin Minsky*

## **1. INTRODUCCIÓN**

### **1.1. OBJETIVOS**

El objetivo de este trabajo es realizar un análisis acerca del desarrollo de la Inteligencia Artificial, identificando los dilemas éticos derivados de éste, para poder hacer una aproximación final de cuáles son los principios que habrían de implementarse para su correcto desarrollo. Para ello, se realizará un estudio de la Inteligencia Artificial, definiendo el concepto y sus características esenciales, para después describir y comparar los distintos tipos. Más adelante se identificarán las diferentes aplicaciones de la Inteligencia Artificial en aspectos económicos, políticos y sociales, analizando sus consecuencias y los dilemas éticos que de ellas derivan. Por último se hará un análisis de los avances legislativos en el ámbito de la ética de la Inteligencia Artificial y cuáles son los principios éticos que deberían servir como base para evitar los posibles efectos negativos de la implementación de la Inteligencia Artificial.

### **1.2. METODOLOGÍA**

Este trabajo se ha llevado a cabo mediante una investigación bibliográfica a través de la cual se ha realizado una recopilación de información para proceder a su análisis. Mediante fuentes de información primarias y secundarias se ha podido investigar y analizar los distintos artículos científicos, éticos, económicos y sociales sobre el tema para así poder tomar y estudiar los datos más relevantes. Del mismo modo se ha accedido a informes y resoluciones de instituciones europeas, como el Parlamento Europeo o la Comisión, así como del gobierno de los Estados Unidos.

Esta bibliografía se ha obtenido a través de herramientas de búsqueda en internet y de plataformas de contenido académico como Google Scholar o la biblioteca de la Universidad Pontificia de Comillas. Del mismo modo, dado que el tema del trabajo es de plena actualidad, se han podido encontrar numerosas fuentes de carácter divulgativo, como revistas o periódicos, lo que ha permitido completar ciertas partes del trabajo a través de un enfoque deductivo.

## **2. ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL? CARACTERÍSTICAS Y SUS DELIMITACIONES. ¿HABLAMOS DE SERES CONSCIENTES?**

Uno de los principales problemas que nos encontramos al acercarnos al campo de estudio de la inteligencia artificial surge de la dificultad de definir el concepto. Por ello, antes de comenzar con el análisis en profundidad de los aspectos y dilemas éticos derivados del desarrollo e implantación de la inteligencia artificial en nuestra vida diaria, debemos empezar definiendo qué es realmente la inteligencia artificial. Para ello, se ha de exponer cuáles son las características que determinan la ‘cualidad’ de inteligencia artificial, pudiendo de esta forma hacer una delimitación del término de forma negativa, es decir, que aquellos dispositivos que no reúnan dichas características, no puedan calificarse como Inteligencia Artificial y se califiquen simplemente como sistemas operativos que sigan las pautas establecidas por el programador.

El concepto de ‘inteligencia artificial’ nace en 1956 de la mano del informático John McCarthy, cuando invitó a un grupo de investigadores de una variedad de disciplinas a una conferencia llamada el Proyecto de Investigación de Verano de Dartmouth sobre Inteligencia Artificial, para discutir lo que finalmente se convertiría en el campo de la inteligencia artificial. En aquel momento, los investigadores se reunieron para aclarar y desarrollar los conceptos en torno a las “máquinas pensantes” que hasta ese momento habían sido bastante divergentes. El propósito de dicha conferencia fue que *“el estudio debe proceder sobre la base de la conjetura de que cada aspecto del aprendizaje o cualquier otra característica de la inteligencia puede, en principio, describirse con tanta precisión que se puede hacer una máquina para simularlo”* (Marr, 2018).

Pero, ¿qué es la Inteligencia Artificial? Para responder esta pregunta parece razonable comenzar por el Diccionario. El Diccionario de la Lengua Española define ‘inteligencia artificial’ 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”* (Diccionario de la Lengua Española, 2014). El propio McCarthy, según un artículo divulgado por la Universidad de Stanford (donde impartió clase desde 1962 hasta el año 2000) en 2007, entendía la Inteligencia artificial

como “*la ciencia y la ingeniería de crear máquinas inteligentes, especialmente programas de computación inteligentes. Está relacionada con la tarea similar de utilizar ordenadores para comprender la inteligencia humana, pero la IA no se limita a métodos que sean observables biológicamente*” (Alandete, 2011).

Una vez presentado el concepto de Inteligencia artificial (en adelante IA), es necesario ahondar en cuáles son las características que han de reunirse para poder ser considerada como tal y no como un simple sistema operativo que sigue las pautas marcadas por el programador. Por ello, parece obvio que las características que han de reunirse son dos: que el sistema sea ‘inteligente’ y que dicha inteligencia tenga un origen ‘artificial’.

En primer lugar, lo que se supone que es “artificial” en la IA, sin duda, tiene relación con sus orígenes y su modo de creación, ya que surge como producto de la invención e ingenio humanos más que como resultado de la influencia natural (especialmente biológica o evolutiva). En otras palabras, las cosas que son artificialmente inteligentes son distintas de las que son naturalmente inteligentes en tanto en cuanto, son artefactos que poseen propiedades especiales que normalmente son características de entes que las adquirieron a través de un proceso natural. Así pues, se trata de cosas que tienen una cierta propiedad (inteligencia) como resultado de un cierto proceso de desarrollo, diseño o fabricación (Fetzer, 1990).

Por ello, la duda surge cuando nos planteamos qué es la “inteligencia”. El Diccionario de la Lengua Española expone varias definiciones de inteligencia, de las que podemos destacar: “*1. Capacidad de entender o comprender; 2. Capacidad de resolver problemas; 3. Conocimiento, comprensión, acto de comprender*” (Diccionario de la Lengua Española, 2014). En base a que la mencionada inteligencia es una característica intrínseca en el ser humano, el problema surge al tratar de aislar aquellos aspectos específicos de la existencia humana que se supone que son “inteligentes”. Dado que los seres humanos exhiben, por ejemplo, ira, celos o rabia, podríamos preguntarnos si un dispositivo sería inteligente en el caso de poder exteriorizar dichos comportamientos. Si los seres humanos son los mejores ejemplos de dicha propiedad, y sin embargo se observa ampliamente que exhiben este tipo de sentimientos, y teniendo en cuenta que hay situaciones, circunstancias y condiciones en las que mostrarlos sería una conducta inteligente, ¿con

qué criterio podemos decir que estos no son algunos de los rasgos característicos de la inteligencia? (Fetzer, 1990).

Dado que la epistemología (como acabamos de observar) es una materia muy compleja y amplia en sí misma y, dado que no es el tema de estudio de este trabajo, no ahondaremos más en cual es la delimitación del concepto de inteligencia. Entonces, basándonos en la definición de Bernard Marr, en este trabajo entenderemos la “inteligencia” en términos de la IA como una simulación de la capacidad de pensamiento abstracto, creativo y deductivo (y particularmente la habilidad de aprender) propio de los humanos, usando la lógica digital y binaria de los ordenadores.

- **¿Estamos ante un ser consciente?**

El debate de si las máquinas pueden tener conciencia no es nuevo, ya que es una cuestión con un gran bagaje a sus espaldas, habiendo dos posturas respecto al tema, una defendida por la “IA fuerte” y otra la “IA específica”. Es por ello que han de exponerse ambas posturas para, saber qué es lo que defiende cada una y por otro lado, para saber cuál es el ‘tipo’ de IA que se ha conseguido desarrollar hasta la fecha.

En primer lugar, la “IA fuerte” (también denominada general) es un término utilizado para describir una cierta perspectiva de desarrollo de la inteligencia artificial. El objetivo de la IA fuerte es desarrollar la inteligencia artificial hasta el punto en que la capacidad intelectual de la máquina sea funcionalmente igual a la de un humano. El ordenador ideal de IA fuerte se construiría en forma de ser humano, tendría la misma percepción sensorial que un humano, pasando por los mismos procesos de educación y aprendizaje que un niño humano. Esencialmente, la máquina “nacería” como un niño y eventualmente se desarrollaría hasta convertirse en un adulto de manera análoga al desarrollo humano (University of California, Berkeley, s.f.).

Este enfoque presenta una solución a los problemas de los intentos simbólicos de crear inteligencia humana en sistemas computarizados. En lugar de intentar dar al ordenador un conocimiento similar al de un adulto desde el principio, sólo habría que darle la capacidad de interactuar con el entorno y la habilidad de aprender de esas interacciones. A medida que transcurriera el tiempo ganaría sentido común y lenguaje por sí mismo.

Este paradigma busca combinar la mente y el cuerpo, mientras que la tendencia común en la programación ha sido la de ignorar el cuerpo en detrimento del intelecto del ordenador.

En la actualidad, el desarrollo de la IA fuerte se ve relegado a las aplicaciones inmediatas de la IA aplicada, ya que ese campo produce resultados más tangibles y cuantificables. Además, el desarrollo de sistemas con IA fuerte no ha avanzado más allá de construcciones básicas, las cuales no alcanzan siquiera los niveles de inteligencia más básicos. Sin embargo, quizás el mayor obstáculo con el que se encuentra la IA fuerte es la dificultad de definición de inteligencia. El objetivo final de la IA fuerte es hacer un ordenador inteligente que pueda pensar y comprender, pero el término de inteligencia, como ya hemos expuesto, siguen siendo ambiguo e indefinible; por lo tanto, no hay una medida determinada de “éxito” en el campo de la IA fuerte.

Por ello, podemos reducir el objetivo de la IA fuerte en tres aspectos: 1) la capacidad de generalizar los conocimientos de distintos ámbitos y aplicar dichos conocimientos adquiridos en ámbitos distintos; 2) la capacidad de hacer planes futuros basados en conocimientos y experiencias; y 3) la capacidad de adaptarse al entorno a medida que se producen los cambios. Además, hay aspectos auxiliares que deben acompañar a estos requisitos principales, como la capacidad de razonar, resolver problemas, la capacidad de incorporar conocimiento y sentido común y la capacidad de planificar (Walch, 2019).

Algunos sectores críticos han cuestionado que la definición anterior de IA fuerte no es suficiente para ser clasificada como verdaderamente inteligente porque el simple hecho de poder realizar tareas y comunicarse como un humano no es realmente una IA fuerte. Reforzando la opinión de estos detractores están los casos de sistemas en los que los humanos son incapaces de distinguir entre un humano y una máquina, de forma muy parecida a la versión física de una prueba de Turing. El test de Turing tiene como objetivo determinar si una máquina es inteligente o no, y consiste en poner a un ser humano, un ordenador y un interrogador en un entorno de conversación, en la que el interrogador se ubicará en una habitación distinta a la del ser humano y el ordenador. El interrogador realizará unas preguntas y las responderán tanto el ordenador como el ser humano, y en el caso de que el interrogador no pueda diferenciar entre el ser humano y el ordenador, se considera que este último puede determinarse como inteligente. Hoy en día, algunos

*chatbots* muy avanzados (como por ejemplo Google Duplex) parecen pasar la Prueba de Turing. Entonces ¿puede considerarse a Google Duplex como un ente en sí mismo inteligente?

Hay varias diferencias entre los “*chatbots*” y un verdadero ejemplo de IA fuerte. Los “*chatbots*” son *bots*<sup>1</sup> especializados y creados para mantener conversaciones y ofrecer respuestas preconcebidas (ESIC, 2018), y se construyen con el único propósito de conversar, utilizando algoritmos<sup>2</sup> y bases de datos de respuesta cada vez más complejos para determinar una respuesta apropiada a una pregunta o declaración. Por ello, los *chatbots* son un ejemplo de IA aplicada, en la que los sistemas se desarrollan y diseñan específicamente para una función concreta.

Pero, no se puede negar que la IA sí que cada día se está volviendo más “fuerte”. Si la IA se está fortaleciendo, es gracias a organizaciones como la ya mencionada Google, u OpenAI, un “*think tank*” fundado por Elon Musk. Podemos afirmar que IA se está fortaleciendo, al menos en el sentido de que es capaz de producir predicciones cada vez más precisas sobre los datos que recibe. Por ejemplo, el progreso en la última década en lo relativo a la visión por ordenador en la última década, se acerca al 100% de precisión reconociendo correctamente los objetos en las imágenes, lo cual es un indicador de una IA cada vez más fuerte. Otro indicador es la capacidad de los algoritmos de DeepMind<sup>3</sup> para superar más videojuegos y para transferir el aprendizaje de un juego a otro. La capacidad de la *GPT-2*<sup>4</sup> de OpenAI para resolver problemas para los que no fue entrenada es otra señal. Pero, a pesar de estos indicadores, aún no hemos llegado al punto de haber

---

<sup>1</sup> Los *bots* son softwares creados bajo el paraguas digital de la inteligencia artificial y sirven para automatizar procesos que se ejecutan sin la necesidad de una intervención humana (ESIC, 2018).

<sup>2</sup> “Conjunto de reglas que, aplicada sistemáticamente a unos datos de entrada apropiados, resuelven un problema en un número finito de pasos elementales” (Peña Marí, 2006).

<sup>3</sup> *DeepMind* es una compañía de IA adquirida por Google, la cual desarrolló una “red neuronal” que aprende a jugar a videojuegos de forma equivalente a la de los seres humanos. Uno de sus programas aprendió a jugar 49 juegos diferentes de Atari desde cero sólo con ver los píxeles y la puntuación en la pantalla, y su programa AlphaGo fue el primero en vencer a un jugador profesional (DeepMind, s.f.).

<sup>4</sup> *GPT-2* es un generador de texto de inteligencia artificial, que basado en una palabra, frase o texto inicial continúa la historia con un estilo y contexto coherente (Pérez, 2019).

conseguido desarrollar una AI fuerte como tal, ya que dichas características por separado, son muestras de IA específica, la cual se expondrá a continuación.

La discusión sobre si seremos capaces de crear una IA fuerte, es a día de hoy fundamentalmente teórica. Por ello, se puede afirmar que la IA fuerte es una visión que a día de hoy queda relegada a la ciencia ficción, ya que no se ha desarrollado ni siquiera de manera aproximada, ningún sistema que pueda reunir los aspectos necesarios para poder considerarlo tan inteligente o más que un ser humano.

La otra aproximación respecto a la IA, es la denominada como “IA débil” (a la que nos referiremos como específica), la cual describe un tipo de IA que está programada para realizar una tarea concreta, ya sea, por ejemplo, reconocer imágenes, jugar al ajedrez o analizar datos en bruto para realizar informes de cualquier índole. Este tipo de IA es la que más se ha desarrollado a día de hoy, y básicamente todos los dispositivos electrónicos con los que interactuamos en nuestro día a día son IA específica, como por ejemplo Siri, Alexa o el traductor de Google, entre muchos otros.

Los sistemas de la IA específica pueden encargarse una tarea en tiempo real, pero extraen información de un conjunto de datos específico. De este modo, estos sistemas no desempeñan ninguna tarea distinta de la única para la que fueron diseñados. La IA específica no es consciente, no es sensible y no se mueve por emociones como hacemos los humanos o como pretende conseguir la aproximación de la IA fuerte, sino que opera dentro de un margen de operaciones predeterminado y predefinido, a pesar de que aparente ser mucho más sofisticada. Algunos podrían asumir que estas herramientas no son “débiles” por su capacidad de interactuar con nosotros y procesar el lenguaje humano, pero la razón por la que se denomina IA “débil” es porque estas máquinas no están cerca de tener una inteligencia similar a la humana, ya que carecen de la conciencia y la inteligencia genuinas para igualar la inteligencia humana. Este tipo de inteligencia presenta la “limitación” de ceñirse únicamente a la función para la que se creó, por lo que, a pesar de realizar dichas tareas de forma mucho más eficiente que un ser humano, en su conjunto, no supera a la inteligencia humana. En otras palabras, se considera que la IA específica es menos inteligente que un ser humano en tanto en cuanto se limita al poder de la organización, más que a la capacidad de comprender o entender la orden que se da, ya que no son conscientes y no pueden pensar por sí mismos (Jajal, 2018).

Desde que en 1997, Deep Blue, el programa de ajedrez desarrollado por IBM, ganara en una partida al legendario Garry Kasparov, el ser humano se dio cuenta del enorme poder y potencial de los ordenadores. Por primera vez, una máquina había superado a un humano. Desde entonces, la IA específica nos ha mostrado que es capaz de superar (con creces) a los humanos en numerosas tareas que anteriormente estaban reservadas a estos últimos. Por ello, aunque la IA específica se denomine “débil”, no implica que lo sea, ya que los sistemas de la IA específica son capaces de procesar datos y completar tareas a un ritmo significativamente más rápido que cualquier ser humano, lo que nos ha permitido mejorar nuestra productividad, eficiencia y calidad de vida en general. Estos sistemas con IA específica, como *Watson*<sup>5</sup> de IBM, por ejemplo, es capaz de aprovechar el poder de la IA para ayudar a los médicos a tomar decisiones basadas en datos, haciendo que la atención sanitaria sea mejor, más rápida y más segura.

Por ejemplo, la IA específica ya se usa en el sector bancario, ya que con una gran capacidad de procesamiento de datos, proporciona una evaluación más rápida, económica y precisa de un posible prestatario, y tiene en cuenta una mayor variedad de factores, lo que conduce a una decisión mejor informada y respaldada por datos. Los bancos digitales y las aplicaciones de emisión de préstamos utilizan algoritmos de aprendizaje automático para utilizar datos que ayuden a evaluar la elegibilidad de los préstamos y ofrecer opciones personalizadas. La puntuación de créditos proporcionada por la Inteligencia Artificial se basa en reglas más complejas y sofisticadas que las utilizadas en los sistemas tradicionales de puntuación de créditos que puede manejar un humano (Bachinskiy, 2019). Por ejemplo, ayuda a los prestamistas a distinguir entre los prestatarios con alto riesgo de impago y los que son solventes pero carecen de un historial de crédito extenso. Además, a diferencia de un ser humano, no es probable que una máquina sea parcial, por lo que la objetividad es otro beneficio del mecanismo impulsado por la IA.

---

<sup>5</sup> IBM Watson Health fue creado para ayudar a resolver algunos de los desafíos de salud más urgentes del mundo a través de datos, análisis e inteligencia artificial. Al combinar expertos humanos con inteligencia aumentada, IBM Watson Health ayuda a los profesionales de la salud e investigadores de todo el mundo a traducir los datos y el conocimiento en percepciones para tomar decisiones más informadas sobre el cuidado de la salud en cientos de hospitales y organizaciones de salud, y un gran y creciente cuerpo de evidencias apoya el uso de Watson en la atención médica (IBM, s.f.).

De la misma forma, en la industria turística, este tipo de IA ayuda a optimizar las ventas y los precios, así como a prevenir las transacciones fraudulentas. Además, permite ofrecer sugerencias personalizadas sobre las fechas, rutas y costos deseados, cuando estamos navegando en sitios de reservación de aviones u hoteles planeando nuestras próximas vacaciones de verano.

Pero, hay que hacer hincapié en que, a pesar de realizar estas tareas de una forma que un humano jamás sería capaz de igualar debido a la apabullante cantidad de datos que emplea, la IA específica se define muy bien por su propio nombre, ya que este tipo de IA sólo realizará esa tarea específica para la que está programada. Por ello, a pesar de que a través del aprendizaje automático podrá aprender de su propia experiencia en el proceso de datos, lo cual le permitirá desarrollar sus capacidades para actuar de una manera más rápida y eficiente, ese conocimiento y esa mejora sólo se aplicarán hacia dichas tareas específicas.

Además, la IA específica nos ha liberado de muchas de las aburridas y rutinarias tareas que no queremos hacer. Desde el aumento de la eficiencia en nuestras vidas personales, como pedirle a Siri que nos pida una pizza, hasta la búsqueda entre montones de datos y su análisis para producir resultados, la IA específica ha hecho que nuestras vidas sean significativamente mejores, por lo que no debemos subestimarla. Con el advenimiento de tecnologías avanzadas como los coches de auto-conducción, los sistemas de la IA también nos aliviarán de realidades frustrantes como estar atrapados en el tráfico, y en su lugar nos proporcionarán más tiempo de ocio (Jajal, 2018).

Así, a pesar de contar con multitud de ejemplos de IA en nuestra vida diaria, todos los ejemplos con los que contamos se limitan a una IA específica, por lo que, tal y como se acaba de exponer, la conciencia no es (ni va a ser, al menos en el corto y medio plazo) una característica de la que dispongan los dispositivos y sistemas con IA. Por ello, parece interesante reflexionar sobre las limitaciones que presentan estos sistemas con IA frente a las capacidades del ser humano.

La falta de conciencia en los sistemas con IA limita mucho la capacidad de los mismos, ya que no realizan ninguna acción por sí mismos de manera voluntaria, ya que para su funcionamiento necesitan en primera instancia haber sido programados y después

necesitan recibir una primera orden por parte del humano que quiere disponer de sus ‘servicios’. Pongamos como ejemplo de estas limitaciones a los famosos vehículos autónomos.

No hablamos de vehículos ‘autónomos’ propiamente dichos, porque, a pesar de realizar el desplazamiento al destino seleccionado sin necesidad de la intervención humana, no podemos hablar de que dicho coche haya decidido por sus propios medios ni el destino, ni el momento de realizar el desplazamiento, ni siquiera la ruta, ya que se decide por un algoritmo que selecciona cuál es la ruta más indicada para ahorrar el máximo de tiempo posible dependiendo de distintos factores. Con esto no se pretende menospreciar o quitarle importancia a dichos vehículos, ya que ofrecen la oportunidad de aprovechar los beneficios de las últimas tecnologías sensoriales y de la IA para tomar decisiones de conducción que mitiguen muchos riesgos asociados a las decisiones de conducción humanas. Dicha combinación entre las tecnologías sensoriales y la IA proporciona una representación topográfica de las carreteras que permite a la IA tomar decisiones de conducción más inmediatas y precisas. Como máquinas, la IA también elimina los riesgos asociados a las fragilidades humanas, como pueden ser la fatiga, la percepción errónea del entorno o las intoxicaciones, además de las decisiones erróneas que los humanos pueden tomar en el contexto de la conducción (Cunneen, Mullins & Murphy, 2019).

Para hablar de vehículos realmente autónomos, estos deberían contar con la capacidad de tomar sus propias decisiones, pero no sólo durante el trayecto entre destinos (cosa que ya realizan), sino que deberían ser capaces de contar con la voluntad de tomar o dejar de tomar ciertas decisiones sin intervención humana. Es decir, no podemos hablar de una ‘autonomía’ del vehículo, ya que no toma ninguna decisión por sí mismo, no hace un ejercicio de reflexión de la razón por la cual realiza sus actos, ni tampoco piensa en si le apetece o no realizar una tarea, simplemente la hace, porque está programado para que si se le solicita, la realizará de la mejor forma posible. Un vehículo autónomo no sale a dar un paseo porque está aburrido, ni se niega a ir a hacer la compra porque llueve, como tampoco se plantea el por qué ha decidido ir por una ruta en vez de por otra, simplemente realiza las cosas en base a lo que le ordenan las directrices marcadas por los algoritmos desarrollados por el ser humano que lo programó. Por ello, sin esa capacidad de decidir por uno mismo, no creo que podamos hablar de sistemas inteligentes que puedan siquiera

asemejarse al ser humano. La falta de conciencia de uno mismo limita de manera total la libertad y la autonomía del sistema con IA, ya que desaparece el libre albedrío y con él desaparece cualquier posible semejanza o equiparación real que pudiera haber entre una máquina y un ser humano.

Es fácil perderse en la filosofía pero debemos tener en cuenta cómo la IA está viviendo una época de ‘madurez’, aumentando las capacidades de la IA y cómo esa madurez puede aplicarse para satisfacer nuevas necesidades. Dada la ambigüedad que encontramos en los términos expuestos, no es posible adentrarnos en la metafísica, realizando un desarrollo pormenorizado de cada uno de ellos, ya que nos ocuparía otro trabajo entero. Por ello, deberíamos más bien mirar las capacidades de los sistemas de IA pueden hacer y tratar de implementarlas en la medida de lo posible en el espectro de necesidades humanas, pero siempre teniendo en cuenta los límites actuales de las tecnologías de IA frente a los seres humanos. De este modo, el objetivo de este trabajo es que se enfoque desde el prisma de las Relaciones Internacionales, viendo cuáles son las aplicaciones de la IA (tanto las ya implementadas como las que aún están en desarrollo) en las sociedades contemporáneas, y cuáles serán los efectos económicos, políticos y sociales que estas acarrearán.

### **3. EL AVANCE Y DESARROLLO DE LA IA Y SUS CONSECUENCIAS**

El desarrollo de tecnologías de IA ha vivido un crecimiento exponencial en la última década, y no parece que vaya a reducirse en las venideras, por lo que deberíamos ser conscientes de los cambios que están por venir y tomar conciencia para intentar prever las consecuencias que estos puedan tener en todos los aspectos de la sociedad. Por ello, en esta sección del trabajo se hará un análisis sobre los cambios que plantea la IA para el futuro y sus posibles consecuencias, haciendo distinción entre el plano económico, el político y el social.

#### **a. Una nueva globalización**

El progreso en muchas áreas ha acelerado el desarrollo de la IA, la cual, como ya se ha visto a lo largo de esta última década, tiene el potencial de remodelar el panorama

competitivo de las empresas, los empleos y el desarrollo económico de los países. En los últimos años, hemos sido testigos del gran avance en el procesamiento de lenguaje o en la visión artificial, entre otros, y este avance parece imparable. Además, muchos productos y servicios emplean ampliamente avances en la IA, como los asistentes personales y los sistemas de reconocimiento facial.

Predecir el impacto económico de la IA o de cualquier otra tecnología es un ejercicio especulativo, ya que no se puede prever con seguridad cuál va a ser su futuro en un mundo en constante cambio. El alcance y el ritmo del despliegue de la automatización dependen de varias variables entre las que podemos destacar la viabilidad técnica, el costo de desarrollar y desplegar tecnologías para usos específicos en el lugar de trabajo, la dinámica del mercado laboral, incluida la calidad y cantidad de la mano de obra y los salarios asociados, los beneficios de la automatización más allá de la sustitución de la mano de obra, y la aceptación normativa y social.

Lo que sí que está claro es que nuevos y amplios mercados pueden desarrollarse a un ritmo extremadamente rápido, ya que la integración de la Inteligencia Artificial en el comercio electrónico está aumentando los estándares de comercialización de muchas empresas. Ha ayudado a crear una experiencia más personalizada al realizar el análisis de grandes conjuntos de datos, y al identificar diferentes patrones de comportamiento del consumidor. Por ejemplo, analicemos el caso de China. A día de hoy, China es el líder mundial en el comercio electrónico, pero hace sólo diez años, China representaba 1 por ciento de las transacciones mundiales de comercio electrónico. En el año 2019, según datos de McKinsey, la venta online al por menor aumentó en 1,5 billones de dólares, lo que representa un cuarto del volumen total de ventas al por menor de China, y representa un volumen de operaciones superior a las ventas al por menor de los diez siguientes mercados más grandes del mundo juntos (McKinsey, 2019).

Los datos son la información que entra en una predicción. Muchos avances recientes en la IA dependen de grandes cantidades de datos digitales para que los sistemas de IA puedan hacer predicciones basadas en ejemplos anteriores. En general, cuantos más ejemplos pasados, más precisas son las predicciones. Por lo tanto, el acceso a grandes cantidades de datos es un activo de gran valor para las organizaciones y empresas debido a su potencial con la IA. El valor estratégico de los datos es relativo, ya que depende de

si los datos son útiles para predecir algo importante para una organización o una empresa, por ello, la capacidad de seguir aprendiendo a través de nuevos datos puede generar una gran ventaja competitiva.

Este conjunto masivo de datos es lo que se conoce comúnmente como “Big Data”. El término Big Data se refiere a la masa heterogénea de datos digitales producidos por empresas y personas cuyas características (gran volumen, formas diferentes, velocidad de procesamiento) requieren herramientas específicas y cada vez más sofisticadas de almacenamiento y análisis informático.

Existe una relación recíproca entre Big Data y la IA, ya que esta última depende en gran medida de la primera para tener éxito, al tiempo que ayuda a las empresas a aprovechar el potencial de sus almacenes de datos de maneras que antes eran muy complejas o imposibles. Esa dependencia nace porque la IA necesita datos para construir su inteligencia, particularmente en cuanto al aprendizaje de la máquina. Una aplicación de reconocimiento de imágenes para el aprendizaje automático, por ejemplo, mira miles de imágenes de un avión para aprender lo que constituye un avión y así poder reconocerlas en el futuro.

Aunque el concepto no es nada nuevo, ya que las empresas solían analizar manualmente sus datos con anterioridad, con los avances a la hora de utilizar Big Data a través de la IA, hay múltiples beneficios como el ahorro de tiempo y costes, la eficiencia, estrategias empresariales más inteligentes, operaciones más eficientes, mayores beneficios y clientes más satisfechos. Utilizando estos datos, las empresas de comercio electrónico pueden aprovechar al máximo sus tendencias pasadas, su rendimiento y su crecimiento para mejorar la satisfacción del cliente, aumentar los ingresos, conseguir nuevos clientes y racionalizar sus operaciones.

La minería de datos<sup>6</sup> con Big Data ayuda a entender a un comprador muy de cerca. El comprador normalmente no suele darse cuenta de que cada clic suyo está siendo

---

<sup>6</sup> La minería de datos es un proceso utilizado por las empresas para convertir los datos en bruto en información útil mediante el uso de programas informáticos empleados para buscar patrones en grandes lotes de datos. (Twin, 2019).

monitoreado continuamente, pero desde que entra hasta que sale, todo lo que explora en la web está quedando registrado. Basándose en eso, el vendedor identifica su interés de navegación y su historial de compras y obtiene una idea de su patrón de compras en cuanto a cuál es su demanda y cuándo compra más. Con el Big Data, los vendedores de comercio electrónico se han beneficiado en la gestión de su inventario también, ya que son capaces de organizar el número de clientes en base a su patrón de compra, detalles demográficos y período de tiempo y así deshacerse de su exceso de stock en consecuencia (Vaghela, 2019). Por ejemplo, por la noche o durante épocas festivas como la Navidad, se producen el mayor número de compras, por lo que algunos propietarios de tiendas online alteran el precio de sus productos/servicios durante ese período de tiempo o publican anuncios en las redes sociales para atraer a los clientes. Además, el Big Data ha facilitado la captación de clientes al recomendar artículos que podrían ser de su interés tras realizar un estudio del patrón de un comprador, por lo que el comercio electrónico está produciendo un mayor nivel de satisfacción del cliente con una experiencia óptima del mismo.

Un ejemplo claro del empleo de estas prácticas y de cómo hacen crecer a una empresa es Amazon. Amazon es el gigante del comercio electrónico a nivel mundial por excelencia, el cual ha prosperado, entre otras medidas, utilizando grandes cantidades de datos recogidos de los clientes mientras navegan para construir y ajustar su motor de recomendación. Cuanto más sepa Amazon sobre ti, mejor podrá predecir lo que quieres comprar y, una vez que sabe lo que puedes querer, agiliza el proceso persuadiéndote para que lo compres, por ejemplo, recomendando varios productos en lugar de hacerte buscar en todo el catálogo. De este modo, queda patente que el papel de la IA en el motor de recomendación de Amazon es enorme, ya que genera el 35% de los ingresos de la empresa (Skuzza, 2020)

La tecnología de recomendación de Amazon se basa en un “filtrado colaborativo”, lo que significa que decide lo que cree que quieres construyendo una imagen de quién eres, y luego ofreciéndote productos que personas con perfiles similares han comprado. La enorme cantidad de datos formados o reunidos se utiliza para organizar una “visión de 360 grados” de un cliente individual. Amazon recopila datos de cada uno de sus clientes mientras utilizan el sitio, ya que, además de registrar lo que compras, la compañía monitorea lo que miras, tu dirección de envío (ya que puede hacer una sorprendente

suposición de tu nivel de ingresos en base a donde vives), y si dejas reseñas o comentarios (Marr, 2019).

Pero Amazon no se limita al uso de la IA en sólo en este aspecto. La empresa elige utilizar la palabra “*flywheel*” para ilustrar cómo las diferentes partes de su extenso negocio funcionan como una única máquina de movimiento continuo. En términos de ingeniería, una “*flywheel*” (en español, volante de inercia) es una herramienta diseñada para almacenar eficientemente la energía rotacional. Funciona almacenando energía cuando una máquina no está trabajando a un nivel constante y en lugar de desperdiciar energía encendiéndose y apagándose, dicha “*flywheel*” mantiene la energía constante y la propaga a otras áreas de la máquina.

Así, la empresa desarrolló una firme “*flywheel*” en torno a la IA, en el que las innovaciones y mejoras en el aprendizaje de la máquina en una parte de la empresa alimentan el trabajo de otros sectores de la misma, los cuales pueden construir productos o lanzar servicios para influir en otras partes de la empresa y en la empresa en general. Esencialmente, lo que se crea en una parte de Amazon actúa como catalizador para el crecimiento de la IA y el aprendizaje automático en otras áreas. Así, la IA no se centra en un departamento en particular en Amazon, ya que digamos que es más como una constante en la empresa que atraviesa todos los departamentos (Skuzza, 2020). La IA y el aprendizaje automático potencian tres productos populares de Amazon: Alexa, el Amazon Go Store, y el motor de recomendación de Amazon ya mencionado anteriormente.

Amazon Echo, que cuenta con el *bot* de IA Alexa, ha sido una de las iniciativas más populares de la compañía en el aprendizaje automático. Amazon se enfrentó a una batalla difícil al principio, especialmente porque fue una de las primeras compañías en intentar crear un asistente virtual con voz que pudiera caber en una encimera. Una vez que la tecnología comenzó a unirse, las divisiones de la compañía se dieron cuenta de que Alexa podría ser beneficiosa para sus productos. Algunas de las primeras habilidades de Alexa fueron la integración con Amazon Music, Prime Video y recomendaciones personalizadas de productos de una cuenta de Amazon (Morgan, 2018).

La tienda de Amazon Go, la cual funciona sin cajeros, también aprovechó la riqueza de datos para rastrear las tendencias de compra de los clientes. Los datos de las cámaras de los teléfonos de los clientes rastrean las actividades de compra y no sólo ayudan a Amazon Go, sino que también pueden ser compartidos con el departamento de aprendizaje automático (o *machine learning* en inglés) para un desarrollo continuo.

Así es como funciona su *flywheel*, a través de los datos de estos tres pilares principales de la compañía trabajan juntos para crear una experiencia cohesiva y completa para el cliente. Pongamos un ejemplo en el que un cliente visita la tienda de Amazon Go para conseguir algunos ingredientes para preparar su comida, y tras pedir a Alexa que busque una receta, el motor de recomendación de productos de Amazon puede determinar que el cliente probablemente necesite comprar un cierto tipo de utensilio para prepararla y sugerírselo. En lugar de pelearse entre sí, las diferentes divisiones comparten los conocimientos que adquieren para proporcionar una experiencia de cliente completa y personalizada.

Este sistema ha sido tan exitoso para Amazon que la compañía ahora vende este enfoque de aprendizaje automático a través de Amazon Web Services a clientes como la NASA, y aprovechando los avances y aplicaciones de la IA en otras áreas de la empresa, ofrece soluciones personalizadas de IA a grandes y pequeñas empresas (Morgan, 2018).

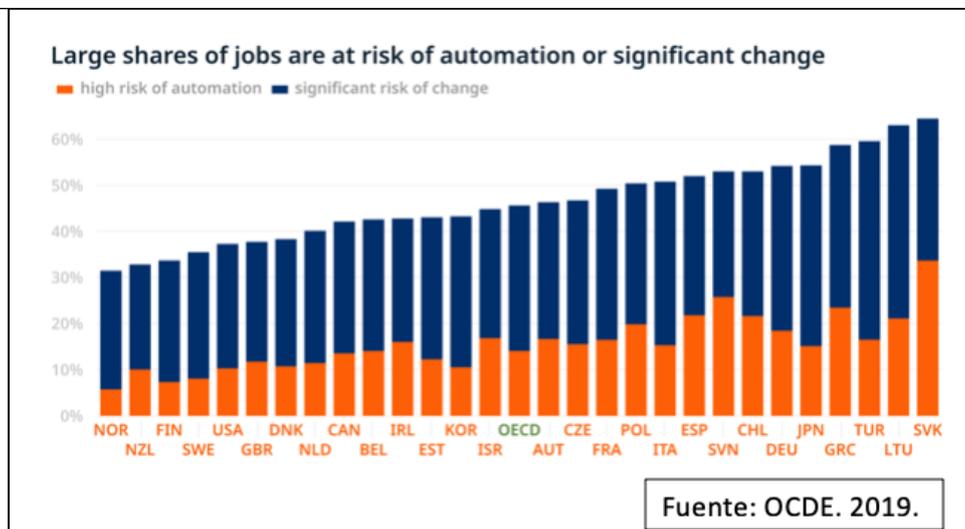
Tal es la importancia y el peso que va a tener este tipo de tecnología con IA enfocada a las empresas, que la consultoría PwC estima que para el año 2030, este sistema de gestión de empresas producirá un impacto económico de 16 billones de dólares. Por ello, no es casualidad que la IA fuera la piedra angular de muchas de las novedades expuestas en la mayor feria de tecnología del mundo, la CES 2020. A pesar de que el nivel actual de implementación de soluciones de Inteligencia Artificial es de alrededor del 40% de las empresas en todo el mundo, esta cifra cambiará drásticamente en los próximos 2 años, llegando hasta el 90% de implementación en las empresas según Ritika Gunnar, vicepresidente de Inteligencia Artificial y Laboratorios de Datos y Aprendizaje de IBM (Riquelme, 2020).

Pero, también hay que tener en cuenta que puede que no todo lo que venga de la mano con la IA sean buenas noticias. La pérdida de puestos de trabajo es una de las enormes

amenazas que presenta la IA. En la última década los humanos han dependido del trabajo físico y de invertir su tiempo para cobrar un sueldo, pero viendo cómo en una amplia variedad de industrias en las que la tecnología ya está presente las máquinas dan mejores resultados que los humanos en cuanto a productividad y eficiencia, en cuanto la adquisición de esta tecnología se abarate, parece que el reemplazo del humano va a ser inevitable para la supervivencia de los negocios. Las consecuencias más drásticas son las que se obtienen por nivel de educación, con tasas de automatización potencial mucho más bajas en promedio para los trabajadores altamente educados con títulos de postgrado o superiores, que para los que tienen niveles de educación bajos o medios (PwC, 2018).

Según informes de Bloomberg, hasta 120 millones de trabajadores en todo el mundo necesitarán ser reciclados debido al impacto de la inteligencia artificial en los puestos de trabajo, basándose en una encuesta de IBM (Hagan, 2019). La cantidad de individuos que serán impactados es inmensa, ya que se cree que más de 50 millones de trabajadores chinos pueden requerir de reentrenamiento como resultado del despliegue relacionado con la IA. Los EE.UU. tendrán que reentrenar a 11,5 millones de personas con las habilidades necesarias para sobrevivir en el mercado laboral (Kelly, 2019). En el caso de España, según un informe de la OCDE publicado en 2019, aproximadamente el 50% de los empleos del país están en riesgo debido a la potencial automatización de las empresas (Ver figura 1) (Sanhermelando, 2019). Y es que según este informe, el 14% de los empleos actuales sufren el riesgo de ser reemplazados por la automatización en los próximos 15-20 años, mientras que un 32% de los empleos podrían sufrir un cambio radical a consecuencia de la automatización de parte de sus tareas.

**Figura 1: Empleos en riesgo por la automatización o por un cambio significativo**



Pero, del mismo modo que esta automatización destruirá gran cantidad de puestos de trabajo, también está creando y creará muchos otros debido a las necesidades de estas nuevas máquinas, no solo por el constante incremento de la demanda, sino que también en lo relativo con su diseño y desarrollo y lo relativo a su mantenimiento y reparación. Según un informe del Foro Económico Mundial publicado en 2018, la automatización destruirá directamente 75 millones de empleos, pero creará 133 millones para 2022. Por su parte, McKinsey estima que, con un crecimiento económico, una innovación y una inversión suficientes en todo el mundo, puede haber suficiente creación de nuevos puestos de trabajo para compensar el impacto de la automatización, aunque en algunas economías avanzadas se necesitarán inversiones adicionales para reducir el riesgo de escasez de puestos de trabajo, estimando que por ejemplo, en los EE.UU., habrá un crecimiento positivo neto del empleo hasta 2030 (Press, 2019).

La estimaciones sobre el aumento del desempleo y sobre la creación de empleos son solo eso, estimaciones, ya que no se puede predecir a ciencia cierta cuál va a ser la tendencia real de la fuerza laboral en este incipiente mercado tecnológico. Lo que sí que está claro es que para bien o para mal, el cambio que se producirá en el mercado laboral es inevitable, aunque no podamos cuantificarlo exactamente, por lo que es necesario anticiparse a él antes de que las consecuencias sean devastadoras tanto a nivel económico como social por una pérdida masiva de empleos.

Como ya se ha expuesto anteriormente, serán los trabajadores poco cualificados los que se verán más afectados por esta tendencia de automatización de las cadenas de producción, ya que serán sus trabajos los que serán más fáciles reemplazar por una máquina. Del mismo modo, según el ya citado informe de McKinsey, se espera que entre el 8 y el 9 por ciento de la demanda de mano de obra en 2030 sea en nuevos tipos de ocupaciones que no han existido antes. Esto requerirá cambios en los programas de formación y educación para preparar nuestra futura fuerza de trabajo, así como ayudar a los trabajadores actuales en la transición a nuevos puestos que utilizarán sus capacidades humanas. Esta tendencia de formación y educación ya está a la orden del día, ya que, por ejemplo, Amazon anunció en 2019 que gastará 700 millones de dólares para formar a unos 100.000 trabajadores en los EE.UU. para el año 2025, ayudándoles a optar a puestos de trabajo más cualificados (Press, 2019).

Por su parte, los trabajadores con mayor cualificación no presentan tanto riesgo, ya que debido a esta cualificación presentan mayor adaptabilidad a los cambios tecnológicos y es más probable que ocupen puestos directivos superiores, los cuales seguirán siendo necesarios para aplicar el criterio humano, así como para diseñar y supervisar los sistemas basados en la inteligencia artificial, por lo que estos trabajadores podrían incluso experimentar un incremento en sus salarios debido a las ganancias de productividad que estas nuevas tecnologías deberían traer. Esto acarrearía una mayor polarización de la riqueza, aumentando consigo las desigualdades sociales, ya que se produciría una tendencia en la que la riqueza se dirija hacia los empleos con mayor cualificación, aumentando la pobreza en las clases menos cualificadas. Por ello, parece necesario que la respuesta a estos cambios se produzca desde la educación y del fomento de la formación, pues en el caso contrario, el desarrollo de la IA nos llevará a un mundo en el que las desigualdades sociales estarán mucho más acentuadas.

#### **b. Una nueva democracia**

Como es obvio, el desarrollo de la IA no sólo traerá consecuencias en la economía global, sino que, como ya se ha podido experimentar en los últimos años, los Estados occidentales están viendo cómo sus sistemas democráticos corren un grave riesgo. Y es que la injerencia de empresas especializadas en el procesamiento de datos a través de la IA ya ha causado estragos en una de las democracias que comúnmente consideramos como una de las más sólidas del planeta, como Estados Unidos. Esto nos lleva a plantearnos cuestiones bastante preocupantes: ¿sobrevivirá la democracia representativa al auge de la IA y del Big Data? ¿Nos llevará hacia democracias más participativas o por el contrario nos llevará a democracias más débiles y más propensas a sistemas totalitarios?

La tecnología digital puede utilizarse al servicio de sociedades liberales o autoritarias, fortaleciendo tanto la responsabilidad del gobierno como la capacidad de represión. Esta tecnología ha dado lugar a un acceso y un intercambio de información sin precedentes y ha amplificado la difusión de la desinformación y la propaganda, contribuyendo así al aumento del populismo y a la polarización de las sociedades democráticas.

Las sugerencias personalizadas de productos y servicios que nos ofrecen los ya mencionados motores de búsqueda provienen de información que se basa en datos personales y meta-datos que han sido recogidos de búsquedas previas, compras y comportamientos de movilidad, así como interacciones sociales. Si bien oficialmente la identidad del usuario está protegida, en la práctica puede inferirse con bastante facilidad. Hoy en día, los algoritmos saben bastante bien lo que hacemos, lo que pensamos y cómo nos sentimos, posiblemente incluso mejor que nuestros amigos y familiares o incluso que nosotros mismos. A menudo las recomendaciones que nos ofrecen encajan tan bien con nuestros intereses que las decisiones que tomamos se sienten como si fueran nuestras propias, aunque en realidad no lo sean. De hecho, nunca habíamos estado siendo controlados con tanto éxito como hoy en día, ya que, cuanto más se sabe de nosotros, menos probable es que nuestras decisiones sean libres y no estén predeterminadas por otros.

Este control sobre nuestros gustos y preferencias se extrapola a otros de los aspectos de nuestra personalidad, como las preferencias políticas o nuestra intención de voto, por lo que no resulta difícil ser capaz de utilizar recomendaciones y sugerencias políticas personalizadas del mismo modo en el que lo realizan los motores de búsqueda como el de Amazon para que nos decantemos por un candidato u otro. Esta fue la estrategia de la empresa británica Cambridge Analytica durante las elecciones presidenciales de Estados Unidos de 2016.

Cambridge Analítica era una empresa de consultoría política creada en 2013 y establecida en Reino Unido, la cual, mediante apropiación indebida de recursos digitales, tenía como objetivo influir en la intención de voto de los electores durante las campañas electorales a través de la combinación de la extracción, intermediación y el análisis de datos con la comunicación estratégica.

Todo comienza en 2013, cuando la empresa recibió 15 millones de dólares de Robert Mercer, un donante multimillonario afín al Partido Republicano estadounidense, y se le encomendó la tarea de identificar las personalidades de los votantes estadounidenses e influir en su comportamiento. Para llevar a cabo esta tarea, Cambridge Analytica diseñó un test de personalidad por cuatro investigadores de la Universidad del Estado de Michigan.

Para realizar dicho test, la empresa desarrolló una encuesta de 120 preguntas, a través de las cuales se calculaba el resultado a través de cinco rasgos de personalidad: Apertura, Conciencia, Extroversión, Aceptación y Neurotismo, lo cual daba a los encuestados una puntuación llamada “OCEAN” (por las siglas en inglés de dichos rasgos). El modelo OCEAN agrupaba rasgos de personalidad a través de rasgos personales que suelen mantenerse en el tiempo, es decir, rasgos de la personalidad de los encuestados que fueran a mantener a lo largo de los años. Al tratarse de una encuesta tan larga, la estrategia de Cambridge Analytica (en adelante, CA) para incentivar a los usuarios a realizarla fue ofrecer un incentivo financiero, es decir, que a cambio de realizar la encuesta se ofrecía un pago de entre 2 y 4 dólares (Hern, 2018).

Para alcanzar el objetivo que se le había encomendado a la empresa (influir a favor del Partido Republicano en las elecciones presidenciales de 2016), CA necesitaba ampliar la red de encuestados. Es entonces cuando Facebook entra en la ecuación. Para que la persona recibiera el pago al realizar la encuesta, se les requería que iniciaran sesión a través de Facebook y aprobar el acceso de la App en la que se realizaba la encuesta. Este era el paso en el que se “engañaba” a los usuarios de Facebook, ya que el usuario solo tenía que iniciar sesión para realizar la encuesta y una vez realizada se le hacía el pago, pero con ese simple ‘click’ CA estaba realizando dos movimientos.

El primero era recolectar toda la información posible del usuario que acababa de iniciar sesión, incluyendo información personal identificable como el nombre real, la ubicación y los detalles de contacto. Esto permitía a CA hacer una especie de “inventario”, ya que podía establecer una relación entre la persona física a la que le acababa de sustraer los datos y el censo electoral donde se encontraba. El segundo movimiento era que la App de la encuesta también recolectaba la misma cantidad de datos de todos los amigos de Facebook del usuario que la instaló e inició sesión (Hern, 2018). De repente, de 270.000 personas a las que se les pagaron un par de dólares al realizar la encuesta se obtienen los datos de 87 millones de usuarios de Facebook de los cuales puede aprovecharse CA para su objetivo (Salinas, 2018).

Aquí es cuando el procesamiento de Big Data a través de la IA comienza. Usando las encuestas realizadas como base, se crearon distintos perfiles de votantes en los que se

encajaban los datos de los 87 millones de usuarios de Facebook. En otras palabras, todos los “likes” y las preferencias que mostraran los datos de los usuarios de Facebook eran procesados y con ellos se podía encuadrar a dichos usuarios en los distintos perfiles creados en base a las encuestas, lo que permitía a CA poder aproximarse a las personalidades y afiliaciones políticas de 87 millones de votantes potenciales (Hern, 2018). Y así, los algoritmos combinaron los datos con otras fuentes, como los registros de los votantes, para crear un conjunto superior de datos.

Una vez procesados todos estos datos y encuadrados en los distintos perfiles de votantes, estos individuos podían ser blanco de una publicidad altamente personalizada basada en sus datos de personalidad. Como explicó Christopher Wylie (el trabajador de CA que dio la voz de alarma de las actuaciones de la empresa), la precisión de los perfiles de personalidad era tan alta que según su personalidad se enviaban anuncios diferentes a personas que fueran votantes del mismo partido político. Es decir, por ejemplo se enviaban distintos anuncios según el tipo de votantes del Partido Demócrata, ya que según su personalidad, el anuncio que a un tipo de demócrata le podría llevar a cambiar su intención de voto o abstenerse de votar, a otro tipo de demócrata le produciría rechazo y afirmaría su voto al Partido Demócrata.

Estos anuncios en línea se difundieron principalmente a través de bots cuyo propósito era difundir propaganda política en las distintas redes sociales, como Twitter, YouTube, e incluso en Google, pero principalmente en Facebook. Los anuncios que más gustaban y más se compartían fueron reproducidos y redistribuidos en función de su popularidad y de quiénes eran sus destinatarios. Así, gracias al procesamiento de estos grandes conjuntos de datos a través del aprendizaje automático, Cambridge Analytica pudo usar esta información en tiempo real para determinar cuáles eran los mensajes que más estaban resonando y dónde, y así, en base a esa información poder dar forma al programa de viajes de Trump (Illing, 2018). De este modo, si había un pico de “likes” en un artículo sobre inmigración en un condado de Colorado o Luisiana, Trump iría allí y daría un discurso centrado en la inmigración.

Pero, toda la estrategia de CA puede llevar al lector a preguntarse si realmente la publicación de anuncios políticos de forma indiscriminada en distintas redes sociales puede llevar a alterar los resultados en unas elecciones con un nivel de electores tan alto

como las de Estados Unidos. Por desgracia, la respuesta es que sí. Ya que el margen de victoria en muchas elecciones estadounidenses es sólo de un pequeño porcentaje, y si además le sumamos que la estrategia de CA actuó principalmente en Estados que se conocen como “*swing states*”, el porcentaje de influencia en las elecciones aumenta. Un “*swing state*” puede definirse como un estado de los Estados Unidos en el que los candidatos republicanos y demócratas tienen niveles de apoyo similares y que se considera que desempeña un papel clave en el resultado de las elecciones presidenciales (Merriam-Webster, s.f.). Así, como explica Brittany Kaiser, una de las directoras de CA que dio la voz de alerta de lo que allí estaba sucediendo, uno de los anuncios más efectivos empleados fue un anuncio en una sitio web de noticias políticas, el cual contenía un gráfico interactivo que parecía un artículo periodístico y que pretendía enumerar “10 verdades incómodas sobre la Fundación Clinton” (es decir, propaganda política deslegitimando a la candidata demócrata, Hillary Clinton), el cual apareció durante varias semanas a individuos de una lista de los principales “*swing states*” cuando visitaban la página web (Lewis & Hilder, 2018).

Además, es importante tener en cuenta que las tecnologías y metodologías de aprendizaje automático que se utilizan hoy en día son todavía relativamente “jóvenes”, y sólo estamos empezando a descubrir el verdadero potencial de estos algoritmos. Teniendo en cuenta la influencia que tuvo esta herramienta utilizada por CA, la cual ha sido (que se sepa) la primera de este tipo en utilizarse, a medida que los algoritmos de aprendizaje automático se vuelvan cada vez más sofisticados, también aumentará su poder de influencia.

Gracias al escándalo de CA en las elecciones presidenciales de Estados Unidos y sus implicaciones en las elecciones de otros países como República Checa, Italia, Ucrania, Kenia, Nigeria, Méjico, Brasil, India o Malasia, entre otros (BBC, 2018), en Reino Unido se ha tomado conciencia de el potencial uso peligroso y poco ético de la tecnología, por lo se ha empezado a tratar el tema de la regulación de la IA.

La comisión de la Cámara de los Lores sobre la IA desarrolló una serie de recomendaciones en lo relativo a la regulación de la IA, y en el centro de dichas recomendaciones hay cinco principios éticos que según la comisión deberían aplicarse en todos los sectores, a nivel nacional e internacional. Estos principios son:

- I. *La inteligencia artificial debe ser desarrollada para el bien común y el beneficio de la humanidad.*
- II. *La inteligencia artificial debería operar sobre principios de inteligibilidad y justicia.*
- III. *La inteligencia artificial no debe utilizarse para menoscabar los derechos a los datos o la privacidad de las personas, las familias o las comunidades.*
- IV. *Todos los ciudadanos deben tener el derecho a recibir una educación que les permita prosperar mental, emocional y económicamente junto con la inteligencia artificial.*
- V. *El poder autónomo de herir, destruir o engañar a los seres humanos nunca debe conferirse a la inteligencia artificial.*

Según la comisión, el objetivo no es incluir estos principios directamente en la legislación, sino más bien tenerlos como una guía en la que basarse para la regulación de la IA. Sin duda, estos principios deberían servir como, valga la redundancia, principios éticos en los que basar el resto de legislación relativa a la IA, y así lo explica Lord Clement-Jones, el presidente de la mencionada comisión: *“Los legisladores tienen que hacer la conexión con la ética, y así es como debemos que creemos que deben hacerlo”* (Hern, 2018).

Entonces, ¿qué debería hacer la gente común para prevenir estas prácticas fraudulentas? Pues aumentar la conciencia sobre quién está usando sus datos y cómo. Antes de aceptar los términos y condiciones al crearse un perfil en redes sociales y empezar a compartir fotos propias o incluso de sus hijos, deberían hacer un ejercicio de información, y si se está de acuerdo con lo que implica y lo que podría significar más adelante, bien, pero al menos se debería adquirir conocimiento primero.

Las universidades deberían también crear espacio en sus programas para los títulos híbridos. Deben incentivar a los estudiantes a estudiar literatura comparada, religiones del mundo, microeconomía, antropología cultural y cursos similares combinándolos con los títulos que se propongan estudiar. Es decir, deberían abogar por programas de doble titulación, por ejemplo, combinando el título en informática con las relaciones internacionales, teología, ciencias políticas, filosofía, salud pública, educación o

similares. Esta medida debería abogar por un estudio profundo de ética, no impartirla como una clase independiente, sino combinándolo con el área de estudio principal, para poder contar con una perspectiva ética a la hora emprender la andadura en el mercado laboral.

Por ello, parece que la supervivencia de la democracia representativa de los estados occidentales depende una vez más de la formación de la sociedad en su conjunto. Cuanto más información se posea acerca de los beneficios y de los riesgos que acarrea la IA, más preparado estará el individuo para hacer frente al futuro que nos aguarda, donde sin duda, la IA tendrá un peso mucho mayor del que tiene hoy en día, cambiando radicalmente nuestra forma de ver el mundo y de interactuar como sociedad. Por ello, el estudio de la ética debería ser uno de los pilares fundamentales para fomentar una tendencia formativa, ya que podríamos considerarla como el pilar en el que se basarían el resto de acciones que realice el individuo, al poder implementar la razón moral como una parte más de la razón humana misma y permitir hacer un juicio de aquello que es ético y lo que no. Esta ética ligada a la IA serviría de base para saber cuales son los comportamientos contrarios a la ética a los que se podría enfrentar el individuo, lo que le permitiría actuar desde una posición de poder, al saber cuales son las consecuencias de los usos fraudulentos de la IA no solo en el ámbito democrático, sino en el resto de aspectos de su vida.

### c. ¿Un nuevo futuro?

Estamos viendo cómo la injerencia de la IA va a cambiar radicalmente el mundo que conocemos, afectando tanto a la economía como a la política, por lo que todos estos cambios inevitablemente van a afectar a nuestra sociedad y la forma en la que interactuamos entre nosotros. La pregunta que nos tenemos que hacer entonces es, ¿en qué medida va a cambiar? Y, sobre todo, ¿hacia qué tipo de sociedad nos llevará la IA? ¿A una sociedad más libre en la que la IA se implementa en la educación, los sistemas sanitarios y en tratar de erradicar la pobreza o el cambio climático o por el contrario a una sociedad más totalitaria en la que la IA se implementa en un control más exhaustivo del ciudadano en pos de la ‘seguridad’ y en la que aumentan las desigualdades sociales?

Cuando en 1949 George Orwell publicó su novela distópica “1984”, nos presentó lo que nos parecía un personaje de ciencia ficción, el “Gran Hermano”. Este personaje,

aunque no aparece directamente, está presente a lo largo de toda la novela a través de carteles omnipresentes que muestran su fotografía acompañándola con el lema “El Gran Hermano te vigila” o a través de dispositivos ocultos en cada habitación que permiten a la “Policía del Pensamiento” vigilar las actividades de todos los ciudadanos.

En estos últimos años hemos visto cómo gracias al uso de la IA, el “Gran Hermano” ha pasado de ser un personaje de una novela distópica a una realidad presente en nuestro mundo. Y es que el empleo de la IA en sistemas de vigilancia que utilizan el reconocimiento facial para un control más exhaustivo de la población ha empezado a utilizarse en países totalitarios como China, convirtiendo lo que antaño nos parecía una situación inconcebible propia de futuro distópico, en una práctica real y ya presente en nuestro planeta, planteando una seria amenaza para nuestras libertades.

Desde hace unos años, China, un país que cuenta en su territorio con más de 200 millones de cámaras de seguridad (Díez, 2019), ha comenzado a implantar cámaras de seguridad con reconocimiento facial gracias a la IA, lo cual ha potenciado un gran debate sobre los límites de la “seguridad” respecto de la privacidad del individuo. De este modo, China está revirtiendo la visión que generalmente se tiene de la tecnología, una tecnología con un gran poder de democratización, que da una mayor libertad a los individuos y que los conecta con el mundo.

En algunas ciudades chinas, las cámaras escanean las estaciones de tren en busca de los delincuentes más buscados de China, los carteles electrónicos muestran las caras aquellos que cruzan la calle indebidamente y los nombres de los que no pagan sus deudas, y del mismo modo, vigilan las entradas a complejos de viviendas. Estas medidas se complementan con otros métodos que controlan el uso de Internet y de las comunicaciones, así como el alojamiento en hoteles, los desplazamientos en avión, en tren e incluso en coche a determinados lugares del país (Mozur, 2018).

John Wethington, un investigador sobre seguridad y ciberseguridad descubrió una gran base de datos relativa a los datos recogidos a través de las cámaras de seguridad con reconocimiento facial. Las personas estaban registradas en base a datos como sus antecedentes penales, datos de reconocimiento facial (como por ejemplo si llevaban barba o una máscara), e incluso de qué etnia eran, diferenciando entre los Han, los cuales son

el grupo étnico mayoritario en China, y los Uigur, una minoría musulmana (Feng, 2019). Esos métodos se están convirtiendo en la corriente principal en toda China, incluyendo ciudades como Beijing.

Al problema de la violación de la privacidad que acarrea este tipo de vigilancia con IA se le suma el hecho de que dichos sistemas de reconocimiento facial se están empleando desde el gobierno chino para controlar y perseguir a la minoría musulmana de los Uigur. Esta campaña está centrada en combatir casi todas las expresiones de la fe islámica uigur como signos de posible extremismo religioso y separatismo étnico y se focaliza en la región autónoma de Xinjiang, donde casi la mitad de sus 25 millones de habitantes son Uigures y donde se estima que más de un millón de ellos han sido enviados a campos de “re-educación” (Byler, 2019).

Este es el primer ejemplo conocido de un gobierno que utiliza intencionadamente la inteligencia artificial para la elaboración de perfiles raciales, marcando en potencia el comienzo de una nueva era de racismo automatizado. El problema reside en que, a pesar de la amplia condena de estas prácticas por parte de la comunidad internacional, no parece que hayan tenido ningún tipo de repercusión en un gobierno totalitario como el chino. Esta ausencia de repercusión por parte del gobierno queda plasmada en nuevas normas encaminadas a un control cada vez más exhaustivo de la población, como por ejemplo a finales de 2019, cuando el gobierno impuso que todos los usuarios de teléfonos móviles en China que registren nuevas tarjetas SIM deben someterse a escáneres de reconocimiento facial. La norma exige que las empresas de telecomunicaciones desplieguen IA y otros métodos técnicos para comprobar la identidad de las personas que registran las tarjetas SIM. La medida, descrita por el Ministerio de Industria e Información como una forma de “proteger los derechos e intereses legítimos de los ciudadanos en el ciberespacio”, facilitando el seguimiento de los usuarios chinos de teléfonos móviles e Internet (Kuo, 2019).

Así, los intentos del gobierno Chino de llevar un control total de la sociedad siguen progresando de manera constante, como hemos podido ver en los últimos meses desde la irrupción de la pandemia del COVID-19, de la que Beijing ha conseguido sacar partido. Desde la llegada de la pandemia se ha visto cómo China está desplegando un sistema de vigilancia masiva sin precedentes, en un esfuerzo por combatir la propagación del

coronavirus. Y es que, a medida que millones de personas regresen a las grandes ciudades y vuelvan a sus puestos de trabajo, serán objeto de una vigilancia total, ya que, la compañía china Hanwang afirma que ha mejorado su tecnología de reconocimiento facial para reconocer a las personas que llevan mascarillas con una precisión del 95% (Wood, 2020), y a las cámaras de seguridad presentes ya en muchos lugares, se les suma la instalación de nuevas tecnologías de reconocimiento facial y de escaneo de temperatura en los sistemas de transporte público (Yu, 2020).

Las tecnologías desarrolladas por China de manera opresiva en Xinjiang y de manera menos agresiva en ciudades como Beijing, se desarrollan de manera implacable bajo un impulso subvencionado por el Estado hacia una posición dominante en el sector de la seguridad. Este programa que ha impulsado el desarrollo de tecnología de vigilancia avanzada cuenta con la “ventaja” de que las empresas y desarrolladores cuentan con una total ausencia de restricciones éticas y legales por parte del gobierno a la hora del desarrollo e implementación de dichas tecnologías. Por supuesto, desde el gobierno se dice que se estos métodos se usan para mantener la seguridad ciudadana, pero la preocupación a nivel internacional sigue creciendo por el uso de los datos por parte del “Gran Hermano”, dando a la policía el poder de crear un estado digital autoritario. El problema es que, una nueva generación de nuevas empresas que atienden a las necesidades autoritarias de Beijing están empezando a marcar las pautas para tecnologías emergentes como la IA, por lo que herramientas similares podrían desarrollarse, llevándonos a una sociedad totalmente controlada, e incluso automatizando los sesgos basados en el color de la piel y el origen étnico, llevando a un racismo automatizado.

Este desarrollo de la IA parece el peor de todos los posibles. El reconocimiento facial potenciado por la IA es barato y fácilmente programable, por lo que es ampliamente disponible, pero también es preocupantemente inexacto. Tanto es así que ha llegado de manera incipiente hasta Reino Unido, donde los ensayos por parte de la policía en Londres y el sur de Gales, mostraron una alta tasa de falsas detecciones de delincuentes, siendo la tasa de inexactitud mucho mayor con las caras negras que con las blancas, por lo que diferentes grupos han pedido la moratoria del uso de estos sistemas (The Guardian, 2019). Una tecnología que en la vida real no puede discernir entre individuos sólo tenderá a aumentar la cantidad de discriminación en la sociedad en su conjunto.

Este tipo de tecnología ha ido mejorando constantemente, y no es descabellado esperar que dentro de unos pocos años sea aceptablemente fiable, así como generalizada. Esto planteará un conjunto de problemas diferentes a los obvios que se plantean cuando no funciona, y éstos no se resolverán con un intento de prohibirlo por completo, ya que, incluso si hay una limitación por parte de los gobiernos, los actores privados irán por delante. Un ejemplo de cómo estos sistemas pueden favorecer al sector privado, es la enorme ventaja con la que contarían los minoristas que tratan de frenar los robos en las tiendas sin mucha ayuda de una fuerza policial sobrecargada.

Por lo tanto, los peligros radicarán no sólo en la recolección de datos, sino también en su posterior almacenamiento. Tanto las cámaras de vigilancia, como la mayoría del llamado “Internet de las cosas”<sup>7</sup>, presentarán muchas vulnerabilidades para determinados hackers, como también lo harán las bases de datos donde se almacena su información. El marco jurídico e institucional en el que deben gestionarse estos peligros todavía no está claro; sin embargo, la urgencia del problema es evidente y, ante el avance imparable de estos sistemas tecnológicos, es necesario empezar a desarrollar un posible marco jurídico que marque unas directrices en las que basarse para lo que previsiblemente se avecina, así como empezar a tomar medidas aunque sea en lo relativo a la prevención de los riesgos que esta tecnología acarrea (The Guardian, 2019).

Pese a este tipo de sistemas que deriva hacia una sociedad cada vez menos libre, hay aplicaciones de la IA que pueden implementarse para el bien social, creando individuos más libres a través de sistemas educativos más completos y especializados, mejorando nuestra calidad de vida a través de su implantación en los sistemas sanitarios o aplicándolo a posibles medidas medioambientales para un planeta más sostenible.

A medida que las aplicaciones educativas de la IA continúan madurando, la esperanza es que la IA pueda ayudar a llenar las lagunas de necesidades en el aprendizaje y la enseñanza y permitir que las escuelas y los maestros hagan más que nunca antes. La IA puede impulsar la eficiencia, la personalización y la racionalización de las tareas administrativas para dar a los profesores el tiempo y la libertad de proporcionar

---

<sup>7</sup> La interconexión a través de Internet de los dispositivos informáticos incorporados en los objetos cotidianos, que les permite enviar y recibir datos (Lexico, s.f.).

comprensión y adaptabilidad, capacidades humanas únicas donde las máquinas tendrían dificultades. Cada estudiante es único y la IA ayudará a adaptar los cursos en función de sus capacidades. La IA y el aprendizaje automático, serán el *modus operandi* a través del cual tendremos programas de estudio individualizados y no sólo el modelo de “la misma clase para todos” que se ha convertido en el estándar mundial (Sharma, 2017). De hecho, en 2016, Facebook ya empezó a probar un software que permite a los estudiantes estar a cargo de sus aprendizajes.

De este modo, en lugar de que los profesores repartan las tareas de clase, el sistema de gestión de aprendizaje *Facebook-Summit* pone a los estudiantes a cargo de la selección de sus proyectos y de marcar su propio ritmo. La idea es animar a los estudiantes a desarrollar habilidades, como la ingeniosidad y la gestión del tiempo, que puedan ayudarles a tener éxito en la universidad. El software ofrece a los estudiantes una visión completa de sus responsabilidades académicas para el año en cada clase y las desglosa en módulos de lecciones personalizables que pueden abordar a su propio ritmo. Así, el sistema invierte la jerarquía tradicional de las clases dirigidas por maestros, requiriendo que las escuelas provean tutoría y entrenamiento intensivo uno a uno para ayudar a cada estudiante a adaptarse (Singer & Isaac, 2016).

Ajustar el aprendizaje en base a las necesidades particulares de cada estudiante ha sido una prioridad para los educadores durante años, pero la IA permitirá un nivel de diferenciación que es imposible para los profesores que tienen que manejar 30 estudiantes en cada clase. Hay varias empresas como Content Technologies y Carnegie Learning que están desarrollando actualmente un diseño de enseñanza inteligente y plataformas digitales que utilizan la IA para proporcionar aprendizaje, pruebas y retroalimentación a los estudiantes desde el nivel pre-escolar hasta la universidad, les proporciona retos para los que están preparados, identifica las lagunas de conocimiento y los redirige a nuevos temas cuando es apropiado. A medida que la IA se hace más sofisticada, podría ser posible que una máquina leyera la expresión del rostro de un estudiante que indica que está luchando por captar un tema y modificará una lección para responder a eso. La idea de personalizar el plan de estudios para las necesidades de cada estudiante no es viable hoy en día, pero lo será para las máquinas impulsadas por la IA (Marr, 2019).

De este modo, las aplicaciones de la IA pueden ayudar a “democratizar” la educación, haciéndola accesible para todos, al poder desarrollar herramientas para llegar a cada estudiante, incluyendo a aquellos que hablan diferentes idiomas o que podrían tener discapacidades visuales o auditivas. Por ejemplo, *Presentation Translator* es un plug-in<sup>8</sup> gratuito para PowerPoint que crea subtítulos en tiempo real para lo que el profesor está diciendo, lo que abre posibilidades a los estudiantes que no puedan asistir al colegio debido a una enfermedad o que requieren un aprendizaje a un nivel diferente o sobre un tema particular que no está disponible en su colegio (Microsoft, s.f.).

Combinando las mejores características de las máquinas con las de los profesores, la implementación de la IA en la educación nos hará obtener un mejor resultado para los estudiantes. La IA puede ayudar a romper los moldes de la educación convencional llevándonos, no sólo hacia una educación al global al alcance de todos, sino que también hacia una educación personalizada para cada estudiante dependiendo de sus capacidades e intereses. Dado que los estudiantes de hoy necesitarán trabajar en un futuro donde la IA es la realidad, es importante que nuestras instituciones educativas expongan a los estudiantes a la tecnología y la usen, pero la labor de los profesores no puede ser reemplazada en su totalidad, ya que el factor humano es básico en la educación, no sólo por el factor emocional y empático del que carece la máquina, sino porque también ayuda al estudiante a adquirir respeto por la autoridad y a respetar normas, que le prepararán para vivir en sociedad.

Del mismo modo que en la educación, la implantación de la IA dentro del ámbito sanitario presenta una vasta cantidad de potenciales aplicaciones. Este potencial de la IA a la hora de prevenir y detectar enfermedades, así como de su aplicación en la investigación científica, no sólo salvarían millones de vidas, sino que ahorraría millones a los sistemas sanitarios de los estados. Y, al igual que en nuestra vida diaria, ya estamos viendo cómo la IA y la robótica están cada vez más presentes en nuestros sistemas sanitarios, ya que se prevé que el valor de mercado de la IA en la industria de la salud alcanzará los 6.600 millones de dólares en 2021 (Mills, 2020).

---

<sup>8</sup> Un plug-in es un componente de un programa de software que puede ser añadido para dar soporte a características o funcionalidades específicas (Techopedia, 2019).

Como ya se presentó anteriormente, una de las aplicaciones de la IA que ya se está probando en la medicina es Watson Health, de IBM (en adelante, Watson). La primera aplicación comercial de Watson se produjo en 2013, cuando el Memorial Sloan Kettering Cancer Center comenzó a utilizar el sistema para recomendar opciones de tratamiento para los pacientes con cáncer de pulmón, a fin de garantizar que recibieran el tratamiento adecuado y, al mismo tiempo, reducir los costos. Desde entonces, proveedores como la Clínica Cleveland, el Centro de Medicina del Cáncer de Maine y el Grupo Médico Westmed también han implementado las herramientas de Watson. Watson procesa a una velocidad de 80 *teraflops*<sup>9</sup> (billones de operaciones de coma flotante por segundo), superando así la capacidad de un humano de responder a las preguntas, ya que, Watson accede a 90 servidores con un almacén de datos combinado de más de 200 millones de páginas de información, que procesa a través de seis millones de reglas lógicas (Rouse, 2018).

De este modo, Watson está ayudando a los sistemas sanitarios a aplicar la tecnología cognitiva para desbloquear grandes cantidades de datos de salud y diagnósticos, ya que Watson puede revisar y almacenar mucha más información médica (cada diario médico, cada síntoma, cada paciente y estudio de los tratamientos y respuestas alrededor del mundo) exponencialmente más rápido que cualquier humano. Y así, como ya se adelantó anteriormente, al combinar expertos humanos con inteligencia aumentada, Watson ayuda a los profesionales de la salud e investigadores de todo el mundo a traducir los datos y el conocimiento en percepciones, lo cual les ayuda a tomar decisiones más informadas sobre la atención en cientos de hospitales y organizaciones de salud (IBM, s.f.).

Otro ejemplo de cómo la IA está siendo aplicada en el ámbito sanitario es para la detección anticipada de enfermedades, y un ejemplo es Google DeepMind (la misma herramienta que se implementó para aprender a jugar a videojuegos a través del aprendizaje automático). Así, los investigadores de la filial DeepMind de Google en Londres, la University College of London y del Moorfields Eye Hospital han utilizado el

---

<sup>9</sup> El término FLOPS (*Floating point operations per second*) significa “operaciones de coma flotante por segundo”, y es una unidad que se suele utilizar para medir los cálculos matemáticos que puede hacer por segundo una CPU y GPU (Fm, 2017).

*Deep learning*<sup>10</sup> para crear un software que identifica docenas de enfermedades oculares comunes a partir de escáneres 3D y que recomienda un tratamiento personalizado para el paciente. Y aunque el software no está listo para su uso clínico, podría ser desplegado en los hospitales en cuestión de años, ya que, tal y como expresó Mustafa Suleyman, jefe de DeepMind Health, el proyecto era “increíblemente emocionante” ya que, con el tiempo podría “*transformar el diagnóstico, el tratamiento y la gestión de los pacientes con afecciones oculares que amenazan la vista (...) en todo el mundo*” (Vincent, 2018).

El programa informático, descrito en un artículo publicado en la revista Nature Medicine, se basa en principios establecidos de *Deep learning*, que utiliza algoritmos para identificar patrones comunes en los datos. En este caso, los datos son escaneos tridimensionales de los ojos de los pacientes realizados mediante una técnica conocida como tomografía de coherencia óptica, o TCO (OCT según sus siglas en inglés). Así, el programa, del mismo modo en el que se ha expuesto anteriormente con su aplicación en los videojuegos, entrenó una red neuronal para detectar con precisión más de 50 tipos de enfermedades oculares simplemente analizando escaneos en 3D, lo cual demuestra el auténtico potencial de la herramienta y lo efectiva que puede ser la tecnología de IA para identificar anomalías reales (Mills, 2020).

Pero las aplicaciones de la IA en el ámbito sanitario no acaban ahí, ya que su implementación en la investigación y desarrollo de nuevos fármacos podría acelerar su avance de manera exponencial. El desarrollo de fármacos es una tarea tediosa que puede llevar años y miles de intentos fallidos, a parte de costar a los investigadores miles de millones de dólares en el proceso. Según estima la FDA (*U.S. Food & Drug Administration*), en promedio, se necesitan aproximadamente ocho años y medio para estudiar y probar un nuevo medicamento antes de que la agencia pueda aprobarlo para el público en general. Eso incluye las primeras pruebas de laboratorio y en animales, así como los posteriores ensayos clínicos con humanos. Y es que sólo cinco de cada 5.000 medicamentos que comienzan las pruebas preclínicas llegan a las pruebas en humanos, y sólo uno de los cinco acaba llegando a las farmacias (FDA, 2015).

---

<sup>10</sup> El Deep Learning (aprendizaje profundo) es un subconjunto del aprendizaje automático en la inteligencia artificial (IA) que tiene redes capaces de aprender sin supervisión a partir de datos no estructurados o no etiquetados (Hargrave, 2019).

Así, algunos gigantes farmacéuticos como Johnson & Johnson o Pfizer se han asociado con las empresas tecnológicas IBM y Google. La idea es construir un programa de descubrimiento de drogas usando el *Deep learning* y la IA. Por lo que, en lugar de utilizar el enfoque tradicional de ensayo y error, el descubrimiento de medicamentos se basa ahora en los datos (Rowe, 2019). Gracias a las simulaciones inteligentes, mejores curas son posibles a través del análisis de la medicina existente, los pacientes y los patógenos, y de este modo, los investigadores han podido incluso redirigir los medicamentos ya existentes para combatir nuevas infecciones. Así, gracias a la implementación de la IA y el Deep learning, un proceso que tomaba años en conseguirse, ahora toma días (Mills, 2020).

Pero el uso y la aplicación de la IA en aras de lograr una sociedad mejor, también habría de implementarse hacia los Objetivos de Desarrollo Sostenible (de aquí en adelante SDGs, según sus siglas en inglés), ya que puede acelerar su cumplimiento y éxito si se desarrolla de manera que beneficie a la humanidad, respete las normas y estándares mundiales y se base en la paz y el desarrollo.

En 2015, la Asamblea de la ONU estableció la Agenda 2030 para el Desarrollo Sostenible, en la cual presenta 17 objetivos (ver en Figura 2) “con 169 metas de carácter integrado e indivisible que abarcan las esferas económica, social y ambiental” (ONU, 2015), con el propósito de alcanzar dichos objetivos para el año 2030, haciendo especial hincapié en la erradicación de la pobreza.

Un estudio de la revista Nature muestra que la IA puede actuar facilitar el cumplimiento de 134 objetivos (79%) en todos los SDG, mediante una mejora tecnológica que permitiría superar ciertas limitaciones actuales. Sin embargo, 59 objetivos (35%, también en todos los SDG) pueden experimentar un impacto negativo por el desarrollo de la IA (Vinuesa et al., 2020)

**Figura 2: OBJETIVOS DE DESARROLLO**



Por ejemplo, en el SDG 1 sobre el fin de la pobreza, el SDG 4 sobre la educación de calidad, el SDG 6 sobre el agua limpia y el saneamiento, el SDG 7 sobre la energía asequible y no contaminante, y el SDG 11 sobre las ciudades y comunidades sostenibles, la IA puede actuar como facilitadora de todos los objetivos apoyando el suministro de alimentos, salud, agua y servicios energéticos a la población. También puede servir de base a los sistemas con bajas emisiones carbono, por ejemplo, apoyando la creación de economías circulares y ciudades inteligentes que utilicen eficientemente sus recursos. Por ejemplo, la IA puede hacer posible la creación de ciudades inteligentes y con bajas emisiones de carbono que abarquen una serie de tecnologías interconectadas, como los vehículos eléctricos autónomos y los aparatos inteligentes que pueden permitir la respuesta de la demanda en el sector de la electricidad (Nerini, Slob, Engström, et al., 2019). La IA también puede ayudar a integrar las energías renovables de forma más generalizada, al permitir redes inteligentes que ajusten parcialmente la demanda eléctrica a los momentos en los que brilla el sol y sopla el viento. Según el mencionado estudio, son más los objetivos que obtienen un resultado positivo gracias a la implantación de la IA que los que obtienen uno negativo.

Sin embargo, la tecnología de IA, tanto la investigación como el diseño de productos pueden requerir recursos computacionales enormes, los cuales sólo son viables a través de grandes centros de computación, cuyas instalaciones tienen una demanda de energía y una huella de carbono muy alta (Jones, 2018). Por ejemplo, las aplicaciones de criptografía como *Bitcoin* están utilizando globalmente tanta electricidad como la demanda eléctrica de algunos países (Truby, 2018), comprometiendo los resultados del SDG 7, sobre energía asequible y no contaminante, pero también en el SDG 13 sobre Acción por el Clima. Algunas estimaciones sugieren que la demanda total de electricidad de las tecnologías de la información y las comunicaciones (TIC) podría requerir hasta el 20% de la demanda mundial de electricidad para 2030, frente al 1% actual (Jones, 2018).

Por lo tanto, es esencial un desarrollo ecológico de la tecnología de las TIC, el cual se plasme en sistemas de refrigeración más eficientes para los centros de datos, una mayor eficiencia energética y un uso de energías renovables en las TIC, lo que contribuiría a contener el crecimiento de la demanda de electricidad (Karnama, Haghghi & Vinuesa, 2019). Además de centros de datos más eficientes y basados en energías renovables, es esencial incorporar el conocimiento humano en el desarrollo de modelos de IA. Además del hecho de que el cerebro humano consume mucho menos energía que la que se utiliza para entrenar los modelos de IA, el factor ético y humano a la hora de la toma de decisiones de las máquinas, no puede ser aportado por otro elemento distinto al humano.

Por ello, aunque la tecnología habilitada para la IA puede actuar como catalizador para lograr la Agenda 2030, también puede desencadenar desigualdades que pueden actuar como inhibidores de los SDG 1, 4 y 5 (el fin de la pobreza, la educación de calidad y la igualdad de género, respectivamente). Esta dualidad se refleja en que, por ejemplo, la IA puede ayudar a identificar las zonas de pobreza y fomentar la acción internacional utilizando imágenes satelitales (Jean, Burke, Xie, et al., 2016) pero por otra parte, también puede dar lugar a requisitos de cualificación adicionales para cualquier trabajo, aumentando así las desigualdades existentes (Nagano, 2018) y siendo un impedimento para el logro de este objetivo.

#### **4. ÉTICA DE LA INTELIGENCIA ARTIFICIAL**

### **a. La roboética**

La roboética, o ética de los robots se define como *“ética aplicada cuyo objetivo es desarrollar herramientas científicas/culturales/técnicas que puedan ser compartidas por diferentes grupos sociales y creencias. Estos instrumentos tienen por objeto promover y alentar el desarrollo de la robótica para el progreso de la sociedad y los individuos humanos, y ayudar a prevenir su uso indebido contra la humanidad”* (Veruggio, 2002).

El término, acuñado en 2002 por Gianmarco Veruggio, puede resumirse como el área de estudio que se ocupa de las reglas que deben crearse para diseñar robots que lleven a cabo un comportamiento ético. Por ello, no debemos caer en el error de pensar que la roboética es la ética de los propios robots o de cualquier ética artificial en sí, si no que se refiere a la ética que deben de aplicar en su programación y diseño tanto los propios diseñadores y fabricantes como los usuarios. El propósito de la roboética es asegurar que las máquinas con IA se comporten de manera que prioricen la seguridad humana por encima de sus tareas asignadas y su propia seguridad, y que también sean acordes con los preceptos aceptados de la moralidad humana.

Antes de la aparición de los sofisticados y potentes ordenadores modernos y del uso a gran escala del Big data, esos sistemas eran programados manualmente y seguían las reglas de la invención humana, pero los avances tecnológicos han llevado al desarrollo de nuevos enfoques. Entre ellos está el aprendizaje automático, siendo actualmente el área más activa de la IA, en el que los métodos estadísticos permiten que un sistema aprenda a partir de datos y tome decisiones, sin estar programado explícitamente.

Las preocupaciones sobre estos avances a veces se centran en el uso de la IA en aviones teledirigidos militares, o en el riesgo de que la IA pueda hacer temblar los sistemas financieros mundiales. Pero no hace falta pensar tan a gran escala, la IA ya ha introducido una preocupación por el desempleo, ya que, por ejemplo, los sistemas autónomos de conducción amenazan con reemplazar a millones de camioneros y hacer que la mano de obra en las fábricas quede obsoleta. Y más allá de estas consideraciones sociales y económicas más amplias, las principales preocupaciones de los analistas se basan en la posible parcialidad y consiguientes desviaciones en las implementaciones éticas de la tecnología, así como la naturaleza de las interacciones entre los sistemas de

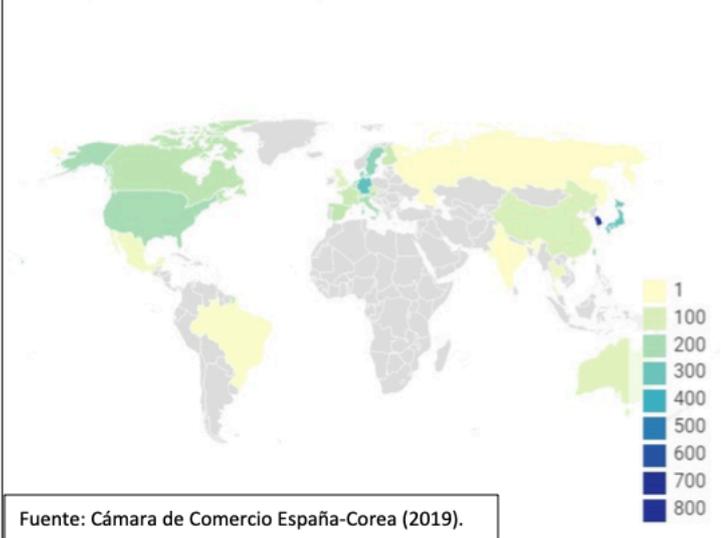
IA y los humanos, en cuanto a si estos sistemas se van a desplegar de forma adecuada y justa en aplicaciones más cotidianas.

La inteligencia artificial puede añadir y evaluar grandes cantidades de datos que están más allá de la capacidad humana, por lo que si a las máquinas se les da el poder de tomar decisiones cotidianas que alteren nuestras vidas nos podríamos encontrar con situaciones tales como que se les permitiera hacer recomendaciones a la hora de contratar, determinar en segundos la solvencia de los solicitantes de préstamos o predecir las posibilidades de que los criminales vuelvan a delinquir. El problema es que tales aplicaciones plantean problemas éticos preocupantes porque los sistemas de IA pueden aplicar lo que han aprendido de los datos del mundo real, incluso amplificando riesgos familiares, como el sesgo racial o de género. Además, los sistemas también pueden cometer errores de juicio cuando se enfrentan a escenarios desconocidos.

Por ejemplo, en 2014, Amazon desarrolló una herramienta de reclutamiento para identificar ingenieros de software a los que contratar. El problema surgió cuando el sistema comenzó a discriminar rápidamente a las mujeres, teniendo la compañía que abandonarlo en 2017. Otro ejemplo claro es Tay, una inteligencia artificial desarrollada por Microsoft en 2016 para mantener conversaciones personalizadas con usuarios de redes sociales, y para ello recababa información sobre ellos y del conjunto de la red. El problema surgió cuando en menos de un día de actividad en Twitter, Microsoft tuvo que desconectarla, debido a que comenzó a publicar tweets con posiciones nazis, racistas y homófobas. Por otro lado, en los últimos años los vehículos autónomos han provocado accidentes mortales cuando se enfrentaron a situaciones y estímulos desconocidos para ellos, que sus sistemas de guía no pudieron interpretar.

**Figura 3.**

Robots por cada 10.000 trabajadores en 2017



Por todos estos problemas que están comenzando a presentarse a medida que se van desarrollando más las inteligencias artificiales, Corea del Sur, el país con más robots por cada 10.000 trabajadores (Cámara de Comercio España-Corea, 2019), ya en el año 2007 comenzó a plantear la creación de un Código Ético para los robots, el cual

surgió con la intención de determinar “directrices éticas sobre el papel y las funciones de los robots”, ya que los robots tienen el potencial de “desarrollar una inteligencia aguda” (Le Monde, 2007), fundándose en la importancia del mercado de la robótica en la economía surcoreana, líder mundial en el número de robots en el mercado laboral. Aunque nunca se ha hecho público, este texto es el primero que se plantea una ley sobre los robots a través del prisma de la ética, y en particular los casos en que se trata de relaciones entre los seres humanos y los robots capaces de tomar decisiones, y por lo tanto de poseer inteligencia artificial.

Otro estado interesado en el tema de los robots es el Reino Unido. En abril de 2016, la agencia británica British Standards Institution (Instituto Británico de Estándares) publicó un documento titulado “*Robots y dispositivos robóticos: Guía para el diseño y aplicación ética de los robots y sistemas robóticos*”<sup>11</sup>. Este texto, aunque carece de cualquier valor jurídico, establece recomendaciones para los creadores de robots, haciendo hincapié en los riesgos éticos asociados al desarrollo de los robots como, por ejemplo, el peligro de que puedan recurrir a comportamientos racistas. Se trata así de evitar una “falta de respeto por la diversidad cultural o el pluralismo” (Devlin, 2016).

Por otra parte, en lo que respecta a la responsabilidad por los robots, los autores de este texto adoptan una postura para que ésta recaiga plenamente en los seres humanos,

<sup>11</sup> “Robots and robotic devices. Guide to the ethical design and application of robots and robotic systems”.

insistiendo en que la persona responsable debe ser claramente identificable. Siendo este texto a día de hoy, uno de los más actualizados en lo que respecta a la ética de los robots.

De todos modos, no es el primer intento realizado por Reino Unido, ya que, en septiembre de 2010 ya se había organizado un grupo de trabajo, llamado “Principios de la robótica”, compuesto por varios académicos y especialistas de renombre reunidos en torno a la idea de debatir sobre la robótica y, en particular, sobre su aplicación en el mundo real (Engineering and Physical Sciences Research Council, s.f.). Sin embargo, este proyecto se pronuncia a costa del reconocimiento de los robots como entidades legales.

En lo que respecta al ámbito europeo, la diputada luxemburguesa Mady Delvaux propuso un informe al Parlamento, el cual lo aprobó en 2017 por amplia mayoría, en el que se pide a la Comisión Europea que trabaje en el establecimiento de normas éticas en materia de robótica e inteligencia artificial “con el fin de explotar todo su potencial económico y garantizar un nivel estándar de seguridad”.

Entre los temas sobre los que el informe pedía a la Comisión que se interesara se encontraba inicialmente la concesión de una personalidad electrónica a los robots, es decir *“crear a largo plazo una personalidad jurídica específica para los robots, de forma que como mínimo los robots autónomos más complejos puedan ser considerados personas electrónicas responsables de reparar los daños que puedan causar, y posiblemente aplicar la personalidad electrónica a aquellos supuestos en los que los robots tomen decisiones autónomas inteligentes o interactúen con terceros de forma independiente”*. Según Mady Delvaux, *“sería el mismo principio que tenemos actualmente para las empresas. Sin embargo, esta solución tardaría un tiempo en hacerse realidad”* (Parlamento Europeo, 2017). Según el informe, una persona electrónica sería potencialmente *“cualquier robot que toma decisiones autónomas de forma inteligente o que interactúa de forma independiente con terceros”*.

Sin embargo, la Comisión aún no ha adoptado esta noción de personalidad electrónica. Además, el Comité Económico y Social Europeo (CESE) se ha opuesto al reconocimiento de cualquier forma de personalidad jurídica para los robots. En un dictamen del 4 de mayo de 2017, consideró que esto sería una *“amenaza para una*

democracia abierta y equitativa”. Sin embargo, el CESE sólo puede emitir dictámenes sin valor jurídico.

Del mismo modo, en abril de 2018, más de 200 expertos procedentes de catorce países de la Unión Europea firmaron una carta abierta dirigida al presidente de la Comisión, Jean-Claude Juncker, en la cual expresaban su rechazo a las peticiones formuladas por el Parlamento a la Comisión, en base a las cuales, la directiva que prepara la Comisión reconocería la “personalidad” de las máquinas. Según los firmantes, el hecho de otorgar el estatus de “personas electrónicas” a las máquinas conllevaría intrínsecamente el reconocimiento de derechos y responsabilidades, lo cual acarrearía una serie de impactos de carácter ético, legal, económico y social (Rius, 2018).

En general, los responsables europeos se deben apresurar a tratar de construir una base de normas sobre los robots que sea aplicable a los Estados miembros, antes de que sean éstos los que empiecen a legislar por su cuenta. Esto plantearía problemas de armonía y, al mismo tiempo, socavaría la estabilidad del marco jurídico europeo, especialmente si los Estados adoptan normas que difieren o incluso se contradicen entre sí de un Estado a otro. Además, también es importante unificar las futuras normas de robótica en vista de la fuerte competencia de otros sectores geográficos que ya están a la vanguardia en este campo, como Asia o los Estados Unidos.

Es cierto que, aunque todavía muy por detrás de Corea del Sur, algunos Estados europeos están considerando el posible reconocimiento de la personalidad jurídica de los robots. Un caso es el de Francia, donde se han hecho algunos esfuerzos para responder a este tipo de preguntas, como por ejemplo el informe adoptado el 29 de marzo de 2017 por la Oficina Parlamentaria de Evaluación de las Elecciones Científicas y Tecnológicas (OPESCT).

Otro documento en el que podemos encontrar expresión de actitud sobre la personalidad jurídica en Europa es la Declaración sobre “*Inteligencia artificial, robótica y sistemas 'autónomos'*” elaborada por el Grupo Europeo sobre Ética en la Ciencia y las Nuevas Tecnologías, de marzo de 2018. Este manuscrito no puede considerarse de carácter político, ya que no expresa una posición oficial de la Unión, pero sin embargo, el personal de los servicios de la Comisión Europea participó en la preparación de la

declaración, por lo que podría considerarse al menos como la opinión experta más cercana a la opinión oficial.

El texto de la mencionada publicación deja claro que sólo una persona puede ser sujeto de responsabilidad y derechos y no puede trasladarla a máquinas autónomas. Tal y como proclama el texto, *“la capacidad y la voluntad de asumir y atribuir la responsabilidad moral es parte integrante de la concepción de la persona en la que se basan todas nuestras instituciones morales, sociales y jurídicas...La responsabilidad moral, en cualquier sentido, no puede asignarse ni trasladarse a la tecnología 'autónoma’”* (European Group on Ethics in Science and New Technologies, 2018). Implica que la responsabilidad moral no puede ser característica de la IA, y esta es una propiedad necesaria de la persona moral y jurídica, por lo que la IA no podría ser considerada como persona jurídica.

En el caso de los sistemas de armas autónomas letales (LAWS) la opinión de los expertos es aún más obvia, ya que, tal y como se proclama en la declaración *“los humanos -y no los ordenadores y sus algoritmos- deben en última instancia mantener el control, y por lo tanto ser moralmente responsables”*. No está clara cuál es la posición oficial de la UE sobre la posible personalidad jurídica de la IA de las fuentes mencionadas. ¿Está la Unión Europea a favor o en contra de la personalidad jurídica de la inteligencia artificial? Se podría hacer al menos una conclusión. La UE considera la personalidad jurídica de la inteligencia artificial sólo en la esfera del derecho civil. De acuerdo con el análisis de los documentos mencionados, la inteligencia artificial podría ser más bien un agente independiente de las transacciones comerciales que un culpable de un delito.

Por su parte, en los Estados Unidos, los robots desempeñan una función dominante en la economía, en particular en el mercado laboral, hasta tal punto que ya constituyen una amenaza para el empleo, estimándose que para 2030 hasta un tercio de la fuerza laboral del país tendrá que adaptarse a este cambio especializándose o cambiando de trabajo, siendo los empleos poco cualificados los que más perjudicados se verán en este aspecto (Della Coletta, 2017). Sin embargo, no existe ninguna ley estadounidense relativa a la condición jurídica de los robots. Probablemente sea sólo cuestión de tiempo, ya que la Administración Nacional de Seguridad del Tráfico en las Carreteras (NHTSA) ha

propuesto que Google conceda la condición de conductor a sus coches autónomos, que por lo tanto serán responsables en lugar de los humanos que transportan (Reuters, 2016).

Otro documento relevante en este aspecto es el Manual de Derecho de la Guerra del Departamento de Defensa de los Estados Unidos<sup>12</sup>. En la disposición 6.5.9.3 “Las obligaciones del derecho de la guerra de distinción y proporcionalidad se aplican a las personas en lugar de a las armas mismas” se puede encontrar una opinión sobre la personalidad jurídica de las máquinas de inteligencia artificial utilizadas en la guerra. El artículo del manual establece que *“el derecho de la guerra no exige que las armas tomen determinaciones jurídicas, aunque el arma (por ejemplo, mediante ordenadores, programas informáticos y sensores) pueda caracterizarse como capaz de tomar determinaciones fácticas, como la de disparar el arma o la de seleccionar y atacar un objetivo”*. Cabe destacar que el Manual no es más que una recomendación oficial sobre la forma en que los oficiales militares deben interpretar el derecho internacional humanitario<sup>13</sup> y como toda interpretación del derecho internacional, puede evolucionar con el tiempo y, por lo tanto, cuando el manual establece que *“no exige”* esto no implica que en un futuro no pueda darse o que tenga que ser prohibido.

En el informe publicado en 2019 por el Comité de Tecnología del NSTC<sup>14</sup> llamado “Plan estratégico nacional de investigación y desarrollo de la inteligencia artificial”<sup>15</sup>, se definen medidas clave para maximizar los beneficios de la tecnología de la inteligencia artificial. Entre otras cosas sugeridas en el informe hay una propuesta para desarrollar robots que podrían seguir las normas éticas y legales. Como se establece en la subdivisión “Construcción de la IA ética” del informe en la “Estrategia 3: Entender y abordar las consecuencias éticas, jurídicas y sociales”, *“dentro de los límites de lo que es tecnológicamente viable, por lo tanto, los investigadores deben esforzarse por desarrollar algoritmos y arquitecturas que sean verificablemente coherentes con las leyes, las normas sociales y la ética existentes, o que se ajusten a ellas, lo que sin duda*

---

<sup>12</sup> Department of Defense: Law of war Manual (Updated 2016).

<sup>13</sup> *“Este manual refleja muchos años de trabajo y experiencia, de parte de abogados civiles y militares de cada Servicio Militar. Refleja la experiencia de este Departamento en la aplicación del derecho de la guerra en operaciones militares reales, y nos ayudará a recordar las duras lecciones del pasado. La comprensión de nuestros deberes impuestos por el derecho de la guerra y nuestros derechos bajo el mismo es esencial para nuestro servicio en la defensa de la nación”*.

<sup>14</sup> National Science and Technology Council.

<sup>15</sup> The National Artificial Intelligence Research And Development Strategic Plan: 2019 Update.

*constituye una tarea muy difícil*". Esto no significa que el texto del informe proponga reconocer la personalidad jurídica de la IA, pero al menos pretende conseguir una máquina autónoma que sea capaz de seguir las normas éticas y jurídicas, es decir, de ejercer los derechos y obligaciones morales y legales (NSTC, 2019).

Las prácticas de la Unión Europea y los Estados Unidos indican que ya se han desarrollado enfoques comunes de la personalidad jurídica de algunos tipos de inteligencia artificial. Ambos países están firmemente en contra de la personalidad jurídica de las máquinas de guerra inteligentes, ya que la responsabilidad por cualquier daño o mala conducta de la IA militar sigue recayendo en los oficiales militares. En caso de aplicación civil de la IA hay dos opciones: la IA podría actuar como persona jurídica o como agente de relaciones comerciales con otras personas jurídicas. Ambos permiten la posibilidad de que la IA se convierta en un sujeto de pleno derecho en las relaciones de derecho civil, lo cual implica que la máquina autónoma será capaz de concluir tratos, vender algo y prestar servicios por sí misma. Pero parece que en el caso del derecho penal este escenario es imposible incluso en el futuro.

#### **b. Principios éticos para una digitalización humanista.**

A lo largo de este trabajo hemos podido ver muchas de las posibles implementaciones de la IA, algunas ya en práctica desde hace tiempo, pero muchas otras están en sus comienzos o ni siquiera están disponibles aún. Así, se han expuesto muchas de las ventajas que esta implementación acarrearía, mejorando nuestras vidas en numerosos aspectos, pero también se ha mostrado cual es la cara más oscura de la aplicación indebida de esta tecnología. El peligro del desarrollo de la IA es que tradicionalmente se basan en las necesidades y valores de las naciones en las que se desarrolla la IA, por lo que, como ya se vio en un apartado anterior, si la IA se utilizan en regiones en las que faltan el escrutinio ético, la transparencia y el control democrático, la IA podría permitir el nacionalismo, el odio hacia las minorías y el sesgo de los resultados electorales. Es por ello que es necesario que, en aras de evitar una digitalización que derive en efectos perniciosos para el ser humano, ha de ser la ética la que sirva como marco de referencia en el que basarse para la implementación de la IA y su futura regulación en los sistemas legislativos de los estados.

Debido a que abarcar todos los principios éticos aplicables a cada una de las implementaciones de la IA sería muy complicado, en esta sección del trabajo nos centraremos en los principios éticos que, en base a las cuestiones tratadas durante el trabajo y a la opinión del autor, son los que habrían de servir de base para el desarrollo de estas tecnologías y para su futura regulación. Ha de garantizarse que el desarrollo de la IA tenga como base fundamental y como objetivo último hacer el bien a los humanos, es decir, ha de poder contarse con una IA confiable, que responda a un fin ético, buscando el bien común. Por ello, los principios éticos en base a los cuales debería implementarse la IA, son el de justicia, los de autonomía y transparencia y el de beneficencia (hacer el bien).

En primer lugar, el principio de justicia lo podríamos definir como la obligación moral de actuar sobre la base de un juicio justo entre demandas en conflicto. Es un principio profundamente vinculado a la equidad, al derecho y a la igualdad. Por ello, desde el prisma de la justicia, una acción será ética en tanto en cuanto sea equitativa para las partes. Es decir, algo será justo en tanto se de a las partes en conflicto lo que se merecen en base a sus condiciones o sus méritos. Por lo tanto, el principio de justicia es, por definición, contrario a la discriminación.

Como hemos visto, la implementación de la IA en la economía traerá numerosos avances, los cuales podrán traducirse en un brutal incremento de los beneficios empresariales, pero por otro lado, también podrán traer consigo un incremento del desempleo para aquellos trabajadores poco cualificados y un aumento de las nóminas de los más cualificados y trabajadores de puestos de dirección. Del mismo modo, la accesibilidad a la tecnología con IA está distribuida de manera desigual, ya que, por ejemplo, es posible que un complejo equipo agrícola mejorado con IA no sea accesible para los pequeños agricultores y, por lo tanto, produzca un aumento de la brecha con respecto a los grandes productores de las economías más desarrolladas.

Por esta razón, la implementación de la IA debería hacerse desde la perspectiva ética del principio de justicia, es decir, que la IA debe ser un mecanismo cuya implementación estuviera encaminada a la reducción de las desigualdades e injusticias sociales, no uno encaminado a su potenciación. La IA, por lo tanto, ha de ser una herramienta que actúe como motor para lograr el fin de las desigualdades económicas, ya que, en caso de no

implementarse en base al principio de justicia, se acentuaría la polarización de la riqueza, y no sólo entre los países desarrollados y los subdesarrollados o en vías de desarrollo, sino que se podría crear una brecha más grande dentro de las sociedades de los países desarrollados, aumentando la riqueza de las clases altas, y aumentando la pobreza de las clases menos pudientes, siendo la clase media la más perjudicada en este proceso de polarización. En otras palabras, el desarrollo de la IA ha de realizarse de manera equitativa, no de manera igualitaria, ya que no todos los individuos tienen la misma facilidad de acceso a la IA y no todas las regiones y estados tienen las mismas necesidades.

Esta lucha contra la desigualdad social, no solo se debe centrar en lo económico, ya que, como hemos visto, la posible implementación de la IA en sectores como el sanitario o educativo, no sólo supondría un desarrollo sin precedentes en ellos, sino que también supondría un incremento de la accesibilidad para aquellos que a día de hoy ven limitadas sus posibilidades debido a que tienen una limitación de recursos. Un desarrollo de la IA basado en el principio de justicia, tendría el potencial de ‘democratizar’ la sanidad o la educación, pudiendo hacerla accesible a un número mayor de individuos e incluso llegar a hacerla universal. Por ello, en lo relativo a aspectos sociales como la sanidad o la educación, no debe de primar la equidad, sino que en este caso, el desarrollo e implementación de la IA debería basarse en el igualitarismo, es decir, que cada ciudadano reciba los mismos recursos para partir de una igualdad de condiciones, y que nadie parte de una posición privilegiada.

Del mismo modo, el principio de justicia no sólo tiene como fin último la justicia social, sino que hay que añadirle la perspectiva medioambiental. Un desarrollo sostenible es un desarrollo justo. Por ello, la IA no sólo debe tener como fin disminuir las desigualdades entre los seres humanos, sino que se además debe centrarse en un desarrollo sostenible medioambientalmente, ya que éste es necesario para crear un futuro justo para las generaciones venideras, porque, en el caso de obviarse la perspectiva medioambiental, las perspectivas de futuro no parecen ser demasiado halagüeñas. El futuro ha de ser sostenible, o no será.

Por otro lado, como hemos podido ver, la IA se ha utilizado recientemente tanto para influir en los sistemas democráticos, reduciendo la capacidad de decisión del individuo

como para desarrollar un control ciudadano más exhaustivo, potenciando el control del comportamiento social. Este control e influencia es un claro ejemplo de amenaza a los derechos humanos debido al mal uso de la IA y uno de sus mayores problemas es la falta de información que reciben los ciudadanos sobre el tipo de datos analizados y las consecuencias que esto puede tener en sus vidas. Por ello, es necesario que la IA se implemente desde el principio ético de la autonomía y en base a la transparencia.

El principio de autonomía lo podemos definir como la obligación moral de respetar las decisiones que tomen los individuos respecto a sus propias vidas, en base a sus propias convicciones e intereses. Nos da un deber negativo de no interferir en las decisiones de adultos competentes, y un deber positivo de dar poder a otros de los que somos responsables. Por lo tanto, este principio está muy relacionado con el de justicia, ya que en el caso de personas que carezcan de autonomía plena, es nuestro deber darles protección, por lo que la equidad es un concepto básico también para este principio.

La aplicación de la IA ha de desarrollarse en base a la creación de ciudadanos libres y conscientes, que puedan tomar decisiones autónomas sin la interferencia de un estado controlador y manipulador. Por lo tanto, es necesario que el desarrollo de la IA no se haga en aras de una mayor manipulación y control del ciudadano, y para que esto sea así, es necesario que el individuo tenga a su disposición la mayor cantidad de información posible, la cual le permita ejercitar sus derechos de forma autónoma y consciente, sabiendo cuáles son las consecuencias de sus decisiones. Así, teniendo a disposición una gran cantidad de información, el individuo tendría que hacer uso de ella y llevar a cabo un proceso de razonamiento e interiorización de dicha información, para ser capaz de realizar sus acciones en base a un conocimiento adquirido a través de la razón y del pensamiento crítico, haciéndole plenamente libre y autónomo a la hora de tomar decisiones. Por ello, debe desarrollarse una IA al servicio de la veracidad, una IA que promueva la transparencia de información en detrimento de la censura y de la manipulación informativa. Un ciudadano con información veraz a su disposición es un ciudadano libre. Por lo tanto, en base a que un desarrollo ético de la IA debe de tener como fin último el bien al hombre, es necesario que el principio de autonomía y la transparencia sean unas de las bases a partir de las cuales se implemente y desarrolle la IA, potenciando de esta manera la soberanía del ciudadano frente a la tiranía de estados totalitarios.

Por último, la implementación de la IA en nuestra sociedad capitalista, ha hecho que, a través de los potentes motores de recomendación de las plataformas on-line, hayamos experimentado un incremento del consumismo. Este consumismo, nos crea la falsa ilusión de que a través de la adquisición de bienes materiales, alcanzaremos la felicidad, es decir, asociamos la felicidad con el bienestar material, siendo la realidad una muy distinta, ya que nunca estamos satisfechos con los bienes materiales que adquirimos, no alcanzando nunca nuestra realización personal ni la felicidad.

Es por ello que el tercer principio ético en el que se debería basar el desarrollo de la IA es el de beneficencia. El principio beneficencia se puede definir como la virtud de hacer el bien (Sánchez-Caro, 2012), es decir, este principio se fundamenta en que cada actuación que realice un individuo ha de estar encaminada a beneficiar a los demás. Así, este principio se ve complementado con otro, el de no maleficencia, el cual se podría definir como el no hacer daño. Así, el filósofo moral William Frankena “actualiza” el principio de beneficencia en cuatro puntos, los cuales son: 1. No infligir mal o daño; 2. Prevenir el daño; 3. Eliminar el daño y 4. Promover el bien (Costa, A. M., & Almendro, C., 2009). Pero, teniendo en cuenta que una de las tres leyes de la robótica planteadas por Isaac Asimov es que “un robot no puede dañar al ser humano ni, por inacción, permitir que un ser humano sufra daño”, el principio de beneficencia debería ser intrínseco a la IA, es decir, no debería plantearse siquiera el debate de si el desarrollo de la IA debe basarse en este principio o no. Por ello, partiendo de la base de que el principio de beneficencia ha de ser necesariamente una de bases del desarrollo de la IA, lo que debería promover la IA es ayudar a que el principio de beneficencia se implante en la forma en la que interactuamos como sociedad, a través del prisma del eudemonismo aristotélico. Así, nuestra sociedad debería actuar en base al principio de beneficencia, es decir, hacer el bien, pero a través de la perspectiva eudemonista.

Por otro lado, el Eudemonismo aristotélico establece que todo ser natural actúa en base a lo que le es propio, es decir, que sus actuaciones se basen en lo esencial del ser en sí mismo y lo distinguen del resto de seres. Así, cada ser ha que actuar en base a lo que le es propio, lo que hace que se pueda distinguir a ese ser del resto. Por ello, si un ser realiza sus actuaciones en base a lo que le es propio por naturaleza, el fin hacia el que tiende cada ser a través de sus actuaciones es el “bien”. En el caso del ser humano, el bien ha de ser

aquello que le es propio, es decir, el bien consiste en aquello que le haga ser “ser humano”. Por lo tanto, teniendo en cuenta que lo que nos distingue del resto de seres naturales es la razón, el bien supremo perseguido por el ser humano ha de ser la racionalidad. Por ello, la excelencia del ser humano, su virtud, se basa en actuar en base a la razón.

De este modo, si lo propio del ser humano es actuar en base a la razón, siendo esto lo que le corresponde hacer en base a su esencia, alcanzará de este modo la “felicidad” (*eudaimonía* en griego), ya que es el fin último del ser humano. Por ello, el ser humano será feliz en tanto en cuanto realice lo que le es propio como ser humano, es decir, el bien supremo del ser humano es que su vida se base en la racionalidad. “*El bien autosuficiente es aquel que por sí solo torna amable la vida, y tal bien es la felicidad*” (Cfr. Ar. Eth. Nic.1097a 15-20).

Dado que el ser humano necesita también satisfacer las demandas corporales y las pasiones del alma, para que pueda vivir en base a la racionalidad, es necesario que actúe en base a la búsqueda del “término medio”, es decir, aprender a gestionar nuestros deseos y pasiones. Por lo tanto, es necesario encontrar un equilibrio, el término medio entre el exceso y el defecto. Así, “*entre la cobardía y la temeridad, hemos de actuar con valentía; frente al despilfarro y la tacañería, hemos de hacerlo con generosidad; frente a la desvergüenza y la timidez, con modestia; frente a la adulación y la mezquindad, con gentileza; etc*” (RecursosTic, s.f.). De este modo, Aristóteles considera que la “virtud” es vivir en base al “hábito” de actuar según este equilibrio o “término medio”. Por lo tanto, se puede decir que un hombre es virtuoso si voluntariamente ha adquirido el hábito de actuar en base a la búsqueda del “término medio”, y así, Aristóteles cree que el ser humano alcanzará la felicidad perfecta y plena en base a un ejercicio constante y prolongado de esta virtud, que es el fin último al que aspira llegar el ser humano.

Por ello, para que el ser humano consiga alcanzar la felicidad, es necesario que la implementación y desarrollo de la IA se haga favoreciendo la promoción del principio de beneficencia a través de una aproximación eudemonista, que contribuya al bienestar y florecimiento humanos, no sólo en las actuaciones individuales, sino en la sociedad contemporánea en su conjunto. La implementación y desarrollo de la IA en la sociedad deberían basarse en hacer el bien a los demás a través de un proceso racional que nos haga actuar en base a un término medio, a través de la virtud. De este modo, a través de

una IA que promueva una sociedad que se rija por este principio, no sólo debe favorecer que los individuos busquen el bien en sus actuaciones individuales, sino que favorezca a la búsqueda del bien común de la humanidad en su conjunto, favoreciendo la cooperación entre seres humanos en aras de buscar el bien supremo. La búsqueda del bien común ha de servir como potenciadora de la cooperación, reduciendo el individualismo de la sociedad contemporánea, para poder así luchar por los problemas a los que nos enfrentamos como Humanidad, contribuyendo al nuestro bienestar ya que, *“el hombre feliz, vive bien y obra bien”*.

## **5. CONCLUSIONES**

El creciente desarrollo de la IA y su implementación en la sociedad contemporánea han traído consigo un nuevo mundo de posibilidades para la Humanidad, pero del mismo modo, también han introducido una serie de amenazas y riesgos para los que, en muchos casos, no estamos preparados.

Como hemos podido ver, debido a la capacidad de los medios actuales, la concepción de una máquina con una inteligencia equiparable a la del ser humano, una IA fuerte, a día de hoy queda relegada a los relatos de ciencia ficción, siendo la IA específica la que predomina en nuestro mundo, ayudándonos a realizar las tareas más mundanas y monótonas, pero también demostrándonos su enorme poder. Por ello, no podemos considerar que el desarrollo de la IA esté cerca de igualar la capacidad intelectual de un cerebro humano, ya que, a pesar de aprender automáticamente en base a la experiencia y de tomar decisiones propias, la IA actual se caracteriza por la realización de tareas específicas, previamente establecidas. Pero, del mismo modo, tampoco ha de subestimarse su capacidad, ya que su poder de procesamiento de datos ha superado y reemplazado, en muchas de sus aplicaciones, la capacidad de los seres humanos.

Pese a los límites derivados de las dificultades que presenta este fenómeno por la alta implicación de conceptos metafísicos, como la inteligencia o el bien común, y debido a que una discusión de los mismos no es el objeto de este trabajo, se ha realizado un análisis de la implicación de la IA en diversos ámbitos.

El impacto de aplicación de la IA en la economía mundial aún es una incógnita, ya que, a pesar de que se pronostican numerosos cambios, no se puede prever a ciencia cierta qué consecuencias acarreará. Gracias a la IA, la ciencia de los datos ha adquirido una gran importancia, pues a través del empleo del Big Data, empresas como Amazon pueden realizar operaciones mucho más precisas, aumentando la productividad y la eficiencia empresarial, e influyendo en las tendencias de los consumidores.

Estos cambios en los mercados pueden acarrear una pérdida masiva de empleos, pero, del mismo modo, también se prevé un crecimiento en el número de empleos gracias a la IA. Por ello, estas dos posibles tendencias en balance general de empleos demuestran que, para evitar un potencial desempleo masivo, la solución pasaría por la potenciación de una mayor formación de los individuos, pues la baja cualificación de los trabajadores es el factor principal de riesgo para sus empleos. Esta mayor educación y formación no sólo dará una mayor libertad a los trabajadores, sino que ayudará también a que la implantación de la IA en el mercado laboral no favorezca a un aumento de la desigualdad, evitándose así una potencial polarización de la riqueza que afecte gravemente a las clases bajas y medias.

Del mismo modo, en el ámbito político, el caso de Cambridge Analytica ya ha puesto de manifiesto cuáles son algunas de las debilidades de las democracias actuales debido al desarrollo de la IA. El uso del Big Data en este ámbito también ha experimentado una tendencia de uso creciente, pues gracias a él se pueden realizar perfiles muy precisos de los votantes, influyendo de esta forma en su intención de voto.

Estas actuaciones demuestran los potenciales peligros de la IA en las sociedades actuales, pudiendo debilitarlas hasta tal punto de llevarnos a una tendencia totalitaria. Por ello, aparte de una mayor protección de los datos personales de los usuarios de internet, ha de fomentarse una mayor conciencia social respecto de cuáles son los usos que se les dan a sus datos, para que, desde el conocimiento, puedan decidir si los aceptan o no. Del mismo modo, una formación de los futuros profesionales a través de la combinación de títulos universitarios y de un estudio más profundo de la ética, permitirá que el profesional tenga una perspectiva más amplia y ética a la hora de realizar su trabajo.

La implementación de la IA en nuestra sociedad sin ningún criterio ético que le sirva de base, ha derivado en sociedades que están experimentando una limitación en sus libertades. El control de la sociedad a través de sistemas de reconocimiento facial o de seguimiento de los ciudadanos implementados en China, han reflejado los peligros y dilemas éticos que acarrea la IA en nuestras sociedades, ya que un uso indiscriminado de esta tecnología puede destruir la libertad individual del ciudadano, e incluso, llevarnos a episodios de discriminación de minorías, como la Uigur. Por lo que, la necesidad de regulación o al menos de un marco ético que sirva de base para el desarrollo de esta tecnología es primordial, ya que la ausencia de límites éticos puede llevarnos a una reducción alarmante de libertades y un aumento de la discriminación.

Por otro lado, la IA también muestra un enorme potencial en cuanto a mejoras sociales. La implantación de la IA en la educación podría llevarnos a sistemas educativos más accesibles y eficientes, y el sector de la salud vería cómo la IA ayudaría a mejorar los sistemas sanitarios. Así, la implementación de esta tecnología puede favorecer el cumplimiento de los Objetivos de Desarrollo Sostenible, ayudando a disminuir la pobreza y el hambre, aportar educación de calidad, o favorecer a crear ciudades con un suministro de energía sostenible, entre otros.

Ante estas circunstancias, es necesario que se trabaje en el desarrollo de un marco ético que sirva de base para el desarrollo de esta tecnología. Los avances que se han producido hasta ahora en este aspecto son insuficientes, ya que, a pesar de que en ciertos países ya se está empezando a plantear códigos éticos o normas que sirvan de marco para un desarrollo ético de esta tecnología, todavía se está debatiendo si los robots con capacidad de tomar decisiones por sí mismos pueden ser sujetos de derecho y tener una personalidad jurídica propia.

Es necesario el desarrollo de normas que establezcan un marco ético, no sólo en el ámbito europeo, sino mundial, fijando una base en la que las legislaciones estatales puedan empezar a desarrollarse siguiendo unos principios comunes, que, a través del prisma de la ética, permitan que un desarrollo e implantación de la IA humanista. El desarrollo de estas normas desde instituciones internacionales como la ONU, serviría para que, estados que no responden a límites legislativos o éticos, realizaran un correcto desarrollo de esta tecnología, evitando usos que vayan en contra de los intereses

generales. Por lo tanto, es necesario el establecimiento de unos principios éticos que sirvan de base en el desarrollo de la IA en aras de una digitalización humanista.

Una IA basada en el principio de justicia se desarrollará una sociedad más igualitaria, buscando una reducción de las desigualdades, fomentando la equidad en las económicas y la igualdad en las sociales, y promoviendo a su vez un desarrollo sostenible de la sociedad. Del mismo modo, una IA fundamentada en el principio de autonomía construirá una sociedad libre, ya que fomentará la información y la transparencia al ciudadano, lo que le permitirá tomar decisiones más libres y autónomas, evitando su manipulación y aumentando su soberanía frente a posibles aproximaciones totalitarias. Por último, dado que el principio de beneficencia es intrínseco a la IA, ésta habrá de desarrollarse en base la fomentación de este principio a través aproximación eudemonista a la forma en la que interactuamos como sociedad. De esta manera, el principio de beneficencia ha de contribuir al bienestar y florecimiento humanos, no sólo en las actuaciones individuales, sino en la sociedad en su conjunto, y de esta forma conseguir una la búsqueda del bien común, cooperando frente a los problemas a los que nos enfrentamos como Humanidad.

La IA tiene el potencial de traernos un nuevo mundo de posibilidades que mejoren y faciliten nuestras vidas, pero para ello, su desarrollo ha de tener a la ética como una de sus bases fundamentales, pues, en el caso contrario, no parece que el futuro vaya a ser demasiado esperanzador.

## 6. BIBLIOGRAFÍA

- Deng, B. (2015). *Machine ethics: The robot's dilemma*. Nature. Recuperado de: <https://www.nature.com/news/machine-ethics-the-robot-s-dilemma-1.17881>
- Bostrom, N. & Yudkowsky, E. (s.f.) *The Ethics of Artificial Intelligence*. Machine Intelligence Research Institute. Recuperado de: <https://intelligence.org/files/EthicsofAI.pdf>
- Conitzer, V; Sinnott-Armstrong, W; Schaich Borg, J; Deng, Y. & Kramer, M. (s.f.). *Moral Decision Making Frameworks for Artificial Intelligence*. Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. Pp. 4831-4835.
- Chanthadavong, A. (2019). *AI and ethics: The debate that needs to be had*. ZDNet. Recuperado de: <https://www.zdnet.com/article/ai-and-ethics-the-debate-that-needs-to-be-had/>
- Duan, Y; Edwards, J.S. & Dwivedi, Y. (2019). *Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda*. Science Direct. Recuperado de: <https://www.sciencedirect.com/science/article/pii/S0268401219300581>
- Sincavage, D. (2019). *How Artificial Intelligence Will Change Decision-Making For Businesses*. Tenfold. Recuperado de: <https://www.tenfold.com/business/artificial-intelligence-business-decisions>
- Araya, D. (2019). *Artificial Intelligence And The End Of Government*. Forbes. Recuperado de: <https://www.forbes.com/sites/danielaraya/2019/01/04/artificial-intelligence-and-the-end-of-government/>
- Samuel, S. (2019). *A quarter of Europeans want AI to replace politicians. That's a terrible idea*. Vox. Recuperado de: <https://www.vox.com/future-perfect/2019/3/27/18283992/ai-replace-politicians-europe-survey>

- Reuters (2014). *Amazon ditched AI recruiting tool that favored men for technical jobs.* Recuperado de: <https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine>
- Chanthadavong, A. (16 de septiembre de 2019). *AI and ethics: The debate that needs to be had.* ZDNet. Recuperado de: <https://www.zdnet.com/article/ai-and-ethics-the-debate-that-needs-to-be-had/>
- Shaw, J. (2019). *Artificial Intelligence and Ethics.* Harvard Magazine. Recuperado de: <https://harvardmagazine.com/2019/01/artificial-intelligence-limitations>
- Rochat, P. (2013). *Self-Consciousness and the Origins of an Ethical Stance.* American Psychological Association. Recuperado de: <http://www.psychology.emory.edu/cognition/rochat/lab/Self-consciousness%20and%20the%20Origins%20of%20the%20Ethical%20Stance.pdf>
- Lin, P., Abney, K. & Bekey, G. A. (2011). *Robot Ethics: The Ethical and Social Implications of Robotics.* MIT Press. Recuperado de: <https://books.google.es/books?hl=es&lr=&id=oBb-lt3l4oYC&oi=fnd&pg=PR5&dq=robot+ethics&ots=yviKB8x1Xl&sig=YB3u-0ImplEko51NMfC-ResEH60#v=onepage&q=robot%20ethics&f=false>
- Cossins, D. (12 de abril de 2018). *Discriminating algorithms: 5 times AI showed prejudice.* New Scientist. Recuperado de: <https://www.newscientist.com/article/2166207-discriminating-algorithms-5-times-ai-showed-prejudice/>
- La Vanguardia. (9 de marzo de 2007). *Corea del Sur adoptará este año el primer código ético para robots del mundo.* Recuperado de:

- <https://www.lavanguardia.com/tecnologia/20070309/51312648066/corea-del-sur-adoptara-este-ano-el-primer-codigo-etico-para-robots-del-mundo.html>
- Cámara de Comercio España-Corea. (s.f.). *Corea es el país del mundo con más robots por cada 10.000 trabajadores*. Noticias de la Cámara. Recuperado de: <https://www.camaracomercioespanacorea.es/es/comunicacion/noticias/837-corea-es-el-pais-del-mundo-con-mas-robots-por-cada-10-000-trabajadores.html>
  - Le Monde & AFP. (7 de marzo de 2007). *La Corée du Sud élabore une "charte éthique des robots"*. Recuperado de: [https://www.lemonde.fr/international/article/2007/03/07/la-coree-du-sud-elabore-une-charte-ethique-des-robots\\_880397\\_3210.html](https://www.lemonde.fr/international/article/2007/03/07/la-coree-du-sud-elabore-une-charte-ethique-des-robots_880397_3210.html)
  - Devlin, H. (18 de septiembre de 2016). *Do no harm, don't discriminate: official guidance issued on robot ethics*. The Guardian. Recuperado de: <https://www.theguardian.com/technology/2016/sep/18/official-guidance-robot-ethics-british-standards-institute>
  - Engineering and Physical Sciences Research Council. (s.f.). *Principles of robotics*. Recuperado de: <https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/>
  - Della Colleta, R. (30 de Noviembre de 2017). *La robótica eliminará hasta 800 millones de empleos en el 2030*. El País. Recuperado de: [https://elpais.com/economia/2017/11/30/actualidad/1512012918\\_284848.html](https://elpais.com/economia/2017/11/30/actualidad/1512012918_284848.html)
  - Reuters. (10 de febrero de 2016). *Google Cars Just Got a Major Boost From U.S. Vehicle Regulators*. Fortune. Recuperado de: <https://fortune.com/2016/02/10/google-nhtsa-driver/>
  - Parlamento Europeo. (16 de febrero de 2017). *European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules*

on Robotics. Recuperado de: [https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051\\_EN.html](https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html)

- Parlamento Europeo. (12 de enero de 2017). *Delvaux propone normas europeas para la robótica y un seguro obligatorio para los modelos grandes*. Noticias Parlamento Europeo. Recuperado de: <https://www.europarl.europa.eu/news/es/headlines/economy/20170109STO57505/delvaux-propone-normas-europeas-para-la-robotica-y-un-seguro-obligatorio>
- Rius, M. (17 de abril de 2018). *¿Urge regular ya los derechos de los robots en Europa?*. La Vanguardia. Recuperado de: <https://www.lavanguardia.com/tecnologia/20180417/442631680924/derechos-robots-ue-persona-electronica-ia.html>
- Pujol, J. (17 de diciembre de 2018). *Robots: ¿Hacia un nuevo concepto de “personalidad jurídica”?*. Confilegal. Recuperado de: <https://confilegal.com/20181217-hacia-un-nuevo-concepto-de-personalidad-juridica/>
- U.S. Department of Defense. (Diciembre de 2016). *Law of War Manual*. Recuperado de: <https://dod.defense.gov/Portals/1/Documents/pubs/DoD%20Law%20of%20War%20Manual%20-%20June%202015%20Updated%20Dec%202016.pdf?ver=2016-12-13-172036-190>
- Consejo Nacional de Ciencia y Tecnología. (Junio de 2019). *The National Artificial Intelligence Research And Development Strategic Plan: 2019 Update*. Recuperado de: <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>
- Comisión Europea. (25 de abril de 2018). *Artificial intelligence: Commission outlines a European approach to boost investment and set ethical guidelines*. Recuperado de: [https://ec.europa.eu/commission/presscorner/detail/en/IP\\_18\\_3362](https://ec.europa.eu/commission/presscorner/detail/en/IP_18_3362)

- Dremluga, R., & Kuznetsov, P., & Mamychev, A. (2019). *Criteria for Recognition of AI as a Legal Person*. Journal of Politics and Law. Recuperado de: [https://www.researchgate.net/publication/335305860\\_Criteria\\_for\\_Recognition\\_of\\_AI\\_as\\_a\\_Legal\\_Person](https://www.researchgate.net/publication/335305860_Criteria_for_Recognition_of_AI_as_a_Legal_Person)
  
- Marr, B. (14 de febrero de 2018). *The Key Definitions Of Artificial Intelligence (AI) That Explain Its Importance*. Forbes. Recuperado de: <https://www.forbes.com/sites/bernardmarr/2018/02/14/the-key-definitions-of-artificial-intelligence-ai-that-explain-its-importance/>
  
- Alandete, D. (27 de octubre de 2011). *John McCarthy, el arranque de la inteligencia artificial*. El País. Recuperado de: [https://elpais.com/diario/2011/10/27/necrologicas/1319666402\\_850215.html](https://elpais.com/diario/2011/10/27/necrologicas/1319666402_850215.html)
  
- Fetzer J.H. (1990) *What is Artificial Intelligence?. In: Artificial Intelligence: Its Scope and Limits*. Studies in Cognitive Systems, vol 4. Springer, Dordrecht. Recuperado de: [https://books.google.es/books?hl=es&lr=&id=77SLBQAAQBAJ&oi=fnd&pg=PT8&dq=what+is+artificial+intelligence&ots=oR7huLIDi\\_&sig=I2U3WsXGiV9mdNBoAJcIXTZvKcI#v=onepage&q=what%20is%20artificial%20intelligence&f=false](https://books.google.es/books?hl=es&lr=&id=77SLBQAAQBAJ&oi=fnd&pg=PT8&dq=what+is+artificial+intelligence&ots=oR7huLIDi_&sig=I2U3WsXGiV9mdNBoAJcIXTZvKcI#v=onepage&q=what%20is%20artificial%20intelligence&f=false)
  
- Walch, K. (4 de octubre de 2019). *Rethinking Weak Vs. Strong AI*. Forbes. Recuperado de: <https://www.forbes.com/sites/cognitiveworld/2019/10/04/rethinking-weak-vs-Bstrong-ai/#22b0252c6da3>
  
- Malpica, J. A. (s.f.). *Inteligencia Artificial y conciencia*. Universidad de Alcalá. Recuperado de: [http://www3.uah.es/benito\\_fraile/ponencias/inteligencia-artificial.pdf](http://www3.uah.es/benito_fraile/ponencias/inteligencia-artificial.pdf)
  
- Pérez, Y. (23 de agosto de 2019). *Open AI lanza versión completa de su generador de textos GPT-2*. Infochannel. Recuperado de:

- <https://www.infochannel.info/open-ai-lanza-version-completa-de-su-generador-de-textos-gpt-2>
- Nicholson, C. (s.f.). *Strong AI, Weak AI & Super Intelligence*. Pathmind. Recuperado de: <https://pathmind.com/wiki/strong-ai-general-ai>
  - Radford, A., & Wu, J., & Amodei, D., & Amodei, D. & Clarck, J., & Brundage, M., et al. (14 de febrero de 2019). *Better Language Models and Their Implications*. Open AI. Recuperado de: <https://openai.com/blog/better-language-models/>
  - University of California, Berkeley. (s.f.). *A Holistic Approach to AI*. Recuperado de: <https://www.ocf.berkeley.edu/~arihuang/academic/research/strongai3.html>
  - Jajal, T. (21 de mayo de 2018). *Distinguishing between Narrow AI, General AI and Super AI*. Medium. Recuperado de: <https://medium.com/@tjajal/distinguishing-between-narrow-ai-general-ai-and-super-ai-a4bc44172e22>
  - IBM. (s.f.). *Watson Health: Get the facts*. Recuperado de: <https://www.ibm.com/watson-health/about/get-the-facts>
  - Cunneen, M., Mullins, M. & Murphy, F. (2019) *Autonomous Vehicles and Embedded Artificial Intelligence: The Challenges of Framing Machine Driving Decisions*. Applied Artificial Intelligence, Vol. 33:8. pp. 706-731. DOI: 10.1080/08839514.2019.1600301
  - Bu, L., & Wang, J., & Wang, K. A., & Zipster, D. (Septiembre de 2019). *China digital consumer trends 2019*. McKinsey Digital. Recuperado de: <https://www.mckinsey.com/~/media/mckinsey/featured%20insights/china/china%20digital%20consumer%20trends%20in%202019/china-digital-consumer-trends-in-2019.ashx>

- Twin, A. (18 de agosto de 2019). *Data Mining*. Investopedia. Recuperado de: <https://www.investopedia.com/terms/d/datamining.asp>
- Vaghela, Y. (13 de febrero de 2019). *Here's How Big Data Analytics Has Changed the eCommerce Industry*. SmartDataCollective. Recuperado de: <https://www.smartdatacollective.com/how-big-data-analytics-has-changed-ecommerce-industry/>
- Morgan, B. (16 de julio de 2018). *How Amazon Has Reorganized Around Artificial Intelligence And Machine Learning*. Forbes. Recuperado de: <https://www.forbes.com/sites/blakemorgan/2018/07/16/how-amazon-has-reorganized-around-artificial-intelligence-and-machine-learning/#5debe4d87361>
- Sakuza, A. (3 de febrero de 2020). *Artificial Intelligence in retail innovation, by Amazon*. Recuperado de: <https://arekskuza.com/the-innovation-blog/amazon-and-artificial-intelligence-in-retail/>
- Hawksworth, J., & Berriman, R., & Cameron, E. (Febrero 2018). *Will robots really steal our jobs?*. PwC. Recuperado de: <https://www.pwc.co.uk/economic-services/assets/international-impact-of-automation-feb-2018.pdf>
- Kelly, J. (30 de septiembre de 2019). *Unbridled Adoption Of Artificial Intelligence May Result In Millions Of Job Losses And Require Massive Retraining For Those Impacted*. Forbes. Recuperado de: <https://www.forbes.com/sites/jackkelly/2019/09/30/unbridled-adoption-of-artificial-intelligence-may-result-in-millions-of-job-losses-and-require-massive-retraining-for-those-impacted/#13c809501de7>
- Hagan, S. (6 de septiembre de 2019). *More Robots Mean 120 Million Workers Need to be Retrained*. Bloomberg. Recuperado de: <https://www.bloomberg.com/news/articles/2019-09-06/robots-displacing-jobs-means-120-million-workers-need-retraining>

- Sanhermelando, J. (25 de abril de 2019). *El auge de los robots: el 50% de los empleos en España, en riesgo por la automatización*. El Español. Recuperado de: [https://www.elespanol.com/mundo/europa/20190425/auge-robots-empleos-espana-riesgo-automatizacion/393711016\\_0.html](https://www.elespanol.com/mundo/europa/20190425/auge-robots-empleos-espana-riesgo-automatizacion/393711016_0.html)
  
- Foro Económico Mundial. (2018). *The Future of Jobs Report*. Recuperado de: [http://www3.weforum.org/docs/WEF\\_Future\\_of\\_Jobs\\_2018.pdf](http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf)
  
- Oracle. (s.f.). *What Is Big Data?*. Recuperado de: <https://www.oracle.com/big-data/what-is-big-data.html#link1>
  
- Press, G. (15 de julio de 2019). *Is AI Going To Be A Jobs Killer? New Reports About The Future Of Work*. Forbes. Recuperado de: <https://www.forbes.com/sites/gilpress/2019/07/15/is-ai-going-to-be-a-jobs-killer-new-reports-about-the-future-of-work/#f61bf9fafb24>
  
- Manyika, J., & Lund, S., & Chai, M., & Bughin, J., & Woetzel, J., & Batra, P., et al. (28 de noviembre de 2017). *Jobs lost, jobs gained: What the future of work will mean for jobs, skills, and wages*. McKinsey Global Institute. Recuperado de: <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages#>
  
- Munera, I. (s.f.). *NO TEMAS A LOS ROBOTS, TU EMPLEO ESTÁ A SALVO SI SIGUES FORMÁNDOTE*. El Mundo. Recuperado de: <https://lab.elmundo.es/inteligencia-artificial/trabajo-robots.html>
  
- Shaban, H. (18 de septiembre de 2018). *Machines will create 58 million more jobs than they displace by 2022, World Economic Forum says*. The Washington Post. Recuperado de: <https://www.washingtonpost.com/technology/2018/09/18/machines-will-create-million-more-jobs-than-they-displace-by-world-economic-forum-says/>
  
- Hern, A. (6 de mayo de 2018). *Cambridge Analytica: how did it turn clicks into votes?*. The Guardian. Recuperado de:

- <https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie>
- Satara, A. (30 de marzo de 2018). *If You Don't Fully Understand the Cambridge Analytica Scandal, Read This Simplified Version*. Inc. Recuperado de: <https://www.inc.com/alyssa-satara/if-you-dont-fully-understand-cambridge-analytica-scandal-read-this-simplified-version.html>
  - Brodwin, E. (19 de marzo de 2018). *Here's the personality test Cambridge Analytica had Facebook users take*. Business Insider. Recuperado de: <https://www.businessinsider.com/facebook-personality-test-cambridge-analytica-data-trump-election-2018-3?IR=T>
  - Illing, S. (4 de abril de 2018). *Cambridge Analytica, the shady data firm that might be a key Trump-Russia link, explained*. Vox. Recuperado de: <https://www.vox.com/policy-and-politics/2017/10/16/15657512/cambridge-analytica-facebook-alexander-nix-christopher-wylie>
  - Swing State. (s.f.). Merriam-Webster. Recuperado de: <https://www.merriam-webster.com/dictionary/swing%20state>
  - Feldstein, S. (22 de abril de 2019). *How artificial intelligence systems could threaten democracy*. The Conversation. Recuperado de: <https://theconversation.com/how-artificial-intelligence-systems-could-threaten-democracy-109698>
  - Körner, K. (22 de agosto de 2019). *Digital politics: AI, big data and the future of democracy*. Deutsche Bank. Recuperado de: [https://www.dbresearch.com/PROD/RPS\\_EN-PROD/PROD0000000000497768/Digital\\_politics%3A\\_AI%2C\\_big\\_data\\_and\\_the\\_future\\_of\\_d.PDF](https://www.dbresearch.com/PROD/RPS_EN-PROD/PROD0000000000497768/Digital_politics%3A_AI%2C_big_data_and_the_future_of_d.PDF)
  - Helbing, D., & Frey, B.S., & Gigerenzer G., & Hafen, E., & Hagner, M., & Hofstetter Y., et. al. (25 de febrero de 2017). *Will Democracy Survive Big Data and*

- Artificial Intelligence?*. Scientific American. Recuperado de: <https://www.scientificamerican.com/article/will-democracy-survive-big-data-and-artificial-intelligence/>
- Lewis, P., & Hilder, P. (23 de marzo de 2018). *Leaked: Cambridge Analytica's blueprint for Trump victory*. The Guardian. Recuperado de: <https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory>
  - Hern, A. (16 de abril de 2018). *Cambridge Analytica scandal 'highlights need for AI regulation'*. The Guardian. Recuperado de: <https://www.theguardian.com/technology/2018/apr/16/cambridge-analytica-scandal-highlights-need-for-ai-regulation>
  - Confessore, N. (4 de abril de 2018). *Cambridge Analytica and Facebook: The Scandal and the Fallout So Far*. The New York Times. Recuperado de: <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>
  - Granville, K. (19 de marzo de 2018). *Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens*. The New York Times. Recuperado de: <https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html>
  - Ingram, D. (20 de marzo de 2018). *Factbox: Who is Cambridge Analytica and what did it do*. Reuters. Recuperado de: <https://www.reuters.com/article/us-facebook-cambridge-analytica-factbox/factbox-who-is-cambridge-analytica-and-what-did-it-do-idUSKBN1GW07F>
  - The Guardian. (13 de Agosto de 2019). *The Guardian view on surveillance: Big Brother is not the only watcher now*. Recuperado de: <https://www.theguardian.com/commentisfree/2019/aug/13/the-guardian-view-on-surveillance-big-brother-is-not-the-only-watcher-now>

- Díez, P. M. (20 de mayo de 2019). *El «Gran Hermano» chino lo ve todo con cámaras que reconocen caras en segundos*. ABC. Recuperado de: [https://www.abc.es/sociedad/abci-gran-hermano-chino-todo-camaras-reconocen-caras-segundos-201905190159\\_noticia.html](https://www.abc.es/sociedad/abci-gran-hermano-chino-todo-camaras-reconocen-caras-segundos-201905190159_noticia.html)
  
- Byler, D. (11 de abril de 2019). *China's hi-tech war on its Muslim minority*. The Guardian. Recuperado de: <https://www.theguardian.com/news/2019/apr/11/china-hi-tech-war-on-muslim-minority-xinjiang-uighurs-surveillance-face-recognition>
  
- Kuo, L. (2 de diciembre de 2019). *China brings in mandatory facial recognition for mobile phone users*. The Guardian. Recuperado de: <https://www.theguardian.com/world/2019/dec/02/china-brings-in-mandatory-facial-recognition-for-mobile-phone-users>
  
- Feng, E. (16 de diciembre de 2019). *How China Is Using Facial Recognition Technology*. NPR. Recuperado de: <https://www.npr.org/2019/12/16/788597818/how-china-is-using-facial-recognition-technology?t=1592402847091>
  
- Yu, K. (18 de febrero de 2020). *Facial recognition: Concerns over China's widespread surveillance*. Al Jazeera. Recuperado de: <https://www.aljazeera.com/news/2020/02/facial-recognition-concerns-chinas-widespread-surveillance-200218111532668.html>
  
- Simonite, T. (3 de septiembre de 2019). *Behind the Rise of China's Facial Recognition Giants*. Wired. Recuperado de: <https://www.wired.com/story/behind-rise-chinas-facial-recognition-giants/>
  
- Taddonio, P. (21 de noviembre de 2019). *How China's Government Is Using AI on Its Uighur Muslim Population*. PBS. Recuperado de: <https://www.pbs.org/wgbh/frontline/article/how-chinas-government-is-using-ai-on-its-uighur-muslim-population/>

- Sample, I. (7 de agosto de 2019). *South Wales police to use facial recognition apps on phones*. The Guardian. Recuperado de: <https://www.theguardian.com/technology/2019/aug/07/south-wales-police-to-use-facial-recognition-to-identify-suspects>
  
- Wood, C. (18 de marzo de 2020). *A Chinese firm claims it has built facial recognition tech that identifies people wearing masks with 95% accuracy*. Business Insider. Recuperado de: <https://www.businessinsider.com/chinese-company-claims-its-facial-recognition-95-accurate-masks-2020-3?IR=T>
  
- Doffman, Z. (3 de mayo de 2020). *China Is Using Facial Recognition To Track Ethnic Minorities, Even In Beijing*. Forbes. Recuperado de: <https://www.forbes.com/sites/zakdoffman/2019/05/03/china-new-data-breach-exposes-facial-recognition-and-ethnicity-tracking-in-beijing/#1ba64c6434a7>
  
- Marr, B. (s.f.). *What Is The Impact Of Artificial Intelligence (AI) On Society?*. Bernard Marr. Recuperado de: <https://bernardmarr.com/default.asp?contentID=1828>
  
- Internet of Things. (s.f.). Lexico. Recuperado de: [https://www.lexico.com/definition/internet\\_of\\_things](https://www.lexico.com/definition/internet_of_things)
  
- Singer, N., & Isaac, M. (9 de Agosto de 2016). *Facebook Helps Develop Software That Puts Students in Charge of Their Lesson Plans*. The New York Times. Recuperado de: <https://www.nytimes.com/2016/08/10/technology/facebook-helps-develop-software-that-puts-students-in-charge-of-their-lesson-plans.html>
  
- Marr, B. (s.f.). *How Is AI Used In Education -- Real World Examples Of Today And A Peek Into The Future*. Bernard Marr. Recuperado de: <https://bernardmarr.com/default.asp?contentID=1541>
  
- Sharma, A. (11 de julio de 2017). *AI and its social impact in the future*. Medium. Recuperado de: <https://becominghuman.ai/ai-and-its-social-impact-in-the-future-707d2049ccd9>

- Plug-In. (28 de mayo de 2019). Techopedia. Recuperado de: <https://www.techopedia.com/definition/4324/plug-in>
- FM, Y. (11 de junio de 2017). *Qué son los teraflops y qué miden exactamente*. Xataka. Recuperado de: <https://www.xataka.com/basics/que-son-los-teraflops-y-que-miden-exactamente>
- Rouse, M. (2018). *IBM Watson supercomputer*. TechTarget. Recuperado de: <https://searchenterpriseai.techtarget.com/definition/IBM-Watson-supercomputer>
- Mills, T. (15 de enero de 2020). *How AI Is Revolutionizing Health Care*. Forbes. Recuperado de: <https://www.forbes.com/sites/forbestechcouncil/2020/01/15/how-ai-is-revolutionizing-health-care/#3846c0fb403c>
- Vincent, J. (13 de Agosto de 2018). *DeepMind's AI can detect over 50 eye diseases as accurately as a doctor*. The Verge. Recuperado de: <https://www.theverge.com/2018/8/13/17670156/deepmind-ai-eye-disease-doctor-moorfields>
- Hargrave, M. (30 de abril de 2019). *Deep Learning*. Investopedia. Recuperado de: <https://www.investopedia.com/terms/d/deep-learning.asp#:~:text=Deep%20learning%20is%20a%20subset,learning%20or%20deep%20neural%20network>.
- U.S. Food & Drug Administration. (27 de abril de 2015). *The Beginnings: Laboratory and Animal Studies*. Recuperado de: <https://www.fda.gov/drugs/drug-information-consumers/beginnings-laboratory-and-animal-studies>
- Rowe, J. (22 de abril de 2019). *TECH GIANTS SHIFT INVESTMENT TO TEST NEW AI MARKETS*. Healthcare It News & Himss Media. Recuperado de: <https://www.healthcareitnews.com/ai-powered-healthcare/tech-giants-shift-investment-test-new-ai-markets>

- IBM. (s.f.). *Watson Health: Get the facts*. Recuperado de: <https://www.ibm.com/watson-health/about/get-the-facts#:~:text=By%20combining%20human%20experts%20with,growing%20body%20of%20evidence%20supports>
  
- Organización de las Naciones Unidas. (25 de septiembre de 2015). *La Asamblea General adopta la Agenda 2030 para el Desarrollo Sostenible*. Recuperado de: <https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/>
  
- Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., Felländer, A., Langhans, S. D., Tegmark, M., & Nerini, F. F. (13 de enero de 2020). *The role of artificial intelligence in achieving the Sustainable Development Goals*. Nature. Recuperado de: <https://www.nature.com/articles/s41467-019-14108-y#:~:text=Although%20AI%20enabled%20technology%20can,action%20using%20satellite%20images5>.
  
- Chui, M., Manyika, J., Harryson, M., Roberts, R., Chung, R., Nel, P., & van Heteren, A. (28 de noviembre de 2018). *Applying artificial intelligence for social good*. McKinsey Global Institute. Recuperado de: <https://www.mckinsey.com/featured-insights/artificial-intelligence/applying-artificial-intelligence-for-social-good#part%202>
  
- Jones, N. (12 de septiembre de 2018). *How to stop data centres from gobbling up the world's electricity*. Nature. Recuperado de: <https://www.nature.com/articles/d41586-018-06610-y>
  
- Karnama, A., Haghighi, E. B., & Vinuesa, R. (Diciembre de 2019). *Organic data centers: A sustainable solution for computing facilities*. ScienceDirect. Recuperado de: <https://www.sciencedirect.com/science/article/pii/S2590123019300635?via%3Dihub>

- Nerini, F. F., Slob, A., Engström, R. E., & Trutnevyte, E. (20 de marzo de 2019). *A Research and Innovation Agenda for Zero-Emission European Cities*. MDPI. Recuperado de: <https://www.mdpi.com/2071-1050/11/6/1692/htm>
  
- Nagano, A. (Abril de 2018). *Economic Growth and Automation Risks in Developing Countries Due to the Transition Toward Digital Modernity*. ACM Digital Library. Recuperado de: <https://dl.acm.org/doi/10.1145/3209415.3209442>
  
- Bachinsky, A. (21 de febrero de 2019). *The Growing Impact of AI in Financial Services: Six Examples*. Medium. Recuperado de: <https://towardsdatascience.com/the-growing-impact-of-ai-in-financial-services-six-examples-da386c0301b2>
  
- Jean, N., Burke, M., Xie, M., Davies, W. M., Lobell, D. B., & Ermon, S. (19 de agosto de 2016). *Combining satellite imagery and machine learning to predict poverty*. Science. Recuperado de: <https://science.sciencemag.org/content/353/6301/790>
  
- Marr, B. (s.f.). *How Is Machine Learning And AI Used In Healthcare – Some Practical Examples*. Bernard Marr. Recuperado de: <https://bernardmarr.com/default.asp?contentID=1450>
  
- UNESCO. (15 de mayo de 2020). *FIRST VERSION OF A DRAFT TEXT OF A RECOMMENDATION ON THE ETHICS OF ARTIFICIAL INTELLIGENCE*. Recuperado de: <https://unesdoc.unesco.org/ark:/48223/pf0000373434/PDF/373434eng.pdf.multi>
  
- Sánchez-Caro J. (2012). *Bioética y sociedad*. Escuela Nacional de Sanidad. Recuperado de: [http://e-spacio.uned.es/fez/eserv/bibliuned:500637/n6.3\\_Bio\\_tica\\_y\\_sociedad.pdf](http://e-spacio.uned.es/fez/eserv/bibliuned:500637/n6.3_Bio_tica_y_sociedad.pdf)

- Costa, A. M., & Almendro, C. (1 de junio de 2006). *Beneficencia y no-maleficencia*. Fistera. Recuperado de: <https://www.fistera.com/formacion/bioetica/beneficencia-no-maleficencia/>
  
- RecursosTic. (s.f.). 4. *La filosofía moral aristotélica*. Recuperado de: [http://recursostic.educacion.es/secundaria/edad/4esoetica/quincena3/quincena3\\_contenidos\\_4b.htm](http://recursostic.educacion.es/secundaria/edad/4esoetica/quincena3/quincena3_contenidos_4b.htm)