



Los sistemas inteligentes están protagonizando una revolución de los datos en las sociedades. Progresivamente van introduciéndose en más y más aspectos de nuestras vidas, automatizando tareas que antes llevaba mucho más tiempo realizar. No obstante, la inteligencia artificial es un arma de doble filo, que puede provocar daños y serios perjuicios en las personas en caso de no ser bien utilizada. La Comisión Europea propone trabajar en este terreno con un enfoque humanista que ponga a las personas siempre en el centro de los procesos de innovación.

La posibilidad de construir máquinas que emulen el pensamiento humano ha sido una aspiración de los científicos desde mediados del siglo pasado, cuando Alan Turing sentó los principios teóricos de la inteligencia artificial. Después de décadas de investigación y evolución, y tras sufrir numerosas decepciones que ralentizaron los avances en este campo, a partir de 2010 vuelve a cobrar fuerza y a concentrar una poderosa atención mediática. En estos años se han conseguido resultados muy relevantes en esta disciplina, especialmente centrados en la especialidad del aprendizaje automático y las redes neuronales, y se aplica con

éxito en campos como el procesado del lenguaje natural, la visión artificial, y en general, en toda aquella tarea basada en la correlación estadística, algo en lo que estos sistemas son extremadamente hábiles.

Pero, más allá de su popularidad mediática y de algún que otro hito espectacular -como cuando en 2016 un algoritmo creado por Google logra vencer al campeón mundial del juego oriental Go-, lo cierto es que la inteligencia artificial ya ha salido de los laboratorios y está muy presente en nuestra vida cotidiana. De hecho, tendemos a automatizar todo aquello susceptible de ser automatizado, intentando simplificar procedimientos y tareas, y hacer la vida más fácil.

Abundan los ejemplos a nuestro alrededor de aplicaciones de la inteligencia artificial, siendo de las más cercanas y cotidianas las que contienen nuestros teléfonos móviles, como el corrector de texto o, en el caso de los más modernos, el reconocimiento facial. Cuando la plataforma YouTube nos sugiere vídeos basándose en nuestro consumo precedente, o cuando Netflix nos recomienda series, existe un algoritmo detrás de esas plataformas aprendiendo sobre nuestras preferencias a medida que hacemos uso del servicio. En el campo de la medicina, gracias a las tecnologías de reconocimiento de imagen existen sistemas inteligentes capaces de realizar diagnósticos rápidos analizando placas de rayos X, algo que ha cobrado especial relevancia en los últimos tiempos para identificar el COVID-19 en pacientes con problemas respiratorios.

Cuando la plataforma YouTube nos sugiere vídeos basándose en nuestro consumo precedente, o cuando Netflix nos recomienda series, existe un algoritmo detrás de esas plataformas aprendiendo sobre nuestras preferencias a medida que hacemos uso del servicio

A pesar de los éxitos en este terreno, se alzan voces que cuestionan el aprendizaje automático como la principal vía de trabajo en inteligencia artificial, como es el caso de Gary Marcus, experto en psicología cognitiva y profesor en la Universidad de Nueva York. Marcus ha plasmado su visión sobre las limitaciones del aprendizaje profundo tanto en un paper, que generó un encendido debate entre la comunidad científica desde que fue publicado en enero de 2018, como en el libro, firmado a medias con Ernest Davis, *Rebooting AI. Buiding Artificial Intelligence We Can Trust* (2019). Su crítica se centra sobre todo en que el aprendizaje profundo requiere de una inmensa cantidad de datos para entrenar a los algoritmos, en que es poco transparente en sus procesos, ofrece pocas oportunidades de transferencia del escenario en que ha sido adiestrado el sistema a otros, y finalmente, en que carece del principio de causalidad, es decir, aunque es capaz de establecer correlaciones entre elementos, no sabe distinguir entre causa y efecto.

Precisamente, en este último aspecto es en el que concentra también su crítica Judea Pearl, otro experto que pone en cuestión la inteligencia artificial basada en el deep learning. En su libro *The Book of Why* lo expresa poéticamente utilizando a Platón:

“Como los prisioneros en la famosa caverna de Platón, los sistemas basados en aprendizaje profundo exploran las sombras en la pared de la caverna y aprenden a predecir con precisión sus movimientos. Carecen del entendimiento de que las sombras observadas son meras proyecciones de objetos tridimensionales que se mueven en un espacio tridimensional. La inteligencia artificial fuerte requiere este entendimiento.”

La rápida difusión que está conociendo la inteligencia artificial despierta no poca preocupación sobre los posibles efectos negativos que pueda tener en las vidas de las personas. La Comisión Europea se ha expresado al respecto en su reciente informe Libro Blanco sobre inteligencia artificial – Una aproximación europea a la excelencia y la confianza. La inteligencia artificial puede socavar los valores sobre los que está construida la Unión Europea, y violar derechos fundamentales de los ciudadanos comunitarios, como la libertad de expresión, de reunión, el respeto a la dignidad humana y la no discriminación por razones de raza, etnia, orientación sexual, discapacidad, edad, o credo o religión. También se considera una potencial amenaza para la protección de los datos personales y el derecho a la intimidad, así como para garantizar el derecho a un juicio justo y la protección del consumidor.

Estos sistemas inteligentes pueden monitorizar con eficacia el comportamiento de las personas, también pueden llegar a identificar a individuos concretos dentro de los grandes volúmenes de *big data* (se conoce como *desanonimizar*), y por tanto, pueden crear fisuras en el derecho a la privacidad y la intimidad. Por otro lado, los algoritmos basados en el *machine learning* han desarrollado en ocasiones sesgos no deseados de género o de raza. La empresa Northpointe desarrolló un algoritmo para predecir la probabilidad de que un delincuente reincidiera y este empezó a juzgar que las personas de color eran más proclives a volver a cometer un crimen que los blancos. Por otro lado, los algoritmos de reconocimiento facial tanto de IBM como de Microsoft y de Amazon funcionan mejor con rostros masculinos que femeninos¹.

La necesidad de contar con sistemas fiables

Además de impulsar la investigación científica y técnica en este campo, las instituciones de la Unión Europea han abierto una línea de trabajo para determinar cómo podemos crear unos sistemas de inteligencia artificial que traigan consigo beneficios a las personas y no perjuicios. De esta forma, desde junio de 2018 el denominado el Grupo de expertos de alto nivel sobre IA trabaja en esta dirección, y uno de sus principales resultados ha sido el informe *Directrices éticas para una IA fiable*, presentado en abril de 2019.

El eje de este análisis es el concepto de fiabilidad de la inteligencia artificial, que debe reposar sobre tres pilares: debe ser lícita, es decir, cumplir todas las leyes y reglamentos aplicables; también ha de ser ética, de modo que se garantice el respeto de los principios y valores éticos; y, finalmente, debe ser robusta, tanto desde el punto de vista técnico como social. Cada uno de estos componentes es en sí mismo necesario, pero no es suficiente para el logro de una inteligencia artificial fiable.

El Grupo de Expertos identifica una serie de directrices éticas que deben acompañar la construcción de máquinas inteligentes:

1. Desarrollar, desplegar y utilizar los sistemas de inteligencia artificial (IA) respetando los principios éticos de respeto de la autonomía humana, prevención del daño, equidad y explicabilidad. Reconocer y abordar las tensiones que pueden surgir entre estos principios.
2. Prestar una atención especial a las situaciones que afecten a los grupos más vulnerables, como los niños, las personas con discapacidad y otras que se hayan visto históricamente desfavorecidas o que se encuentren en riesgo de exclusión, así como a las situaciones caracterizadas por asimetrías de poder o de información, como las que pueden producirse entre empresarios y trabajadores, o entre empresas y consumidores.
3. Reconocer y tener presente que, pese a que aportan beneficios sustanciales a las personas y a la

sociedad, los sistemas de IA también entrañan determinados riesgos y pueden tener efectos negativos, algunos de los cuales pueden resultar difíciles de prever, identificar o medir (por ejemplo, sobre la democracia, el Estado de derecho y la justicia distributiva, o sobre la propia mente humana). Adoptar medidas adecuadas para mitigar estos riesgos cuando proceda; dichas medidas deberán ser proporcionales a la magnitud del riesgo.

En relación con el marco ético establecido, se proponen siete requisitos que debe cumplir una inteligencia artificial fiable:

1. Acción y supervisión humanas
2. Solidez técnica y seguridad
3. Gestión de la privacidad y de los datos
4. Transparencia
5. Diversidad, no discriminación y equidad
6. Bienestar ambiental y social
7. Rendición de cuentas.

El Libro Blanco sobre inteligencia artificial

A principios de 2020, la Comisión Europea publicó su libro blanco sobre inteligencia artificial, intitulado *White Paper On Artificial Intelligence - A European approach to excellence and trust*, que, basado en el análisis publicado el año precedente por el grupo de expertos, plantea distintas opciones políticas para alcanzar dos objetivos: por una parte, promover la implantación en Europa de sistemas inteligentes, y por otra, ser capaces de controlar los riesgos que lleva asociados.

El documento destaca cómo la inteligencia artificial -una combinación de datos, algoritmos y capacidad de computación- puede combinarse con las fortalezas tecnológicas e industriales europeas para que la UE se convierta en líder mundial de la economía del dato. Sin embargo, subraya que resulta indispensable que la inteligencia artificial europea esté fundamentada en nuestros valores y derechos fundamentales, como la dignidad humana y la protección de la privacidad. La Comisión defiende también que los sistemas inteligentes deben colaborar en la consecución de los Objetivos de Desarrollo Sostenible -así como los procesos democráticos y los derechos sociales-, y apoyar el Pacto Verde Europeo, la hoja de ruta de Europa para desarrollar una economía sostenible.

La infraestructura de políticas que persigue esta estrategia debe llevar a la creación de lo que los autores denominan “un ecosistema de excelencia” en la cadena de valor de la economía de la Unión Europea, en torno a la adopción de soluciones de inteligencia artificial, movilizando recursos tanto públicos como privados. Para ello, en paralelo se pretende instaurar un “ecosistema de confianza”, con el fin de garantizar a los ciudadanos la seguridad y confianza necesarias para que usen sin miedo las aplicaciones basadas en inteligencia artificial, y también, para que las empresas y organismos públicos puedan innovar en este campo enmarcados por un entorno legal.

Se pretende instaurar un “ecosistema de confianza” con el fin de garantizar a los ciudadanos la seguridad y

confianza necesarias para que usen sin miedo las aplicaciones basadas en inteligencia artificial

El enfoque defendido por la Comisión Europea es antropocéntrico: la inteligencia artificial no es un fin en sí misma, sino una herramienta para servir a las personas e incrementar su bienestar. En concreto, debe apoyar a los individuos para realizar mejores elecciones y mejor informadas de acuerdo con sus intereses y objetivos. De esta forma, los sistemas inteligentes deben ser habilitadores de una sociedad equitativa, apoyando la acción humana y los derechos fundamentales, sin disminuir, limitar o extraviar la autonomía de los seres humanos.

Foto de [Markus Spiske](#) en [Pexels](#)