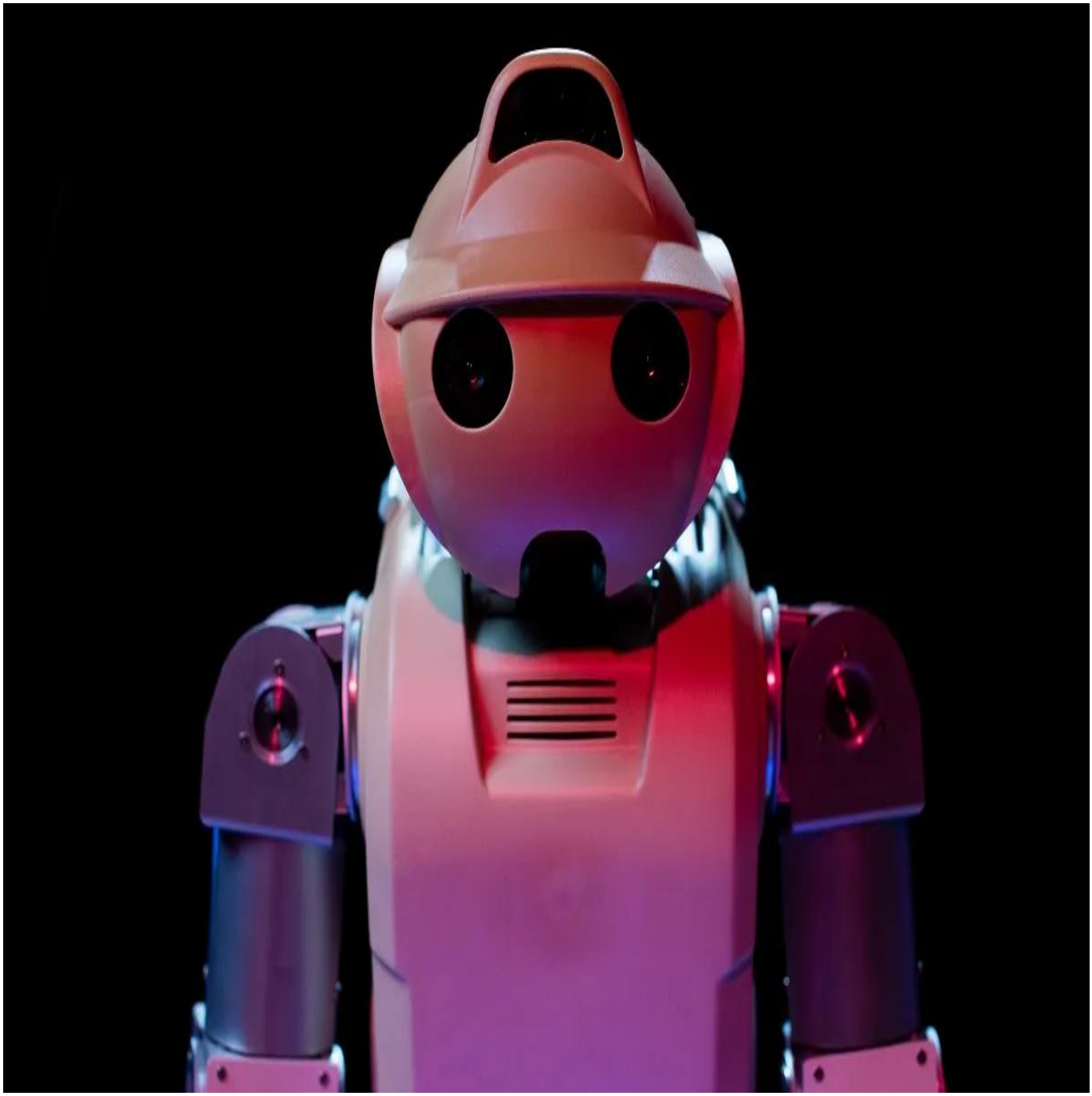
La apuesta europea por una inteligencia artificial fiable



La Unión Europea está impulsando la inteligencia artificial para no quedarse atrás en términos de innovación y competitividad. No obstante, el desarrollo de sistemas y máquinas inteligentes debe ir regido por un marco de directrices éticas, que garanticen que todos los europeos salen beneficiados de su aplicación.

Inteligencia artificial, ese término de moda que causa a la vez admiración y un oscuro temor reverencial. De alguna forma, no nos resulta cómoda la idea de que proliferen máquinas capaces de hacer las cosas que hacemos los humanos, y en ocasiones, mejor que los propios humanos. La revolución digital se plantea actualmente en determinados foros como un proceso mucho más disruptivo que la revolución industrial que tuvo lugar a lo largo del siglo XIX. Entonces también se temía los efectos no deseados del progreso, aunque, con el tiempo, el maquinismo acabó por traer una mejora en las condiciones de vida de las personas.

Nadie niega que la llegada de las máquinas inteligentes acarrea innumerables beneficios para la economía y para la sociedad en su conjunto, en términos de mejoras en la productividad de las empresas y en la eficiencia

de determinados procesos. Se trata de sistemas que poco a poco van formando parte de nuestras vidas cotidianas, como los asistentes virtuales alojados en los altavoces inteligentes o en nuestros teléfonos, que aprenden de nosotros, y que procuran ofrecernos servicios e información útiles. Además, cada vez resulta más fácil relacionarse con ellos, pues ahora los algoritmos informáticos han aprendido a reconocer el habla humana: entienden lo que les decimos.

«La economía digital será el motor principal para el crecimiento de la economía mundial y los chips de inteligencia artificial más avanzados conformarán la infraestructura de la nueva época», ha llegado a afirmar Robin Li, presidente del gigante tecnológico chino Baidu, en el marco de la edición 2019 de la Conferencia Mundial de Internet[1]. Todos los grandes tecnólogos coinciden en este juicio, pero también muestran su cautela ante posibles problemas relacionados con la gestión de la seguridad y la privacidad que presentan estas tecnologías.

> La economía digital será el motor principal para el crecimiento de la economía mundial y los chips de inteligencia artificial más avanzados conformarán la infraestructura de la nueva época

No son estas las únicas cuestiones que nos deben preocupar en relación con los sistemas inteligentes. Lejos de visiones distópicas de robots que controlan el mundo -más propias de la ciencia ficción que de un futuro probable-, lo cierto es que la inteligencia digital plantea amenazas reales que pueden causar daños a personas y a colectivos específicos. En la medida en que vamos dejando cada vez más acciones y procesos de toma de decisiones en manos de las máquinas, abrimos la puerta a que se produzcan daños físicos, por un fallo de funcionamiento –por ejemplo, del piloto automático de un avión o de un coche autónomo-, o la posibilidad de que se produzcan situaciones abiertamente injustas y discriminatorias.

Es por todo ello que, en paralelo a la investigación para el desarrollo de sistemas inteligentes cada vez más perfectos y eficaces, se debe trabajar en garantizar que su uso y aplicación no puedan perjudicar a nadie. Las autoridades de la Unión Europea, conscientes de los riesgos que entraña este campo de la tecnología, han encargado a un grupo de expertos de alto nivel la redacción de unas directrices para promover una inteligencia artificial fiable, tarea que han plasmado en un documento que ha visto la luz en abril de 2019.

Las amenazas de las máquinas inteligentes

A pesar de los logros que tienen lugar en el campo de la inteligencia artificial, presenta numerosos riesgos para las personas, tanto por culpa de fallos o errores, como por ser utilizada deliberadamente para destruir y hacer daño. Karen Hao y Will Knight, expertos del MIT en este campo, identifican las siguientes amenazas actuales.

La inmadurez de la tecnología del vehículo autónomo. A pesar de que numerosas empresas, tanto tecnológicas como de automoción, están inmersas en la carrera por lanzar el coche autónomo, los accidentes

que suceden relacionados con este tipo de vehículos hacen pensar que la conducción autónoma está aún muy verde. El pasado abril tuvo lugar el primer accidente de este tipo de vehículos en nuestro país, en concreto en Galicia, cuando un Tesla de conducción automática chocó contra un coche que estaba estacionado en el arcén.

Y también existe la posibilidad de que la inteligencia de uno de esos vehículos resulte "engañada" a propósito, algo que afirman que han conseguido hacer los investigadores de un laboratorio de la empresa china Tencent con un vehículo Tesla. Parece ser que, alterando los datos que reciben los sensores del coche, han logrado confundir al algoritmo que lo gobierna.

Bots dedicados a la manipulación de la opinión pública. En marzo de 2018 saltó a los medios el escándalo de la consultora Cambridge Analytica, que utilizó los datos personales de los usuarios de Facebook para influir en la intención de voto de las elecciones presidenciales de Estados Unidos de 2016. Este caso puso en evidencia cómo los algoritmos de inteligencia artificial que determinan la información que se ofrece en redes sociales pueden ser manipulados para que promuevan la desinformación, para evitar el debate y el intercambio de opiniones, y para aislar a los ciudadanos en sus propios *filtros burbuja* –utilizando la terminología acuñada por Eli Pariser- de opinión.

Relacionado con lo anterior, en España en mayo el Tribunal Constitucional derogó el artículo 58 bis de la Ley Orgánica 5/1985, de 19 de junio, del régimen electoral general (conocida como LOREG), que trataba sobre la utilización de medios tecnológicos y datos personales en actividades electorales. En concreto, permitía a "*los partidos políticos, coaliciones y agrupaciones electorales recopilar datos personales vinculados a las opiniones políticas de los ciudadanos para la realización de todo tipo de actividades políticas durante el periodo electoral"*. En suma, autorizaba a todos esos agentes a espiar al ciudadano, recogiendo sus datos libremente para poder usarlos en las campañas electorales.

La posibilidad de crear armas letales. Los empleados de Google se rebelaron contra la dirección de la empresa cuando descubrieron que ésta estaba colaborando con las fuerzas aéreas estadounidenses para crear drones inteligentes de ataque. El denominado Proyecto Maven fue abandonado por la compañía tecnológica, pero eso no quiere decir que no estén en marcha por todo el mundo investigaciones tendentes a crear armas inteligentes.

Los empleados de Google se rebelaron contra la dirección de la empresa cuando descubrieron que ésta estaba colaborando con las fuerzas aéreas estadounidenses para crear drones inteligentes de ataque.

La amenaza de las herramientas de reconocimiento facial. Aunque es presentada como una tecnología clave para la identificación de usuarios en un futuro cercano –que probablemente sustituya al uso actual de claves de acceso-, el utilizar inteligencia artificial para verificar que somos quienes decimos que somos entraña numerosas dudas éticas relacionadas con el derecho a la privacidad y, en general, con las libertades

individuales.

Deep fake, una plaga difícil de erradicar. Más allá de las *fake news*, la manipulación de vídeos mediante sistemas inteligentes, para humillar o desacreditar a determinadas personas y personalidades falseando las imágenes, se presenta como uno de los grandes problemas mediáticos de los últimos tiempos. Una tecnología cada vez más sofisticada que permite realizar falsificaciones de imágenes –o *deep fakes*– cada vez más convincentes.

Los algoritmos que discriminan. El uso sin control de algoritmos para tratar los datos en ámbitos tan diversos como el mundo financiero, la justicia, la educación o la selección de personal, puede causar un daño inmenso a determinados colectivos y a ciudadanos individuales. En este sentido, Cathy O´Neil habla de *armas de destrucción matemática*[2], para resaltar lo destructivos que pueden llegar a ser. Para ella, un arma de destrucción matemática es un algoritmo que se utiliza en la toma de decisiones importantes que afectan a muchas personas, como, por ejemplo, qué trabajo pueden conseguir, qué tipo de tarjeta de crédito deberían usar o en qué universidad deberían estudiar.

El problema es que, aunque el algoritmo es muy relevante en la vida de las personas a las que se dirige, su funcionamiento es un secreto para ellas. De hecho, suele ser un sistema de evaluación que no entienden, y, con frecuencia, ni siquiera saben que están siendo evaluadas por él. La inteligencia artificial a veces comete errores: trata injustamente a ciertas personas, pero sus decisiones son inapelables.

La apuesta europea por la IA

La inteligencia artificial está en el corazón de la agenda científica de la Unión Europea. La Comisión reconoce la capacidad de este conjunto de tecnologías para mejorar la vida de las personas, y para generar beneficios para la sociedad y la economía, a través de cuestiones como el impulso del cuidado de la salud, el aumento de la eficiencia de la administración pública, haciendo el transporte más seguro, inyectando competitividad en la industria o generando una agricultura más sostenible.

> A través del programa Horizon 2020, la Comisión Europea ha dedicado 2 600 millones de euros a las áreas relacionadas con la inteligencia artificial, como la robótica, el big data, el transporte, la sanidad y las tecnologías emergentes.

El impacto económico de la automatización del trabajo intelectual, de los robots y de los vehículos autónomos, se calcula que alcanzará entre los 6,5 y los 12 billones de euros anuales para 2025. Sin embargo, Europa se queda atrás en relación a la inversión privada en inteligencia artificial, pues dedicó en 2016 entre 2 400 y 3 200 millones de euros frente a los entre 6 500 y 9 700 millones de Asia, y los entre 12 100 y 18 600 millones de Norteamérica.

Es por ello, que la Comisión Europea impulsa con grandes inversiones aspectos como los sistemas cognitivos, la robótica, el *big data* y las tecnologías emergentes del futuro, con el fin de garantizar el mantenimiento de la competitividad del tejido económico del continente.

A través del programa Horizon 2020, ha dedicado 2 600 millones de euros a las áreas relacionadas con la inteligencia artificial, como la robótica, el *big data*, el transporte, la sanidad y las tecnologías emergentes. La investigación en robótica ha recibido hasta 700 millones de fondos públicos, a los que hay que sumar 2 100 millones de financiación privada. Los Fondos Estructurales y de Inversión europeos también han incidido en la formación y el desarrollo de capacidades con 27 000 millones de gasto en ese tema, de los cuales 2 300 han sido dedicados al desarrollo de capacidades digitales.

La necesidad de contar con unos sistemas fiables

Las autoridades europeas son conscientes de los peligros que entraña el desarrollo descontrolado y no razonado de la inteligencia artificial. En consecuencia, paralelamente al desarrollo científico y tecnológico, han visto la necesidad de abrir un debate para esclarecer cómo conseguir que los sistemas inteligentes traigan consigo beneficios a las personas, y no perjuicios.

A tal efecto, en junio de 2018 la Comisión creó el Grupo de expertos de alto nivel sobre IA, que en abril de 2019 presentó el documento Directrices éticas para una IA fiable. El informe pretende ofrecer algo más que una simple lista de principios éticos, y proporciona orientación sobre cómo poner en práctica esos principios en los sistemas sociotécnicos.

La inteligencia artificial debe ser lícita, ética y robusta.

Los autores se centran en el concepto de fiabilidad de la inteligencia artificial, que hacen 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.

Las directrices que propone en el trabajo se dirigen solamente a los dos últimos aspectos, la ética y la robustez. 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 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. Garantizar que el desarrollo, despliegue y utilización de los sistemas de IA cumpla los requisitos para una IA 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, y 7) rendición de cuentas.
2. Para garantizar el cumplimiento de estos requisitos, se deberá estudiar la posibilidad de emplear tanto métodos técnicos como no técnicos.
3. Impulsar la investigación y la innovación para ayudar a evaluar los sistemas de IA y a promover el cumplimiento de los requisitos; divulgar los resultados y las preguntas de interpretación abierta al público en general, y formar sistemáticamente a una nueva generación de especialistas en ética de la IA.
4. Comunicar información a las partes interesadas, de un modo claro y proactivo, sobre las capacidades y limitaciones de los sistemas de IA, posibilitando el establecimiento de expectativas realistas, así como sobre el modo en que se cumplen los requisitos. Ser transparentes acerca del hecho de que se está trabajando con un sistema de IA.
5. Facilitar la trazabilidad y la auditabilidad de los sistemas de IA, especialmente en contextos o situaciones críticos.
6. Implicar a las partes interesadas en todo el ciclo de vida de los sistemas de IA. Promover la formación y la educación, de manera que todas las partes interesadas sean conocedoras de la IA fiable y reciban formación en la materia.
7. Ser conscientes de que pueden existir tensiones fundamentales entre los diferentes principios y requisitos. Identificar, evaluar, documentar y comunicar constantemente este tipo de tensiones y sus soluciones.

Comisión Europea (2019) “Directrices éticas para una IA fiable”.

European Commission (2019) “Artificial Intelligence For Europe”.

Hao, K. y Knight, W. (2019) “Never mind killer robots—here are six real AI dangers to watch out for in 2019” en *MIT Technology Review*. Disponible en: https://www.technologyreview.com/s/612689/never-mind-killer-robotshere-are-six-real-ai-dangers-to-watch-out-for-in-2019/

Hao, K. (2019) “Confiamos mucho en softwares de IA que no merecen esa confianza» en *MIT Technology Review*. Disponible en: https://www.technologyreview.es/s/11511/confiamos-mucho-en-softwares-de-ia-que-no-merecen-esa-confianza

Knight, W. (2019) “Military artificial intelligence can be easily and dangerously fooled” en *MIT Technology Review*. Disponible en: https://www.technologyreview.com/s/614497/military-artificial-intelligence-can-be-easily-and-dangerously-fooled/?utm_source=newsletters&utm_medium=email&utm_campaign=the_download.unpaid.engagement