

# Los sesgos y la manipulación encabezan el «lado oscuro de la IA»

**La inteligencia artificial marca el inicio de una nueva Revolución para la Humanidad. Sus innumerables ventajas son innegables en la era de datos en la que vivimos, pero también entraña ciertos riesgos. Los sesgos y la manipulación encabezan el “lado oscuro de la IA” y su prevención es hoy más importante que nunca.**

[ILUSTRACIÓN: DAN4 / [ISTOCK](#) ]

La inteligencia artificial (IA) ya incide en múltiples ámbitos de nuestra vida cotidiana, desde el reconocimiento de voz en nuestros teléfonos hasta el análisis de datos en la investigación médica. Su avanzada tecnología ofrece innumerables ventajas, como la automatización de tareas, eficiencia, capacidad de procesamiento de grandes volúmenes de datos y personalización de múltiples servicios.

Sin embargo, a medida que se propaga, sin aparentes límites, el uso y desarrollo de esta tecnología, se vislumbra un lado no tan positivo derivado de las propias características que hacen especial a la IA: la ‘toma de decisiones’ y su ‘capacidad’ para ejecutarlas de manera autónoma. Cuando un ser humano toma decisiones, normalmente, la envuelve de un contexto ético (o, al menos, tiene esa capacidad de discernir entre lo bueno y lo malo de su aplicación).

**Los sistemas de IA son software, es decir, programas informáticos sin responsabilidad ni intención**

La IA, como software que es, carece por ahora de ese contexto. Por tanto, es importante que nosotros, las personas, acompañemos las grandes bondades de esta tecnología, aplicando nuestra ética en estos procesos en función del caso. Pero veamos brevemente cuáles son algunos de los principales riesgos actuales, y por qué son cada vez más voces las que claman sobre la necesidad de una regulación más dura frente a la -a veces- laxa ética.

## Ética para máquinas

Se habló mucho del caso de un algoritmo de Amazon que estuvo durante un tiempo actuando de manera discriminatoria a la hora de seleccionar perfiles para un puesto de trabajo de desarrollador de software. Parece que sólo admitía a hombres jóvenes y de raza blanca, descartando automáticamente al resto. [Este hecho saltó en su día a los periódicos](#), ensalzando la amenaza que podría suponer el uso de IA en determinados procesos, por ejemplo la selección de recursos humanos. Fue solo el principio de la consciencia de algunos de los riesgos de la IA asociados de los sesgos del algoritmo.

**Las personas debemos acompañar las grandes bondades de la IA aplicando nuestra ética**

Los sistemas de IA aprenden a partir de datos. Y si esos datos reflejan prejuicios existentes en la sociedad, existe el riesgo de perpetuar o incluso exacerbar esos sesgos, especialmente sin la supervisión adecuada. Esta cuestión se ha manifestado ya, en varias ocasiones, en sistemas de contratación, créditos y sistemas judiciales. En 2019, por ejemplo, [se conoció que el algoritmo de crédito de Apple Card](#), gestionado por Goldman Sachs, asignaba límites de crédito significativamente más bajos a mujeres en comparación con hombres, incluso si compartían los mismos perfiles de ingresos y crédito.

A la hora de combatir estos desafíos, la IA no es diferente a otras tecnologías de gran impacto como lo fueron el automóvil, en su día, o internet. Éstas también necesitaron regularse para prevenir consecuencias negativas. En este caso, y teniendo en cuenta su rápida evolución, sin duda se requiere una regulación dinámica y adaptativa.

Por otro lado, surge la cuestión de la responsabilidad en caso de errores o accidentes causados por sistemas de IA. Sin un marco legal claro, determinar quién es responsable puede suponer un desafío. Un claro ejemplo es el del coche autónomo. ¿De quién sería la culpa si atropellara a un peatón? Pese a que esta cuestión ha sido ya ampliamente debatida a nivel internacional, aún no existe una respuesta totalmente unánime, más allá de hablar de una 'responsabilidad compartida' entre la empresa que desarrolla el automóvil (y su software de IA) y, por supuesto, el conductor.

## **Regulación en pañales**

En la mayoría de países, las leyes en materia de IA se encuentran en etapas incipientes. En Estados Unidos, por ejemplo, la normativa ha sido más sectorial y depende, en gran medida, de los estados individuales, aunque existen ciertos marcos federales en áreas específicas, como la privacidad o la discriminación.

### **En la mayoría de países, las leyes en materia de IA se encuentran en etapas incipientes**

Desde un punto de vista global, el reto radica en equilibrar la innovación con la protección del ciudadano. La [Carta de Derechos Digitales](#) que puso en marcha el Gobierno español en 2021 es un ejemplo de esta tarea pendiente de concienciación y prevención ciudadana. Sin embargo, aún falta accionarla para que sea realmente eficaz, y esto no es una tarea fácil. Una regulación excesiva podría sofocar la innovación, mientras que una falta de regulación, y de medidas concretas y prácticas, podría dejar a las personas desprotegidas en muchos sentidos.

En esta línea, la Unión Europea aprobó, en agosto de 2024, la Ley de Inteligencia Artificial, en un esfuerzo común entre los organismos reguladores europeos, empresas, expertos en IA y sociedad civil. Su objetivo: proteger los derechos fundamentales de las personas, garantizar la transparencia en la toma de decisiones de los sistemas de IA y establecer mecanismos de rendición de cuentas y supervisión humana adecuados, entre otras cuestiones básicas.

No pretenden regular la tecnología en sí misma, ya que esto supondría un problema para su implementación y desarrollo en la industria de la UE, sino casos específicos que pueden suponer un riesgo. Están los usos y desarrollos prohibidos (como puntuación social de la población.), los de alto riesgo (aquellos cuya implementación puede afectar a los derechos fundamentales de la persona, como uso de la IA para incorporación a un puesto de trabajo), los de riesgo medio (obligaciones de transparencia, como en el caso de chatbots) y los de riesgo bajo (automatismos sin riesgo, como filtros de spam).

## La manipulación, uno de los principales riesgos reales

Uno de los más claros fenómenos que ya se está empezando a vislumbrar es la manipulación, esa tendencia a dejarnos llevar por ideas no consolidadas por nosotros mismos, incluso de manera inconsciente.

El empleo, principalmente a través de sistemas de recomendación, está empezando a incidir en la capacidad de decisión de unos consumidores cada vez más pasivos. Pero, en pocas líneas, cabría plantearse quién es 'el verdadero culpable' de una situación que podría desembocar en una merma para la actual democracia. ¿Es la inteligencia artificial, como tecnología? ¿Son las máquinas que 'toman decisiones' por nosotros y nos incitan a un consumo premeditado?

## Software sin intención ni responsabilidad

En primer lugar, los sistemas de IA son software, es decir, programas informáticos, con peculiaridades específicas. Por tanto, sin responsabilidad ni intención. Sí la tienen, sin embargo, las empresas u organismos - tanto públicos, como privados- que las usan o las desarrollan. Clarificar esta cuestión es clave para dejar de personificar en la IA las responsabilidades humanas con el empleo de cualquier tecnología.

## Sin espíritu crítico y conocimiento, la era de la inteligencia artificial podría llegar a relacionarse con la de la manipulación masiva inconsciente

Por otro lado, tenemos la cuestión de la preparación real de nuestra sociedad para asumir el impacto de la IA. Aunque, a partir de la salida de ChatGPT empezó a democratizarse su uso, el conocimiento general sobre sus repercusiones en nuestro día aún es muy vago.

Es necesario que se involucren los poderes públicos, las empresas y los centros educativos para trabajar conjuntamente en una revolución cultural y educativa que fomente el conocimiento en impacto de la IA y el espíritu crítico de las personas. En caso contrario, la era de la inteligencia artificial podría llegar a relacionarse con la de la manipulación masiva inconsciente.

En definitiva, la IA presenta un potencial transformador inmenso. Sin embargo, como cualquier herramienta poderosa, conlleva tanto oportunidades como peligros. Aunque existen esfuerzos en curso para desarrollar una regulación sólida, todavía hay un largo camino por recorrer para garantizar su uso seguro y ético en nuestra sociedad.

Eso sí, debemos recordar que el peligro no es la IA, en sí misma, sino el uso que nosotros, los humanos, hacemos de ella. La clave está en aumentar la conciencia popular y profesional sobre los riesgos. Por otro lado, aunque hoy día no es una opción negarse a la incorporación de esta tecnología en nuestra vida, se debe incentivar su aplicación y desarrollo ético y responsable desde el diseño, es decir, desde el planteamiento inicial del proyecto con IA. Asimismo, se debe realizar un seguimiento durante todo su ciclo de vida. Solo de esta manera conseguiremos un impacto positivo de esta revolucionaria tecnología.

**Argyle, LP; Bail, CA; Busby, EC; Gubler, JR; Howe, T; Rytting, C. Sorensen; T. Wingate, D.** "Leveraging AI for democratic discourse: Chat interventions can improve online political conversations at scale" en Proc Natl Acad (2023, 120, 41). Disponible en: <https://www.pnas.org/doi/10.1073/pnas.2311627120>  
**Comisión Europea.** Ley europea de IA. Comisión Europea, 2025. Disponible en: <https://digital-strategy.ec.europa.eu/es/policies/regulatory-framework-ai>  
**CDE.** Inteligencia artificial: amenazas y oportunidades. Centro de Documentación Europea de la Universidad

de Almería, 2020. Disponible en: <https://www.cde.ual.es/inteligencia-artificial-amenazas-y-oportunidades/>  
**Parlamento Europeo.** Inteligencia artificial: oportunidades y desafíos. Parlamento Europeo, 2020. Disponible en <https://www.europarl.europa.eu/news/es/headlines/society/20200918STO87404/inteligencia-artificial-oportunidades-y-desafios>

**Salazar, I; Benjamins, R** (2020): El mito del algoritmo: cuentos y cuentas de la inteligencia artificial. ANAYA multimedia. ISBN-10: 8441542805

**Salazar, I** (2019): La Revolución de los robots. Cómo la inteligencia artificial y la robótica afectan a nuestro futuro. Asturias, Ediciones Trea. ISBN: 1 978-8417767-34-1

**Salazar, I; Gómez de Agreda, Á.** “Sesgos y perspectiva cultural en el entrenamiento de los algoritmos de inteligencia artificial” en Revista Privacidad y Derecho Digital. (2019, 15, 29-63). ISSN: 2444-5762

**Salazar, I.** “Los robots y la inteligencia artificial: nuevos retos del periodismo” en Doxa Comunicación (2018, 27, 295-315). ISSN: 1696-019X