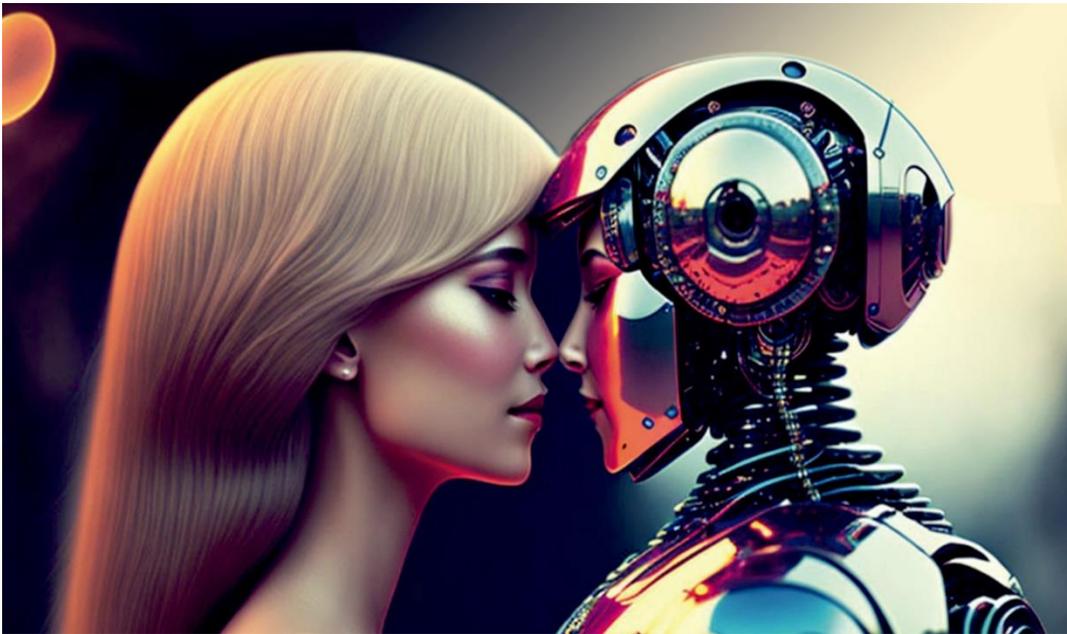




Madrid, 18 de enero 2024

“Hemos normalizado tanto la digitalización que no pensamos en sus implicaciones sociales y morales”

- Sara Degli-Esposti, experta en ética e inteligencia artificial firma el nuevo libro de la colección ¿Qué sabemos de? (CSIC-Catarata)
- La publicación plantea el debate sobre los riesgos y las oportunidades de la IA



Todos tenemos móviles inteligentes, pero casi no conocemos las tecnologías que los hacen posibles ni que los datos que se van almacenando pueden ponernos en peligro. / Wikimedia Commons

“El mundo de la inteligencia artificial (IA) que conocemos hoy se basa en una simple constatación: la importancia de los datos. El *big data* es la materia prima que alimenta los modelos de IA basados en aprendizaje automático, y **estos modelos inundan los dispositivos digitales que colonizan cada detalle de nuestras vidas**, desde lo laboral a lo más íntimo y personal. Se trata de un proceso diario tan normalizado que rara vez nos



paramos a pensar en las implicaciones que conlleva”. Este es el punto de partida del ensayo que la investigadora experta en ética e inteligencia artificial del CSIC **Sara Degli** publica en la **colección ¿Qué sabemos de? (CSIC-Catarata)**.

La ética de la inteligencia artificial se centra en el **impacto social y en los retos morales de la digitalización y de la economía basada en los datos**. Lejos de dar respuestas concluyentes, Sara Degli-Esposti plantea el debate sobre la IA con ejemplos concretos que permiten entender sus aplicaciones más comunes. La autora reflexiona sobre la **confianza que depositamos en estas tecnologías**, a pesar de que la mayoría desconocemos su funcionamiento, así como en **las personas y entidades que están detrás de su desarrollo e implantación**. También habla de la **relación entre bienestar y digitalización**, de los **principios éticos** que se están intentando adoptar desde Europa para desarrollar una IA fiable, y de las posibles **soluciones a nivel legal y técnico** propuestas para promover el buen desarrollo de una tecnología que ha experimentado un crecimiento fulgurante en los últimos años.

Y es que muchas de las consecuencias del uso de la IA **ya no son parte de una serie de ciencia ficción, sino que pertenecen a una realidad cotidiana**, y, a veces, algo inquietante. “De las aplicaciones de la inteligencia artificial dependen algunas de nuestras oportunidades en la sociedad, tanto de encontrar trabajo como pareja, porque **nuestra reputación se mide en las redes sociales**. Nos preocupa quedarnos obsoletos en un mundo diseñado por la informática, así como el efecto de la desinformación en las democracias occidentales” explica la científica del Instituto de Filosofía del CSIC. También nos alerta que se automatice la injusticia y que, a través de los sesgos, **se perpetúen dinámicas de discriminación y segregación social**. “Por otro lado, crece la intranquilidad por las consecuencias de los entornos digitales sobre la capacidad de aprender, expresarse y relacionarse de las nuevas generaciones”, añade.

Ética como escudo para la IA

Durante los últimos años, muchas entidades públicas y privadas se han esforzado en elaborar líneas maestras para el desarrollo ético de la IA. Son, según el filósofo y director del Instituto de Filosofía del CSIC Txetxu Ausín, “como un amparo y escudo frente a la fragilidad y vulnerabilidad que podemos experimentar como individuos y como sociedad con relación a la IA”.

En Europa existe la **Lista de Evaluación de la Inteligencia Artificial Confiable (ALTAI)**, una especie de guía creada en 2020 para evaluar la confiabilidad de un sistema de IA en desarrollo. La lista incluye siete puntos entre los que se encuentra **la protección de la autonomía del ser humano**. “Aunque sean ayudados por una IA, los seres humanos deben conservar la autonomía de debatir las sugerencias de la IA y decidir finalmente



por ellos mismos”, apunta Sara Degli-Esposti. También indica que **los sistemas de IA han de ser resistentes y seguros**, es decir, garantizar un plan de contingencia que permita la continuidad de las operaciones en caso de que surjan problemas inesperados y evitar que la IA dañe algo o a alguien.

La **protección de la privacidad y el buen gobierno de los datos** es otro de los puntos de ALTAI. El diseño de sistemas de IA **que no sean “cajas negras”** y que, por tanto, se conozcan sus capacidades y limitaciones, **que no reproduzcan prejuicios sociales** (los famosos sesgos), que promuevan **el bienestar social y ambiental**, y que permitan establecer **mecanismos para rendir cuentas** para así identificar responsabilidades y sanciones, en caso de que se genere algún perjuicio o daño como consecuencia del uso de un sistema de IA, también integran estos principios generales.

Medidas como ALTAI son una buena base, pero Sara Degli-Esposti advierte que “un debate centrado únicamente en principios abstractos no permite entender la materialidad, y con ella, la complejidad, los conflictos y las inercias asociadas a la producción de la IA”. Porque, **¿cómo puede evitarse que los sistemas de IA reflejen la discriminación, los prejuicios y las injusticias sociales existentes a partir de sus datos de entrenamiento?**, o **¿cómo proteger la intimidad de las personas**, si los datos personales se recogen y analizan con tanta facilidad? Para responder a estas cuestiones, hay quien propone elaborar **catálogos de responsabilidades y deberes** por parte de las entidades y profesionales que están detrás de la IA. Por su parte, la autora defiende que **“el desarrollo ético de la IA es facultad y responsabilidad de todos los actores involucrados**, desde los mismos diseñadores y desarrolladores de estas tecnologías hasta sus usuarios, porque con sus datos y su uso contribuyen a su perfeccionamiento”. En la primera línea de ese “todos” se sitúa la **comunidad investigadora**, pues “deben tener en cuenta que los riesgos para los sujetos se sopesan contra los beneficios para la sociedad, y no en beneficio de los investigadores individuales o de los propios sujetos de investigación”, advierte.

IA generativa: ¿es necesario pisar el freno?

La IA generativa, cualquier sistema de IA cuya función principal sea crear contenidos, permite, entre otras aplicaciones, crear los famosos *deepfake*, es decir, vídeos o audios falsos, manipulados digitalmente, producidos mediante aprendizaje automático con redes neuronales profundas. Su utilización puede llevarnos a contenidos más o menos ingeniosos o divertidos, incluso a **promover causas humanitarias**. Por ejemplo, la tecnología de doblaje desarrollada por la empresa de IA Synthesia ha permitido a David Beckham ‘hablar’ nueve idiomas para una campaña de concienciación contra la malaria.

Pero **el potencial uso malévolo de esta tecnología es enorme**, ya que se puede utilizar para tergiversar las declaraciones de figuras públicas, para suplantación de identidad

con fines lucrativos o para inicios de sesión controlados por voz, entre otras aplicaciones cuando menos cuestionables moralmente. De hecho, **“se ha demostrado que la IA generativa puede incurrir en un plagio, en infracciones de derechos de autor o en la creación de contenidos nocivos** que se pueden usar en campañas de desinformación”, comenta Sara Degli-Esposti.

En muy poco tiempo **los modelos de IA generativa han alcanzado un grado de sofisticación tan alto** que hasta preocupa a algunos de sus creadores. Es el caso de **Geoffrey Hilton**, vicepresidente e ingeniero de Google hasta mayo de 2023, cuando abandonó la empresa y expresó públicamente su preocupación por el peligro potencial de la IA para lanzar **campañas de desinformación**. Y parece que esta alerta es compartida. “En una carta abierta publicada el 22 de marzo de 2023 y firmada por más de 30 mil expertos se hace un llamamiento a todos los laboratorios de IA para **poner en pausa durante al menos 6 meses el entrenamiento de sistemas de IA** más potentes que GPT-4 con el fin de dedicarse a la investigación técnica sobre la seguridad de la IA y promover el desarrollo de sistemas de gobernanza que incluyan un ecosistema de auditoría y certificación que reduzca los riesgos de manipulación de esta tecnologías”, explica Degli-Esposti.

Nuevamente, la investigadora del Instituto de Filosofía pone el foco en la ciudadanía, porque **“la rapidez de la evolución de la IA generativa se debe en gran medida a usuarios como nosotros dispuestos a jugar con estos sistemas y a contribuir al desarrollo de los mismos”**. Muchos se ponen a disposición de las personas para que interactúen con ellos mientras operan. Y no se trata de un gesto de generosidad por parte de las empresas, sino de una invitación a participar en un experimento social de mejora de producto. “En el caso de los *chatbots*, por ejemplo, las preguntas que introducimos (conocidas como *prompt*) también son datos que contribuyen a entrenar el modelo”, añade.

Normativa para una IA en permanente cambio

El marco normativo que regula la economía de los datos y de la IA está en constante evolución, como la propia IA. En este contexto, desde Europa se habla del **“efecto Bruselas”** para referirse a que la incorporación de una regulación estricta en la Unión Europea puede incentivar la innovación en empresas europeas que respeten la normativa, y así crear oportunidades de exportación y de convertirse en referentes mundiales.

“La **Ley de Inteligencia Artificial** es un reglamento propuesto por la Comisión Europea que pretende regular el desarrollo, despliegue y utilización de la inteligencia artificial, la robótica y las tecnologías conexas dentro del mercado único europeo”, indica la investigadora. En relación a estas cuestiones, nuestro país parece bien posicionado,



pues España contará con la primera **Agencia Española de Supervisión de la Inteligencia Artificial** (AESIA) con sede en A Coruña.

Por su parte, en Estados Unidos se ha redactado un Anteproyecto de Carta de Derechos de la Inteligencia Artificial aplicado a sistemas automatizados que tienen el potencial de afectar significativamente a los derechos, oportunidades o acceso del público estadounidense a recursos o servicios críticos.

Tras presentar los muchos ejes que conforman esta mirada ética de la IA, Sara Degli-Esposti plantea una pregunta final: **“¿antes de empatizar con los androides no sería oportuno empatizar con los demás seres vivos que tenemos alrededor y cuya vida (y calidad de vida) depende tanto de nuestras decisiones?”**. Tal y como concluye la experta, “saber que, aún en 2022, según Naciones Unidas, unos 50 millones de personas vivían en condiciones de esclavitud moderna —28 millones en trabajos forzados y 22 millones en matrimonios forzosos—, puede ayudarnos a entender por qué dentro de la academia algunas personas consideran el debate sobre la autonomía de la IA un espejismo o un ejercicio de narcisismo típico del primer mundo”.

[La ética de la inteligencia artificial](#) es el número 150 de la **colección ¿Qué sabemos de?** (CSIC-Catarata). Para solicitar entrevistas con la autora o más información, contactar con: comunicacion@csic.es (91 568 14 77).

Sobre la autora

Sara Degli-Esposti es investigadora científica en Ética e Inteligencia Artificial en el Instituto de Filosofía del Consejo Superior de Investigaciones Científicas. Es coinvestigadora principal del proyecto “Ingeniería de valores en sistemas de IA – VAE” y Ethics Advisor del proyecto H2020 NEST. Su investigación se centra en inteligencia artificial fiable y confiable, riesgos de privacidad de las tecnologías digitales y cumplimiento normativo con las leyes de protección de datos en el contexto del capitalismo de la vigilancia.

CSIC Cultura Científica