

Luces, sombras y riesgos de la inteligencia artificial

SANTIAGO ÁLVAREZ CANTALAPIEDRA

La inteligencia artificial (IA) está en nuestras vidas de forma silenciosa y apenas visible. Se encuentra en las plataformas desde las que adquirimos productos o contratamos servicios, en las redes sociales, en los buscadores y traductores de uso cotidiano, en los asistentes virtuales que prestan ayuda en los hogares, en los robots que actúan como sirvientes domésticos, en algunos termostatos que regulan la temperatura de la vivienda, en los electrodomésticos programables a distancia, en las pulseras que miden los signos vitales, en el *smartphone* del que no nos despegamos ni en las horas de descanso. Nos acompaña de forma imperceptible a pesar de determinar profundamente nuestro modo de vida y la reproducción de las dinámicas sociales y relaciones de poder.

La proliferación de espacios digitalizados hace que vivamos con naturalidad en ámbitos crecientemente artificializados. Esa dificultad de percibir el medio en el que nos desenvolvemos es un poderoso obstáculo que impide pararnos a pensar en las implicaciones ecológicas, sociales, económicas y políticas que la digitalización conlleva y, más aún, las específicas que se desprenden de complejas tecnologías que –como el *big data* y la IA– impulsan su desarrollo.

Introducción

Un misterio encerrado en un arcano

Otra dificultad para evaluar las consecuencias de la IA viene de la mano de su complejidad y opacidad. No resulta evidente identificarla y, mucho menos, definirla y conocer cómo funciona. Aunque se viene desarrollando desde mediados del siglo pasado, la mayor parte de la población empieza a ser consciente de su presencia y uso masivo con la noticia del lanzamiento del ChatGPT, cuando se situó el asunto en el centro de la conversación social. Desde entonces, raro es el día en que no aparezca alguna noticia, crónica o artículo sobre el tema, pero en lo fundamental sigue siendo un arcano indescifrable para la ciudadanía.

Una definición básica de IA hace referencia a sistemas y programas que pretenden imitar las capacidades cognitivas propias de los humanos a partir de la suposición de que todas esas funciones (como el aprendizaje, el cálculo, la racionalidad instrumental, la percepción, la memorización e incluso la investigación científica y la creatividad artística) pueden describirse con tal precisión que resultaría posible programar un ordenador para que las replicara.

El desarrollo de la matemática de la comunicación y de la economía de los datos ha permitido que la información se convierta en una unidad cuantificable que no dependa de su contenido para ser transmitida. La informática que ofrece dispositivos de entrada y salida (*inputs/ outputs*), unidades de procesamiento (mecanismo de control) y memorias (unidades de almacenamiento de las unidades de control), unida al desarrollo de los algoritmos y las teorías de redes, han hecho posible que, reunidos los conocimientos y los avances de esos campos, surgiera la inteligencia artificial. Los avances en la neurociencia han ofrecido espejos en los que inspirarse. La IA basada en modelos neuronales consiste en superponer capas de neuronas artificiales: «Cada neurona recibe muchos *inputs* (aportaciones) de la capa inferior y emite un único *output* (producto resultante) a la capa superior, como hacen las muchas dendritas y el único axón de las neuronas naturales. Capa a capa, la información se va haciendo más abstracta, como ocurre en nuestro córtex (corteza cerebral) visual».¹ Lograr el reconocimiento de imágenes, la interpretación del lenguaje hablado o la generación de textos coherentes requiere ingentes cantidades de datos y gran potencia de computación que solo se ha alcanzado en

¹ Javier Sampedro, «Inteligencia artificial: el miedo no es un argumento», *El País*, 5 de enero de 2024. Disponible en: <https://elpais.com/proyecto-tendencias/2024-01-05/inteligencia-artificial-el-miedo-no-es-un-argumento.html>

tiempos recientes. Estas circunstancias son las que han permitido la aparición de la IA generativa, cuya función principal es generar contenidos en forma de texto, imagen, audio o video indistinguibles de los generados por un ser humano.

Resultados espectaculares acompañados de preocupantes sombras

Pero estos logros espectaculares vienen acompañados de nuevas opacidades que cubren con un misterio aún mayor el que provoca en el entendimiento de la ciudadanía de a pie la complejidad propia del funcionamiento de la IA. En efecto, los modelos de lenguaje extenso (*large language models*, LLM) que posibilita la IA generativa requieren un laborioso entrenamiento que emplea, como hemos dicho, una gran cantidad de datos cuyo manejo no está claro que cumpla un criterio mínimo de transparencia capaz de garantizar los derechos de los autores o el consentimiento en la entrega de la información de los usuarios digitales con los que se alimenta al sistema.

Otra cuestión tiene que ver con la fiabilidad de los contenidos generados. El sistema se entrena con información y a sus respuestas se les exige coherencia y credibilidad, pero poco más. Por eso hay que tener presente, como sostiene el profesor de ética de la Universitat Ramon Llull, Xavier Casanovas, «que la IA ni sabe por qué sabe lo que sabe ni entiende lo que dice o responde. Su respuesta busca simular el lenguaje humano y (...) puede dar lugar a un giro epistémico definitivo en nuestra escala de valores, en el cual la verdad deja de tener importancia y lo que cuenta es la verosimilitud, es decir, la apariencia de verdad».² Situación que se agrava si, además de pretender utilizar la IA para avanzar en el conocimiento, aspiramos también a obtener recomendaciones y a automatizar decisiones. Lo señala el profesor de ciencias políticas Joan Subirats: «Estamos entrando en escenarios en los que los sistemas computacionales más avanzados son capaces de decirnos qué tenemos que hacer, qué alternativa de acción es la “correcta”, o dicho de otra manera, cuál es su “verdad”»,³ para subrayar a continuación la necesidad de (re)politizar el debate tecnológico, pues ante los procesos de cambio acelerado que vivimos como consecuencia del rápido despliegue de las nuevas tecnologías de la

² Xavier Casanovas, «De la inteligencia Artificial a la inmortalidad del alma», *Papeles* núm. 269 (Suplemento del Cuaderno CJ núm. 234), septiembre del 2023.

Disponible en: <https://www.cristianismeijusticia.net/sites/default/files/pdf/papes269.pdf>

³ Joan Subirats, «¿Qué hay de nuevo en la incertidumbre actual?», *La maleta de Portbou*, núm.º 49, 2021, p. 16.

información no podemos dejar de incorporar de manera sistemática la incertidumbre y la complejidad al debate social y a la toma de decisiones colectivas. Asumir ambas cosas en el proceso colectivo de toma de decisiones exige considerar la pluralidad de subjetividades, la combinación de diferentes “inteligencias”, la contraposición de distintos intereses, perspectivas e ideas y la confrontación de valores, es decir, deliberar políticamente, para así poder encontrar soluciones comúnmente aceptadas más allá de la respuesta “correcta” propuesta por una máquina que no hace sino reproducir los criterios y sesgos con los que ha sido construida.

Sesgos y desigualdades automáticas

Nos encontramos ante una de las cuestiones centrales de la proliferación de la IA: la implantación de sistemas automáticos que deciden a quién se vigila, qué familias son merecedoras de los recursos públicos, a quién se preselecciona para un empleo, qué clientes deben obtener la cobertura de un seguro, a quién se le debe conceder un préstamo o investigar por fraude.⁴ La IA contribuye a preservar y profundizar viejas formas de segregación social al mismo tiempo que crea otras maneras de discriminación de la mano de los sesgos digitales. Si el «sesgo de la automatización» consagra la discriminación al dar por válido y usar una recomendación obtenida de un sistema automatizado, el «sesgo algorítmico» ya viene incorporado en el propio mecanismo en la medida en que los datos que se utilizan para entrenar ese sistema reflejan y perpetúan injusticias ya existentes.

La reproducción de injusticias y desigualdades a través de las tecnologías nos muestran que estas no tienen nunca un carácter neutro ni una evolución autónoma al encontrarse incrustadas en una realidad social. Aunque el automatismo y los resultados inesperados que ofrece la IA puedan inducirnos a pensar en la emergencia de una realidad paralela plenamente autónoma de nuestra voluntad, lo cierto es que el desarrollo de esta tecnología está completamente integrado en las estructuras sociales, políticas y económicas de la sociedad. En este sentido, la IA no hay que verla únicamente como el producto de un proceso tecnocientífico, sino también como el resultado de la red de interconexiones de intereses, relaciones de poder, valores, normas, prácticas, costumbres y sesgos preexistentes durante

⁴ Virginia Eubanks en su libro *La automatización de la desigualdad* (Capitán Swing, Madrid, 2021) analiza innumerables casos en los EEUU donde los sistemas de elegibilidad automatizados, los modelos de predicción de riesgos o los algoritmos de clasificación están sirviendo, sobre todo, para supervisar y castigar a los más pobres.

el proceso de diseño, desarrollo y adopción del artefacto, de manera que su funcionamiento suele generar resultados discriminatorios por motivos de género, origen étnico o clase social. A su vez, la propia tecnología que forma parte de ese entramado de relaciones contribuye a su redefinición e influye, cuando no determina, los comportamientos en una sociedad.

El papel de la economía y de las finanzas

Se ha señalado que los programas de IA inspirados en la estructura del cerebro humano solo pudieron avanzar de forma significativa cuando dispusieron de enormes cantidades de datos y suficientes capacidades de computación para procesar esa información. Gracias a la popularización de las redes sociales y otras plataformas, los usuarios y clientes generamos diariamente los datos que el análisis automatizado de la IA convierte en una ventaja competitiva para las empresas. De esa apropiación de datos surgen nuevos modelos de negocio sobre los que asientan su poder las grandes tecnológicas como Meta, Amazon o Alphabet. Un modelo que tiene como objetivo captar permanentemente nuestra atención para mantenernos eternamente enganchados a los dispositivos digitales que permiten poder cumplir adecuadamente con nuestro papel de suministradores de información. Este imperativo extractivo eleva las operaciones de suministro a la categoría de función definitoria del llamado «capitalismo de la vigilancia».⁵

Pero para entender el poder de las grandes empresas tecnológicas es preciso examinar también cómo se han financiado. Pocas empresas del sector han crecido sin el impulso inicial de los fondos de capital riesgo. A fin de cuentas, el poder digital es parte de la economía financiarizada.⁶ Aunque venía desarrollándose desde la segunda mitad del siglo pasado, el auge y despliegue de la IA solo ha podido darse ahora, cuando las finanzas ocupan un papel central en el funcionamiento de la economía.

Implicaciones en todos los ámbitos

El aprendizaje automático es la parte de la IA que mayor influencia va a tener en nuestras vidas, ya que al reconocer patrones de datos permite anticipar, manipular

⁵ Shoshana Zuboff, *La era del capitalismo de la vigilancia*, Paidós, Barcelona, 2020.

⁶ Nils Peter, «Control económico: el papel de la financiación en las grandes empresas tecnológicas», *Estado del poder 2023*, TNI/ FUHM/ CLACSO, 2023. Disponible en castellano en la página web de FUHEM: <https://www.fuhem.es/2023/11/23/control-economico-el-papel-de-la-financiacion-en-las-grandes-empresas/>.

y monetizar el comportamiento de los usuarios, pudiéndose además aplicar prácticamente a cualquier ámbito. La educación, el mundo del trabajo, la política o el funcionamiento de las democracias se están viendo afectadas por su presencia. En el ámbito educativo, se resaltan las ventajas y los riesgos que puede proporcionar tanto a los docentes como al alumnado.⁷ En el campo científico se promete que su aplicación ayudará a acelerar la investigación de forma exponencial. En el mundo laboral se nos dice que transformará el sentido del trabajo y se remarca el impacto que podrá tener sobre el empleo y las condiciones de su desempeño.⁸

Tal vez sea en el ámbito de la política donde se observa con mayor preocupación los efectos de la presencia masiva de la IA en la vida social. Podemos recordar la intención de influir en los electores, a partir de millones de perfiles de Facebook, de la empresa Cambridge Analytica en el referéndum sobre el Brexit y en las elecciones estadounidenses de 2016 que auparon al gobierno a Donald Trump. Sin embargo, es probable que lo peor esté aún por llegar. La combinación de la IA generativa con unas redes sociales cada vez más absorbentes u otras realidades virtuales como el Metaverso auguran un escenario en el que podrán hacer acto de presencia –como ya ocurre, pero a una escala más grande y con un grado de penetración incomparablemente mayor– contenidos tóxicos y sensacionalistas que inunden y enrarezcan con confusión y mendacidad la atmósfera del debate político, profundizando la polarización y la desinformación. Lo hemos ido señalando: si la IA suprime la verdad y la transparencia, es desarrollada al servicio y bajo el control de grandes corporaciones oligopólicas y actúa reproduciendo y simplificando sesgos de género, étnicos o de clase, puede llegar un momento en el que la aspiración a una idea de democracia real e inclusiva se convierta simplemente en una quimera.

El impacto ecológico de la IA

Por otro lado, no cabe obviar su materialidad. La IA emplea grandes cantidades de recursos y energía con amplias consecuencias ecológicas. Es importante considerar toda la cadena de suministro y producción de la IA. Empieza, como todas las dinámicas extractivistas, en las minas de las que obtenemos los recursos. La

⁷ Nacho Meneses, «Estas son las tendencias y los riesgos que marcarán el desarrollo de la IA en la educación», *El País*, 5 de enero de 2024.

⁸ Daniel Susskind, «Trabajo y sentido en la era de la inteligencia artificial», *Nueva Sociedad*, núm. 307, septiembre-octubre de 2023. Disponible en: <https://nuso.org/articulo/307-trabajo-sentido-era-inteligencia-artificial/>

materialidad de la IA (como la de toda la economía digitalizada) está asociada a 17 elementos metálicos conocidos como «tierras raras», cuya disponibilidad es un factor limitante con relevantes implicaciones geopolíticas, como se ha señalado en otras ocasiones en esta misma revista. Antes de llegar a la fase de consumo está la fase de entrenamiento de los algoritmos, que conlleva una huella de carbono considerable. Continúa con la energía que exige su funcionamiento y los consumos propios de agua, energía y materiales (particularmente litio y cobre) de todas las infraestructuras que la hacen posible, como los centros de datos o las grandes instalaciones que albergan los servidores informáticos y los cientos de miles de kilómetros de cableado.⁹

Regular: ¿será suficiente?

Una investigación del Centro de Resiliencia de Estocolmo ha identificado catorce posibles “trampas evolutivas” que podrían convertirse en callejones sin salida para la humanidad.¹⁰ Somos increíblemente creativos como especie, capaces de innovar y adaptarnos a diversas circunstancias y de cooperar a escalas sorprendentemente grandes. Pero estas capacidades resultan tener consecuencias no intencionadas. El estudio muestra cómo la humanidad podría quedar atrapada en esas trampas evolutivas que se producen a partir de innovaciones inicialmente exitosas.¹¹ La IA y la robótica quedan identificadas entre esas trampas potenciales. Sea o no así, la amenaza real que se vislumbra en estos momentos tiene que ver con el tipo de mentalidad de quienes viven y gobiernan en ese particular ecosistema de Silicon Valley. Para muestra un botón. Han elaborado un indicador –con

⁹ Sobre los efectos en el mundo rural de nuestro país, véase: Ana Valdivia, «El coste medioambiental de construir minas y centros de datos para la IA en la España rural», *The Conversation*, 16 de julio de 2023. Disponible en: <https://theconversation.com/el-coste-medioambiental-de-construir-minas-y-centros-de-datos-para-la-ia-en-la-espana-rural-208385>. Una estimación reciente de la huella eléctrica e hídrica de los servicios de datos digitales en Europa en: Javier Farfán y Alena Lohmann, «Gone with the clouds: Estimating the electricity and water footprint of digital data services in Europe», *Energy Conversion and Management*, Vol. 290, 15 de agosto de 2023.: <https://doi.org/10.1016/j.enconman.2023.117225>

¹⁰ Peter Søgaard Jørgensen et al., «Evolution of the polycrisis: Anthropocene traps that challenge global sustainability», *Philosophical Transactions of Royal Society of London B*, 379, The Royal Society Publishing, 2023. <https://doi.org/10.1098/rstb.2022.0261>

¹¹ La simplificación de los cultivos o la generalización de los antibióticos son algunos ejemplos de esas trampas evolutivas. La implantación de modelos agrícolas simplificados que permiten pocos cultivos altamente productivos (como el trigo, el arroz, el maíz o la soja) ha significado que las calorías producidas se hayan disparado durante el último siglo, pero con la contrapartida de depender de un sistema alimentario altamente vulnerable a cambios ambientales (como los fenómenos meteorológicos extremos o las nuevas enfermedades) o a riesgos en las cadenas de suministro. Otro ejemplo es la resistencia a los antibióticos, que mata anualmente a más de 1,2 millones de personas (más que el sida o la malaria).

nombre P(doom)– de la probabilidad de que la IA acabe con la civilización y la especie humana. Dado que no es una medida científica sino únicamente una forma de volcar la percepción que tiene esa gente del riesgo de la actividad que llevan entre manos, lo que verdaderamente muestra es la calentura vanidosa que padecen como narcisistas deseosos de imaginarse capaces de provocar el fin del mundo.¹² De ser así, que la Unión Europea sea la primera región del mundo en regular los usos de la IA no dice gran cosa mientras no se adopte un enfoque que aborde desde la interseccionalidad las discriminaciones sexistas, racistas y de clase que comportan el desarrollo de esta tecnología, la materialidad de los procesos desde la «cuna hasta la tumba» y la complejidad de las tramas económicas e institucionales que distribuye desigualmente oportunidades y riesgos, beneficios y perjuicios, entre uno y otros. Como acierta a señalar Sara Degli-Esposti, «romper estas prácticas requiere un cambio social profundo que no puede materializarse simplemente en un código de buenas prácticas».¹³

¹² Jesús Díaz, «Una nueva estimación del peligro de la inteligencia artificial para el futuro de la humanidad», *El Confidencial*, 13 de diciembre de 2023. Disponible en: https://www.elconfidencial.com/tecnologia/novaceno/2023-12-13/p-doom-apocalipsis-inteligencia-artificial_3791208/

¹³ Sara Degli-Esposti, *Lá ética de la inteligencia artificial*, CSIC/Catarata, Madrid, 2023, p. 44.