

# Esto no es inteligencia

Una nueva filosofía de la IA

**LUCIANO FLORIDI**

TRADUCCIÓN DE JAVIER ANTA

*Esto no es inteligencia. Una nueva filosofía de la IA*

Título original: *La differenza fondamentale*

© 2025 Mondadori Libri S.p.A., Milano

© Traducción: Javier Anta

Este libro ha sido traducido gracias a una subvención del Ministerio de Asuntos Exteriores y de la Cooperación Internacional de Italia.

© de esta edición, Shackleton Books, S. L., 2026

Shackleton  
— b o o k s —

   @Shackletonbooks  
shackletonbooks.com

Realización editorial: Ailen Abdala

Diseño y maquetación: Ana Montero

Los créditos relativos a las imágenes se indican al final de cada pie de foto, a menos que se trate de fotografías de dominio público.



ISBN: 978-84-1361-795-4

Depósito legal: B 16126-2026

Impreso por ELCOGRAF (Italia).

Reservados todos los derechos. Queda rigurosamente prohibida la reproducción total o parcial de esta obra por cualquier medio o procedimiento y su distribución mediante alquiler o préstamo públicos.

*«Todos sabían lo laborioso que es el método habitual para alcanzar las artes y las ciencias; mientras que, mediante su invención, la persona más ignorante, por un precio razonable y con un poco de esfuerzo físico, podía escribir libros de filosofía, poesía, política, derecho, matemáticas y teología, sin la menor ayuda del genio ni del estudio».*

— JONATHAN SWIFT, *Los viajes de Gulliver*



# Contenido

|                                                                                                                 |     |
|-----------------------------------------------------------------------------------------------------------------|-----|
| PREFACIO                                                                                                        | 9   |
| PRIMERA PARTE — AGENCIA ARTIFICIAL                                                                              | 13  |
| Capítulo 1 – IA: el imperativo ético                                                                            | 15  |
| Capítulo 2 – La expectación por la IA: una nueva burbuja tecnológica                                            | 31  |
| Capítulo 3 – Más allá de la expectación: separando los hechos de la ficción                                     | 57  |
| Capítulo 4 – El cortocircuito: máquinas como mentes<br>y mentes como máquinas                                   | 97  |
| Capítulo 5 – Repensando la IA como agencia sin inteligencia                                                     | 115 |
| Capítulo 6 – Una conjetura sobre cómo equilibrar certeza<br>y alcance en la IA simbólica y generativa           | 159 |
| SEGUNDA PARTE — NOTAS SOBRE UNA CULTURA POST-IA                                                                 | 177 |
| Capítulo 7 – Repensando la movilidad: una transformación basada en la IA                                        | 179 |
| Capítulo 8 – El engaño en la era de la IA                                                                       | 195 |
| Capítulo 9 – La utilidad de la IA: usando la tecnología para el bien común                                      | 215 |
| Capítulo 10 – El reto de la empleabilidad tras la IA                                                            | 229 |
| Capítulo 11 – El liderazgo de la UE en la IA confiable:<br>la ventaja del código abierto ( <i>open source</i> ) | 261 |
| Capítulo 12 – La era del borrado: IA y la sobreabundancia de información                                        | 279 |
| Capítulo 13 – La era post-Vitruviana: IA, contenido y creatividad                                               | 295 |
| Capítulo 14 – Hiper-persuasión: el poder manipulativo de la IA                                                  | 313 |
| Capítulo 15 – El control de lo digital: el giro hacia el <i>hardware</i>                                        | 331 |
| Capítulo 16 – Realidad: el eclipse de lo análogo                                                                | 343 |
| Capítulo 17 – Una cultura del <i>proxy</i>                                                                      | 375 |
| AGRADECIMIENTOS                                                                                                 | 387 |
| LISTA DE ABREVIACIONES Y ACRÓNIMOS                                                                              | 389 |
| LISTA DE IMÁGENES                                                                                               | 391 |
| LISTA DE TABLAS                                                                                                 | 395 |
| BIBLIOGRAFÍA                                                                                                    | 397 |



## PREFACIO

La inteligencia artificial (IA) forma parte de la revolución digital que está transformando nuestras vidas y nuestras realidades de manera tan profunda que merece calificarse de «epocal». Dado su impacto transversal y amplio, la IA también es un tema que se ha analizado y debatido extensamente durante décadas, y aún más en los últimos años. Entonces, ¿qué justifica otro libro sobre ella? Espero que el título y el subtítulo ofrezcan una respuesta convincente, o al menos una excusa razonable para contribuir a una literatura ya abundante.

Comencemos con el título. En la primera parte del libro, ofrezco una interpretación poco ortodoxa —pero espero que pronto resulte trivial— de la IA como una nueva forma de «agencia artificial», y no como algún tipo de «inteligencia». La idea es que en el futuro, más pronto que tarde, utilizaremos la expresión «IA» del mismo modo en que hoy usamos «HP»: nadie busca pezuñas ni crines para evaluar la potencia en caballos de fuerza de un motor (volveré sobre este punto). Este cambio de interpretación, de la inteligencia a la agencia, no resta nada a la enorme importancia de este punto de inflexión crucial. Al contrario, estamos ingeniando nuevas y extraordinarias formas de agencia sin inteligencia: el cielo de nuestra inteligencia es el límite de dicha agencia, y esto es tan revolucionario e impactante como si

estuviéramos creando nuevas formas de actores inteligentes. El «milagro», por así decirlo, es de la misma magnitud, pero de una naturaleza muy distinta y con otras consecuencias e implicaciones. Si interpretamos mal el fenómeno, no podremos planificar ni diseñar el futuro.

En consonancia con esta interpretación, la segunda parte —reflejada en el subtítulo— incluye varias exploraciones de las transformaciones significativas que ha provocado la agencia artificial en nuestra sociedad y cultura. No son las únicas, pero sí las que considero más influyentes. Sin embargo, este es un libro filosófico sobre el presente y el futuro de nuestra cultura pos-IA, más que especulativo, pues pretende analizar de forma realista y normativa los desarrollos de nuestra sociedad digital y de sus habitantes artificiales. El objetivo es comprender qué está ocurriendo o es probable que ocurra, y qué podría y debería ocurrir. El lector no encontrará aquí especulaciones sobre escenarios hipotéticos. Con gusto dejo eso a futurólogos de todo tipo, nuestros nuevos astrólogos. Este libro tampoco es una introducción a la IA, y solo incluye un breve capítulo sobre la ética en este campo. Sobre el primer tema existen ya algunos textos excelentes. En cuanto al segundo, he publicado previamente un libro titulado *Ética de la inteligencia artificial* (Floridi, 2024), que puede considerarse un complemento de este, aunque ambos pueden leerse de manera independiente. Idealmente, habría publicado ambos como un único volumen largo, pesado y poco lucrativo, si no fuera porque las editoriales necesitan mantener un negocio rentable.

Una breve nota sobre los capítulos antes de continuar: los capítulos comienzan con resúmenes, lo que permite a los lectores que deseen saltarse algunas secciones seguir la historia general apoyándose en estas síntesis y centrarse solo en los temas que les resulten más interesantes. Todos los capítulos han sido publicados con anterioridad, pero aquí han sido actualizados, revisados, mejorados y, en ocasiones, combinados

para crear un flujo coherente. Sigo creyendo que este enfoque es el más eficaz. Comienzo con los títulos y luego paso al mosaico para el que han sido diseñados. Los lectores familiarizados con mis trabajos anteriores quizá reconozcan la estructura y el enfoque.

Una última aclaración: no espero que lo que he argumentado en este libro —ni en ningún otro lugar— tenga un impacto significativo. Parafraseando a Hegel, lo único que aprendemos de la filosofía es que no aprendemos de la filosofía. Por ello, este libro no está escrito con la esperanza de que las cosas mejoren —lo que, desde un punto de vista epistemológico, sería poco razonable a la luz de lo que sabemos sobre el pasado y la naturaleza humana—, sino para darle una oportunidad a la esperanza por sí misma. Y eso, quizá, sea moralmente loable dado lo que podríamos querer lograr y evitar en el futuro.



## PRIMERA PARTE — AGENCIA ARTIFICIAL

En esta primera parte del libro se plantea la tesis de que la IA debe interpretarse no como una forma de *inteligencia*, propiamente, sino como un nuevo tipo de *agencia*, y que esta es intrínsecamente falible. El argumento se despliega a lo largo de seis capítulos. El primer capítulo ofrece una visión general de los principales problemas éticos que plantea la IA. Los capítulos segundo y tercero constituyen la sección crítica (*pars destruens*). En ellos, sostengo que el entusiasmo desmedido en torno a la IA está contribuyendo a la progresiva inflación de una suerte de nueva burbuja tecnológica, y que ha surgido un círculo vicioso de la suma de tal burbuja y los relatos de ciencia ficción que presentan la IA como una forma de inteligencia capaz de evolucionar hacia una ultrainteligencia y generar riesgos apocalípticos. El capítulo cuarto marca el inicio de la etapa constructiva (*pars construens*) de esta propuesta, donde se afirma que este círculo vicioso se sustenta en una retroalimentación entre, por un lado, la antropomorfización de las máquinas en el discurso sobre la IA y, por otro, la computarización de la mente en las ciencias cognitivas y neurobiológicas. Así, el capítulo quinto presenta un análisis detallado del concepto de «agencia», aclarando cómo la IA ejemplifica una nueva forma de agencia sin inteligencia. Finalmente se planteará una conjetura según la cual la IA generativa sería inherentemente falible.



## CAPÍTULO 1 — IA: EL IMPERATIVO ÉTICO

La inteligencia artificial está transformando rápidamente nuestro mundo. A medida que los sistemas de IA se vuelven más autónomos y se integran cada vez más en diversos sectores y aspectos de la vida cotidiana, algunas cuestiones éticas fundamentales como la responsabilidad, la transparencia, los sesgos y la privacidad se ven agravadas o adquieren nuevas formas. En este capítulo ofrezco una perspectiva del panorama ético actual de la IA; en él analizo la necesidad urgente de abordar los sesgos en los sistemas de IA, salvaguardar la privacidad individual, garantizar la transparencia y la rendición de cuentas y gestionar los impactos sociales más amplios de la IA en los mercados laborales, la educación y las interacciones sociales. Asimismo, subrayo el carácter global de los desafíos que plantea la IA, incluidos su impacto medioambiental y los riesgos para la seguridad, destacando la importancia de la cooperación internacional y de directrices éticas culturalmente sensibles. Además, expongo tres desafíos sin precedentes que la IA plantea: para el derecho de autor y la propiedad intelectual; para nuestra comprensión de la autenticidad, la originalidad y la creatividad, a través de la influencia transformadora de los contenidos generados por IA; y para la autonomía individual, con la «hiperpersuasión» de la IA. Profundizo en cada uno de estos temas en la segunda parte del libro. En la conclusión, enfatizo

la importancia de una vigilancia crítica continua, del diseño conceptual innovador y de los esfuerzos colaborativos entre diversos actores para abordar las complejidades éticas de la IA y configurar un futuro sostenible y socialmente deseable. El capítulo destaca el papel fundamental de la filosofía en la identificación y el análisis de los problemas más acuciantes, así como en el desarrollo de soluciones convincentes y viables, y aboga por un enfoque nuevo, comprometido y constructivo de la investigación filosófica en la era digital.

## INTRODUCCIÓN: EL CONTEXTO ÉTICO

En los últimos años (especialmente a partir de la publicación de GPT-3, el modelo de lenguaje que impulsó en un inicio ChatGPT, en junio de 2020, y el lanzamiento público de ChatGPT el 30 de noviembre de 2022) la inteligencia artificial, a menudo equiparada de manera amplia con el llamado «aprendizaje automático» (*machine learning*, o «ML»), ha emergido como una fuerza pionera y disruptiva anunciando una nueva era de capacidades y transformaciones. Su crecimiento extraordinario en cuanto potencia y alcance señala el inicio de una época caracterizada por sistemas computacionales capaces de aprender, adaptarse y operar de manera autónoma; por ejemplo, generar y analizar contenidos, tomar decisiones y controlar otros sistemas. La profunda influencia de la IA se hace sentir en virtualmente todos los sectores de la actividad humana, desde la revolución de los diagnósticos sanitarios hasta la optimización de los sistemas financieros. Está así reconfigurando industrias y economías; el empleo y la educación; los conflictos, culturas y políticas, y hasta el propio tejido mismo de las sociedades a escala global. Todo ello agrava dilemas éticos preexistentes, modifica algunos de ellos y, por supuesto, da lugar a otros nuevos. A medida que los sistemas de

IA adquieren una mayor autonomía y se integran en diversos procesos, ciertas preocupaciones éticas como la responsabilidad, la transparencia, los diversos sesgos y la privacidad adquieren nueva relevancia y alcance, y acaban por transformarse en nuevos desafíos sociales. Sin duda, estos problemas requieren una comprensión sofisticada de la intersección entre tecnología, ética e innovación en toda su complejidad. He abordado extensamente estas cuestiones previamente (Floridi, 2024), y no trataré aquí de resumir su contenido. Naturalmente, podría haberle pedido a ChatGPT que lo hiciera, aunque el propio lector también puede hacerlo. En su lugar, lo que puede resultar útil para orientarse en este complejo panorama e interpretar los siguientes capítulos de este libro es ofrecer algunos antecedentes que los sitúen en el debate en curso, junto con algunos elementos destacados que indiquen hacia dónde podría dirigirse la discusión.

## PROBLEMAS EXACERBADOS Y RENOVADOS

En un momento en que los sistemas de IA adquieren autonomía en la toma de decisiones, aprenden a mejorar su rendimiento e interactúan con los seres humanos cada vez con más sofisticación, las consideraciones éticas relativas a la responsabilidad y la transparencia se vuelven urgentes. ¿De quién es la responsabilidad cuando un sistema de IA autónomo comete un error grave, incluidas alucinaciones manifiestas? ¿Cómo podemos garantizar que los sistemas de IA sean transparentes y responsables, respeten estándares éticos y eviten causar daños? De hecho, ¿son siquiera explicables en primer lugar? Cualquier lector de este libro ya sabe que uno de los principales desafíos éticos en el desarrollo y despliegue de la IA es la prevalencia de sesgos. Los sistemas de IA entrenados con conjuntos masivos de datos suelen reflejar y reforzar

los sesgos sociales presentes en dichos datos. Llamémoslo el problema «BIBO» (por *bias in, bias out*), reformulando el conocido dicho «*garbage in, garbage out*» (o «GIGO»), que se refiere a la calidad de los datos. Con BIBO y GIGO en mente, no resulta sorprendente que algoritmos desinformados y sesgados produzcan resultados discriminatorios, afianzando aún más las desigualdades e injusticias sociales.

Por poner un ejemplo, la tecnología de reconocimiento facial ha sido criticada reiteradamente por no identificar correctamente a las personas y además hacerlo de manera claramente sesgada. No se trata únicamente de un problema técnico, sino de un fracaso ético profundo que perpetúa prejuicios sociales existentes. Para combatir los sesgos en la IA, es esencial un esfuerzo coordinado que promueva la equidad, la transparencia y la rendición de cuentas en todas las etapas del desarrollo de la IA, desde la recopilación de datos y el entrenamiento de modelos hasta la aplicación en contextos reales y la evaluación continua. Debemos adoptar medidas proactivas, como la recopilación de datos diversos e inclusivos, el control de calidad, auditorías periódicas de los sistemas de IA para detectar resultados sesgados y el desarrollo de algoritmos capaces de identificar y corregir sesgos. Esto implica también diversificar el propio campo de la IA, de modo que los equipos que desarrollan estos sistemas reflejen una pluralidad de perspectivas y trayectorias. Por supuesto, como siempre, las decisiones que realmente importan son, en última instancia, decisiones humanas.

Las implicaciones éticas de la IA se extienden asimismo a cuestiones de privacidad, consentimiento y protección de datos. Conviene recordar que nuestro pasado no es en absoluto ejemplar. Estamos acostumbrados, por ejemplo, a que las plataformas de redes sociales utilicen nuestros datos, y la vigilancia digital es hoy una realidad cotidiana para varios millones de personas. Sin embargo, el futuro se presenta aún más problemático si cabe, ya que las preocupaciones relativas a la privacidad

se intensifican en la era de la IA debido a su capacidad para recopilar, analizar y cruzar datos a una escala, velocidad y precisión sin precedentes. Datos personales que antes se mantenían más aislados y eran menos accesibles para su correlación pueden ahora combinarse para crear perfiles detallados de la vida, las preferencias, las elecciones, los gustos y las inclinaciones de los individuos. Incluso para, de alguna manera, predecirlos. Esto genera una situación de temores generalizados en torno a la vigilancia y la seguridad de los datos, especialmente en países y contextos no democráticos o iliberales. A medida que los sistemas de IA recopilan, analizan y manipulan enormes volúmenes de información personal (que a menudo controlan unas pocas empresas) surgen ciertos interrogantes sobre los derechos individuales y la protección de datos sensibles, particularmente en ámbitos como la educación, el trabajo y la salud pública. Sabemos, por ejemplo, que un sistema de IA diseñado para la atención sanitaria debería priorizar valores como la confidencialidad del paciente, el consentimiento informado y la no maleficencia. Estos valores deben integrarse en todos los aspectos del sistema, desde la forma en que se recopilan y almacenan los datos de los pacientes hasta los algoritmos que los analizan e interpretan para ofrecer sugerencias diagnósticas. Los desarrolladores y reguladores de la IA deben colaborar para garantizar que la privacidad no se sacrifique en nombre de nuestra comodidad (culpa que todos compartimos), la eficiencia de los proveedores de servicio o el control por parte de quienes ostentan el poder. Esto puede implicar el desarrollo de sistemas de IA sensibles a la privacidad; el debate sobre la ética por diseño está en expansión. También la implementación de medidas sólidas de protección de datos y la garantía de que las personas, o las instituciones que actúan en su nombre, mantengan el control sobre sus datos y sobre cómo la IA los utiliza, recordando siempre que lo que debe protegerse son las personas, no meramente sus datos.

Eliminar los sesgos y salvaguardar la privacidad exige responsabilidad y transparencia, que son también principios fundamentales de la ética de la IA. A medida que los sistemas de IA y sus procesos de toma de decisiones se vuelven más complejos, resulta cada vez más urgente garantizar su transparencia y la rendición de cuentas de quienes los operan. Por esta razón, la IA explicable «XAI» se ha convertido en un campo crucial: se dedica a hacer comprensibles las decisiones de la IA para los seres humanos. Se trata de algo esencial para generar confianza, como también garantizar la responsabilidad cuando surgen problemas, y asimismo mejorar los sistemas de IA para que funcionen de manera distinta o más eficaz (véase, por ejemplo, el «*machine unlearning*»).

La idea de responsabilidad va más allá de comprender las decisiones de la IA e incluye la responsabilidad moral y la responsabilidad jurídica que derivan de dichas decisiones. Si un sistema de IA causa un fallo o un daño, determinar quién es responsable, y en qué medida, puede resultar complejo. Los modelos tradicionales de responsabilidad pueden resultar insuficientes, ya que el «decisor» puede no ser un ser humano en el circuito o supervisándolo (*in the loop / on the loop*), sino un sistema de IA que opera de forma autónoma, influido por desarrolladores, científicos de datos, usuarios e incluso otros sistemas de IA, y moldeado por su propio proceso de aprendizaje. A modo de ilustración, con el auge de la IA militarizada y de los sistemas de armas autónomas en los escenarios bélicos (los conflictos en Ucrania y Gaza han proporcionado trágicos casos de experimentación), la atribución de responsabilidad se vuelve notablemente más difícil. Determinarla, no obstante, resulta también más crucial, y en especial cuando los resultados son impredecibles o causan daños no intencionados o injustificables. Puede que sea necesario desarrollar modelos de responsabilidad más distribuidos que incluyan a los múltiples agentes

implicados en el diseño, desarrollo, despliegue y operación de los sistemas de IA, y que aborden la llamada «brecha de responsabilidad» en el marco del derecho internacional humanitario y de los estándares éticos. De manera similar, podrían ser necesarios nuevos marcos de responsabilidad moral y jurídica que reflejen la naturaleza distribuida de la toma de decisiones en la IA. Por ejemplo, el futuro podría dar lugar a ideas innovadoras sobre lo que se entiende por responsabilidad objetiva (*strict liability*) (Floridi, 2016a).

Hasta aquí he esbozado algunos problemas éticos antiguos que se han agravado o transformado con la revolución de la IA. Sin embargo, cualquier panorama ético de la IA estaría incompleto sin considerar sus impactos sociales más amplios. Al igual que otras tecnologías en el pasado, los sistemas de IA están asumiendo tareas que tradicionalmente han realizado los seres humanos y, por tanto, esta irrupción tiene el potencial de reconfigurar de manera drástica los mercados laborales, los sistemas educativos y las interacciones sociales. La diferencia con revoluciones anteriores, como la agrícola o la industrial, es que estas tardaron milenios y siglos, respectivamente, en transformar plenamente las sociedades. En cambio, la revolución digital, incluida la IA, está transformando el mundo en el transcurso de una sola generación: la nuestra. Los cambios económicos, sociales y culturales resultantes no solo son profundos, sino que se producen a una velocidad vertiginosa. Debemos anticiparlos y gestionarlos ahora para evitar que se agraven las desigualdades y las fracturas sociales. No podemos permitirnos abordarlos *a posteriori*, cuando los costes humanos, medioambientales y financieros de corregir cualquier impacto negativo serían astronómicos.

Los «cuellos marrones» (agricultura, o *bio-ware*) y los «cuellos azules» (industria, o *hard-ware*) se vieron reemplazados con la llegada del motor. Los «cuellos blancos» (servicios, o *soft-ware*), que representan la mayor parte de la fuerza laboral en países como Estados Unidos, parecían

seguros y destinados a crecer.<sup>1</sup> Sin embargo, la revolución de la IA está automatizando los empleos de cuello blanco y podría desplazar a muchos trabajadores, generando desempleo y dificultades económicas. No se trata de un desempleo causado por falta de demanda (hay abundancia de empleos en cualquier sociedad avanzada de la información con una fuerte presencia de IA, como Japón, Corea del Sur, Singapur o Estados Unidos), sino de un desajuste entre oferta y demanda (personas que carecen de las competencias o conocimientos adecuados) y de una brecha creciente, conocida como «brecha digital», y que aquí representa la desigualdad en educación y en el acceso a los servicios de IA. Los empleos del futuro estarán vinculados a la gestión de tecnologías digitales por parte de los «cuellos verdes» (Floridi, 2023). No obstante, para tratar de evitar o mitigar efectos adversos, los responsables políticos y los líderes industriales deben considerar estrategias como reformas educativas, programas de recapitación y otras redes de protección social para la generación que soportará el mayor impacto de esta rápida revolución. En otras palabras: parte de las futuras ventajas financieras de la IA debe redistribuirse para afrontar nuestros problemas actuales.

El impacto de la IA no solo amplifica problemas digitales existentes e introduce nuevos desafíos sociales, sino que también «democratiza» (en el sentido informático del término, es decir, afecta a todos) y globaliza riesgos y problemas. Al fin y al cabo, la ética de la IA no puede limitarse a un único punto de vista cultural o nacional. Consideremos los tres ejemplos siguientes.

---

1 El uso de una categorización de los tipos de trabajos mediante colores de los cuellos de la ropa empleada en el trabajo es popular en el ámbito angloparlante. Si bien Floridi recurre al uso estándar de cuellos blancos y azules, este habla de «cuellos verdes» (expresión usualmente empleada en referencia a labores medioambientales y vinculadas a energías renovables, para los que prefiere «cuellos marrones») denotando trabajos directamente vinculados a ámbito digital y de la IA. (N. del T.)

El primero de ellos tiene que ver con el impacto medioambiental de la IA. Conciérne a la sostenibilidad de las tecnologías de IA, incluido el consumo energético y las emisiones de carbono de los centros de datos, así como a la producción y operación de los sistemas de IA y de la infraestructura digital más amplia que los sustenta. Los centros de datos constituyen el núcleo de la IA y requieren enormes cantidades de energía eléctrica para su funcionamiento y refrigeración, lo que contribuye a las emisiones de gases de efecto invernadero. La fabricación de hardware y dispositivos de IA implica procesos intensivos en recursos, como la extracción de tierras raras y el uso de grandes volúmenes de agua, lo que provoca degradación de los ecosistemas y contamina. Además, la replicación y difusión generalizadas de las tecnologías de IA amplifican estos efectos a través de su impacto acumulativo: a más dispositivos, un mayor consumo energético y más emisiones. Todo ello plantea interrogantes urgentes sobre prácticas sostenibles en la industria tecnológica, como también subraya la necesidad de soluciones de computación más ecológicas. Entre las estrategias posibles se incluyen mejorar la eficiencia energética, adoptar fuentes renovables en los centros de datos, el reciclaje y la reutilización de sistemas y dispositivos de IA, o el uso de la propia IA para alcanzar los Objetivos de Desarrollo Sostenible de la ONU (Roberts et al., 2021). Una perspectiva responsable y ética de la innovación en IA (que alinee el avance tecnológico con la protección del medio ambiente) puede y debe perseguirse para reducir la huella ecológica de la IA. La alianza entre «lo verde», que representa todos nuestros entornos naturales y artificiales, y «lo azul», que simboliza nuestras tecnologías digitales, en particular la IA, debería convertirse en el proyecto humano del siglo XXI.

He aquí un segundo ejemplo: el avance de las tecnologías de IA está generando riesgos significativos para la seguridad al aumentar la vulnerabilidad de los sistemas frente a ciberataques cada vez más sofisticados.

Actores malintencionados pueden explotar sistemas de IA para desarrollar técnicas de piratería más avanzadas; ataques de *phishing* impulsados por IA altamente personalizados y difíciles de detectar; explotación automatizada y a gran escala de vulnerabilidades de software, y uso del aprendizaje automático para vulnerar sistemas de cifrado con mayor eficacia. No solo eso, sino que además los propios sistemas de IA pueden convertirse en objetivos. Los atacantes podrían intentar manipular los algoritmos mediante técnicas como el envenenamiento de datos o los ataques adversarios, con el fin de degradar el rendimiento del sistema o provocar resultados erróneos. En ambos casos, nos enfrentamos a desafíos globales que deben abordarse teniendo en cuenta, además, la complejidad añadida de la diversidad de normas y valores culturales. Se requiere cooperación internacional para establecer directrices éticas, estándares técnicos y marcos jurídicos que respeten los distintos contextos culturales (a menudo complementarios entre sí) y que, al mismo tiempo, salvaguarden los derechos humanos universales.

Como tercer ejemplo —en este caso trágico— nos encontramos el uso de la IA en la guerra (Taddeo, 2024). Los conflictos en Ucrania y Gaza han sido testigos de un uso extensivo de la IA para el análisis de inteligencia, la selección de objetivos, la logística y la comunicación. Esta integración acelera el ritmo de las operaciones hasta el punto de reducir el tiempo de toma de decisiones humanas a meros segundos, lo que plantea profundas cuestiones éticas sobre el control humano significativo y la responsabilidad en el campo de batalla.

Sin embargo, la IA también está generando cuestiones que no tienen precedente, algunas de las cuales ya son claramente visibles. Tres de ellas resultan más urgentes que las demás; principalmente, en términos de cómo transformarán nuestra cultura y experiencia. Las esbozaré por separado, aunque están interconectadas, y volveré sobre ellas más adelante en el libro.

## PROBLEMAS SIN PRECEDENTES

### *Copyright y propiedad intelectual*

La IA plantea desafíos sin precedentes al derecho de autor y a la propiedad intelectual («PI»). Este panorama se está transformando de manera fundamental, hasta el punto de que los marcos éticos y jurídicos existentes tienen dificultades para acomodar las nuevas problemáticas. La capacidad de la IA para producir obras que antes pertenecían al ámbito exclusivo del intelecto humano cuestiona las nociones tradicionales de creatividad y autoría, dando lugar a importantes dilemas jurídicos y éticos. Tres ámbitos resultan especialmente urgentes.

En primer lugar, en el lado de los *inputs*, la IA complica la aplicación y el cumplimiento de los derechos de propiedad intelectual. Los sistemas de IA se entrenan con conjuntos masivos de datos (y pueden procesarlos y aprender de ellos) que incluyen materiales protegidos por derechos de autor, sin que existan directrices claras sobre su legalidad. Con frecuencia, estas prácticas infringen los derechos de autor existentes (o, al menos, entrañan el riesgo de hacerlo), ya que la IA puede generar contenidos sustancialmente similares a obras incluidas en los conjuntos de entrenamiento. Determinar la responsabilidad en estos casos, si recae en los desarrolladores, los usuarios o la propia IA, sigue siendo una cuestión controvertida.

En lo que respecta a los *outputs*, por otro lado, la IA genera obras—desde textos e imágenes hasta vídeos y música—que plantean la cuestión de si las creaciones producidas exclusivamente por sistemas de IA, o solo bajo supervisión humana, pueden estar protegidas por las leyes de derecho de autor, las cuales exigen tradicionalmente la existencia de un autor humano. El concepto mismo de autoría se ve así ampliado, lo que obliga a los legisladores a reconsiderar los criterios de elegibilidad para la

protección por derechos de autor. Esta problemática se ve agravada por la dificultad de rastrear y demostrar la originalidad de los contenidos generados por IA, dada la capacidad de esta tecnología para producir grandes cantidades de obras derivadas de manera fácil, rápida y barata.

La tercera cuestión, es el papel de la IA en la creación de los llamados *deepfakes* (es decir, falsificaciones digitales hiperrealistas), que plantea desafíos urgentes en la protección de los derechos individuales, como también de las marcas registradas, en la lucha contra la desinformación y en la salvaguarda de las reputaciones. También se debe prevenir la manipulación informativa y de daños potenciales al público. Abordar todos estos desafíos exige repensar los marcos actuales de propiedad intelectual para adaptarlos a la realidad de la IA. Hacerlo podría implicar varias respuestas: adoptar estándares de derecho de autor más flexibles; considerar nuevos tipos de derechos de PI específicamente diseñados para contenidos generados por IA; desarrollar acuerdos internacionales para gestionar los problemas transfronterizos de PI impulsados por la IA; incorporar marcas de agua u otras herramientas de software para verificar la autoría y la procedencia, y actualizar el marco de uso justo (*fair use*), como *Creative Commons*, en relación con los contenidos utilizados para entrenar sistemas de IA. Estos son solo algunas instancias de debates actualmente en curso. Las soluciones deberán integrarse en un marco que proteja la creatividad humana, los derechos individuales y el patrimonio cultural.

Paso ahora a dos problemas ciertamente importantes y sin precedentes históricos sobre los que nos detendremos más adelante.

### *El futuro del contenido*

Que la IA pueda producir contenidos de forma automática y de calidad cada vez más alta (de manera fácil, rápida y barata) plantea desafíos nuevos y complejos cuyas implicaciones epistemológicas, éticas y

socioculturales apenas comenzamos a comprender. Hemos visto que la creación de contenidos por parte de la IA suscita interrogantes epistemológicos sobre la autenticidad, la originalidad y la creatividad. Debemos reconsiderar y actualizar estos conceptos (junto a otros afines) a la luz de la revolución de la IA. No podemos simplemente descartarlos, como tampoco fingir que las soluciones antiguas —por exitosas que hayan sido— apenas necesiten extender su alcance para afrontar nuevos desafíos.

Desde una perspectiva ética, consideremos, por ejemplo, que procedemos de una cultura consumista, y que durante casi un siglo esta ha sobredimensionado a todas luces el valor del producto o del resultado final: no importa quién produjo una camiseta, cómo o por qué; solo se evalúan su calidad y su precio. Esta cultura se ve de un tiempo a esta parte obligada a replantearse la importancia del proceso, del origen y de la relación entre la fuente y el producto. No son solo el contenido o el estilo de una página de diario lo que puede importar, sino también (y, sobre todo) quién la escribió, por qué y cómo. Dado que puede no ser posible distinguir si un texto breve ha sido generado por un ser humano o por una IA (de hecho, así sucede o pronto lo hará), resulta profundamente relevante si quien lo escribió fue Anne Frank o ChatGPT.

Por último, ya hemos comenzado a experimentar el impacto social de la IA. La automatización de la producción de contenidos no debe subestimarse. Si bien estas tecnologías democratizan la capacidad para crear contenidos de alta calidad, también amenazan con desestabilizar industrias y medios de vida tradicionales. Y, además, amplificarán el siguiente desafío: la hiperpersuasión.

### *Hiperpersuasión*

Al ofrecer una personalización, eficiencia y capacidad predictiva sin precedentes, la IA (por ejemplo, los sistemas de recomendación) está

transformando también las dinámicas sociales, la comunicación y la vida sociopolítica. Se plantean así nuevos desafíos relacionados con la erosión de la autonomía, la manipulación de la libertad de elección y el poder de la persuasión. Esta capacidad generalizada de limitar y modelar la facultad de los individuos para tomar decisiones libres e independientes, guiando de manera sutil sus pensamientos, creencias, emociones y acciones según lo determinado por algoritmos subyacentes puede denominarse «hiperpersuasión»: la hiperpersuasión que la IA puede ejercer sobre cada uno de nosotros. La hiperpersuasión de la IA puede influir en la información, las noticias y los puntos de vista a los que las personas están expuestas. Al controlar los contenidos que los individuos consumen y con los que interactúan, los sistemas de IA pueden crear cámaras de eco y burbujas de filtrado que refuerzan creencias y prejuicios existentes, polarizando opiniones y debates, manipulando puntos de vista y, en consecuencia, dificultando el diálogo constructivo y la comprensión entre distintos grupos. La hiperpersuasión de la IA también puede sugerirnos opciones que influyen en las preferencias, desde elecciones triviales (como la próxima película que ver esta noche) hasta decisiones significativas (como la mejor escuela o el mejor empleo futuro para una persona); desde preferencias alimentarias y de estilo de vida hasta orientaciones políticas presentadas como más convincentes. Ser inteligente suele implicar estar dispuesto a cambiar de opinión; sin embargo, esta flexibilidad puede convertirse en maleabilidad cuando se explota de forma poco escrupulosa. Tal manipulación se extiende al ámbito político, donde la hiperpersuasión de la IA puede inclinar la opinión pública y el comportamiento electoral mediante la provisión selectiva o la ocultación de información; influir en elecciones a través de campañas microdirigidas y *deepfakes*; y moldear narrativas políticas y percepciones públicas, socavando en última instancia la integridad de los sistemas democráticos. Esta fragmentación de la esfera pública pone en peligro la

cohesión social y el proceso democrático, especialmente cuando la IA se utiliza para difundir desinformación o contenidos polarizantes.

### *La necesidad de la filosofía*

De cara al futuro, resulta evidente que el panorama ético de la IA seguirá evolucionando en paralelo a los avances tecnológicos. En este breve capítulo me he limitado a esbozar algunos problemas éticos bien conocidos. He destacado unas pocas cuestiones que merecen una atención especial porque su impacto aumentará de manera significativa en un futuro próximo, dando forma a nuestras sociedades y culturas. Debemos mantener no solo una vigilancia crítica constante para garantizar que nuestros marcos éticos puedan acomodar e influir en los nuevos desarrollos, sino también ser imaginativos e innovadores en el diseño conceptual —mi definición de la filosofía— con el fin de elaborar las mejores soluciones y ofrecer así una oportunidad real de fomentar un futuro sostenible y socialmente preferible. Todo esto es alcanzable. Sin embargo, el análisis de los problemas y la síntesis de sus soluciones tienen más probabilidades de éxito si implicamos a un amplio espectro de actores (desde tecnólogos y especialistas en ética hasta responsables políticos y la ciudadanía) en la tarea de abordar las complejidades morales de la IA con lucidez y tolerancia. Esta es una noticia alentadora para la filosofía: existe un amplio margen para realizar un trabajo conceptual excelente y crucial de manera colaborativa, fundamental e innovadora. Necesitamos con urgencia una nueva filosofía de nuestro tiempo y para nuestro tiempo, que no se limite a excavar, deconstruir, analizar, sospechar o desvelar problemas antiguos y nuevos, sino que también produzca, construya, articule, justifique, respalde y por último defienda soluciones novedosas. La filosofía es desesperadamente necesaria para identificar y analizar los desafíos más acuciantes que plantea la revolución digital y para idear respuestas convincentes y viables. Existe, no

obstante, una advertencia final. Para que esto ocurra, la filosofía debe salir de su torre de marfil. En este sentido, espero que los capítulos siguientes contribuyan a romper el escolasticismo actual y a «dejar entrar la luz», como diría Leonard Cohen.

## CAPÍTULO 2 — LA EXPECTACIÓN POR LA IA: UNA NUEVA BURBUJA TECNOLÓGICA

En el capítulo 1 expuse algunos de los desafíos éticos más urgentes que plantea la IA. En este capítulo sostengo que el entusiasmo desmedido en torno a la IA forma parte del ruido que impide abordar dichos desafíos de manera eficaz. Este entusiasmo presenta las características propias de una burbuja tecnológica, estableciendo paralelismos con cinco burbujas tecnológicas anteriores: la burbuja de las puntocoms, la burbuja de las telecomunicaciones, la burbuja tecnológica china, el auge de las criptomonedas y la burbuja de las acciones tecnológicas. El ciclo de entusiasmo en torno a la IA comparte con ellas rasgos clave, entre los que se incluyen la presencia de una tecnología potencialmente disruptiva, una especulación que supera a la realidad, nuevos paradigmas de valoración, una fuerte participación de inversores minoristas y una regulación insuficiente. Asimismo, señalo otras similitudes más específicas, como la proliferación de *start-ups* de IA, las valoraciones infladas y las preocupaciones éticas asociadas a esta tecnología. Si bien se reconoce el potencial histórico y transformador de la IA, el capítulo aboga por una cautela pragmática, una planificación basada en evidencias y un pensamiento crítico a la hora de relacionarse con el