



## LOS DESAFÍOS ÉTICOS DE LA INTELIGENCIA ARTIFICIAL

*Jorge Enrique Linares Salgado*

**H**a causado mucho revuelo el deslumbrante desarrollo de la inteligencia artificial (IA) en sistemas informáticos y robóticos, en aplicaciones de todo tipo (que van desde lo lúdico a lo trascendente), en *chats* de información y conversación pública, así como en un ámbito creciente de aplicaciones para monitorear y controlar hasta las emociones y los estados de ánimo. A la vez, ha provocado estupor que expertos en IA, directores de empresas y creadores de esas aplicaciones hayan expresado sin ningún pudor (a diferencia de lo que acostumbran), sus dudas y temores sobre los alcances de esta tecnología y sus posibles efectos negativos en los individuos y la sociedad. Durante la industrialización contemporánea, quienes se especializan o impulsan las innovaciones tecnológicas nunca han expresado sus preocupaciones ni mucho menos sus miedos; tampoco han revelado información sobre los enormes riesgos sociales y ambientales que han implicado y que ocasionaron daños onerosos o letales desde comienzos del siglo XX. Así que esta repentina oleada de *sinceridad* de los gurúes de la IA es sospechosa, por decir lo menos, porque no parece estar motivada por la iluminación moral sino, quizá, por un objetivo de mercadotecnia.

Como examiné en mi libro *Ética y mundo tecnológico*,<sup>1</sup> el crecimiento desmedido del poder de la tecnología siempre ha requerido de una contención ético-política que establezca límites y regulaciones para prevenir y reducir sus peligros. Lamentablemente, eso no ha sucedido en

<sup>1</sup> Jorge Enrique Linares, *Ética y mundo tecnológico*, UNAM-FCE, Ciudad de México, 2008.

toda la era industrial. En cambio, han prevalecido tanto el afán de innovación como la ambición de ganancias jugosas por parte de los inversionistas. Dos ejemplos bastan: las mentiras y la publicidad engañosa por parte de las industrias tabacalera y farmacéutica, la cual fabricó opioides cada vez más nocivos y adictivos.

Gracias a la propaganda, solemos percibir este tipo de innovaciones como frutos del ingenio y la creatividad, pero olvidamos que se trata de mercancías y bienes productivos pertenecientes al capitalismo industrial. El principal objetivo de estos productos es incrementar, mediante su uso intensivo y extendido, la plusvalía de los capitales invertidos en ellos. Se dice que la IA se expandirá por el mundo mu-

cho más rápido que la internet, debido a la infraestructura informática y a las bases de datos precedentes, así como por el interés comercial que la impulsa. Su desarrollo acelerado y su uso totalitario, pues no deja lugar a alternativas, obedece a una “razón de fuerza mayor” que se impone globalmente, como diría Eduardo Nicol.<sup>2</sup> Así, los objetivos de la expansión mundial de la IA responden a valores tecnoeconómicos, y no a fines humanitarios.

La gran innovación tecnológica que cambió la historia, la internet, muy pronto fue controlada por el mismo afán de obtener ganancias

<sup>2</sup> Eduardo Nicol, *El porvenir de la tecnología*, FCE, Ciudad de México, 1972.

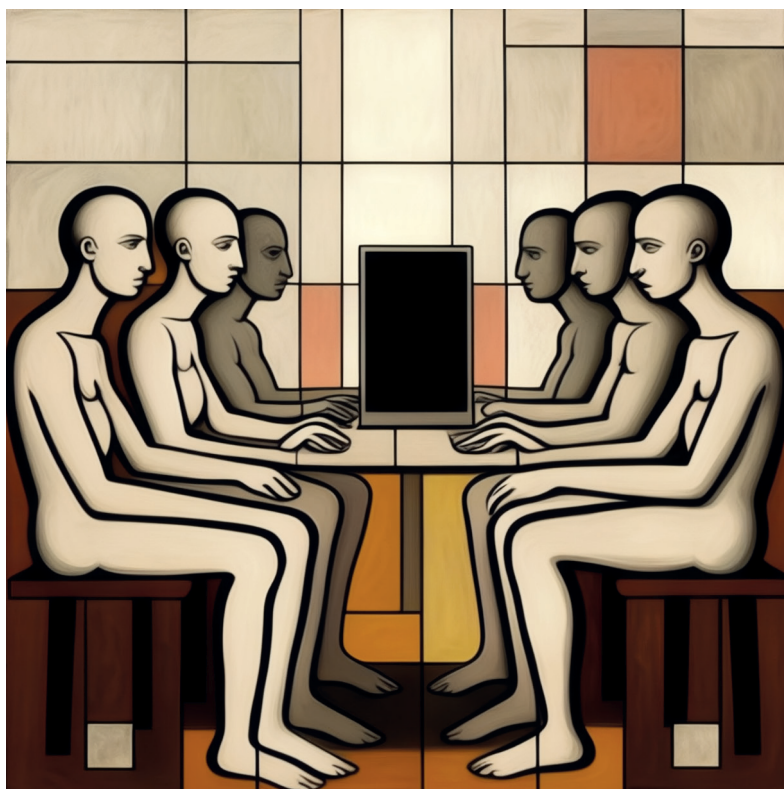


Imagen generada con Imagine.art

globales, convirtiendo a la mayoría de la población en usuarios dependientes de su uso cotidiano. Los sistemas informáticos y de telecomunicaciones que derivaron de la internet nos transformaron en una multitud de procesadores gratuitos de datos al servicio de las principales compañías, todas estadounidenses, que gobiernan el mundo digital.<sup>3</sup> Nos convirtieron en *siervos digitales* de los nuevos "señores del aire",<sup>4</sup> es decir, de esas nubes de información que no son etéreas, sino que residen en centros de datos gigantescos que consumen cantidades ingentes de energía y agua.

A pesar de todo, es innegable que los sistemas de IA pueden generar beneficios sociales al hacer que las intervenciones tecnológicas, tanto en la naturaleza como en la sociedad, sean mejores y más precisas. Si consideramos que muchas acciones tecnológicas involucran capacidades cognitivas y que sus operaciones son demasiado complejas para la mente humana, los algoritmos de la IA pueden optimizar diversos sistemas sociales (como el médico o el legal, al automatizar algunas de sus decisiones), evitar los errores y la corrupción, potenciar la producción industrial, mejorar todo tipo de servicios (desde los bancarios hasta los psicoterapéuticos), afinar la predicción de fenómenos climáticos, geológicos y biológicos (como los que causan pandemias) e incluso anticipar conflictos y revueltas sociales, pues estas tecnologías serían capaces de "observar" en tiempo real el movimiento de grupos de personas, identificar emociones en sus rostros y detectar mensajes en redes que activen alarmas por sus intenciones *peligrosas*. La IA también po-

dría automatizar la enseñanza de cualquier disciplina (sustituyendo a los frustrados profesores), perfeccionar la traducción en tiempo real de lenguas y libros, ayudar en la escritura de textos, generar y modificar imágenes y videos, y quizá incluso inducir o alterar nuestros sueños; en fin, toda una gama de actividades que requieren de la cognición compleja y que, hasta hace unos años, solo seres humanos capacitados y experimentados en esas tareas podrían realizar.

Ante el imponente avance de la IA, ¿qué será de las humanidades, la filosofía, la literatura o la narración histórica? ¿Cómo será, en unas cuantas décadas, una clase en la Facultad de Filosofía y Letras de la UNAM? ¿Habrá profesores que dialoguen con sus alumnos?, ¿sobre qué hablarán?, ¿acaso los seres humanos de carne y hueso seguiremos conversando? En general, los sistemas de IA que están a prueba (alimentados y fortalecidos por cada uno de nosotros al usarlos repetidamente) tienden a sustituir capacidades cognitivas y trabajos desahucados por humanos, lo que se denomina eufemísticamente como *automatización inteligente* de procesos y tareas, algunas ciertamente repetitivas o enfadosas. El ejemplo más célebre son los vehículos *inteligentes*. Ya en 1954 Jacques Ellul reparó en que el automóvil particular es la imagen primordial de la ideología tecnológica en torno al empoderamiento egocéntrico, además de crear la ilusión de una vida de confort y una *autonomía* ficticia, porque este invento nos hace dependientes del *sistema técnico global*.<sup>5</sup> Hoy en día los vehículos totalmente *autónomos* son una posibilidad. En el futuro próximo los humanos ya no conduciremos. La liberación de la responsabilidad individual de

<sup>3</sup> Shoshana Zuboff, *La era del capitalismo de la vigilancia*, Paidós, Barcelona, 2020.

<sup>4</sup> Javier Echeverría, *Los señores del aire: Telépolis y el tercer entorno*, Destino, Madrid, 1999.

<sup>5</sup> Jaques Ellul, *La edad de la técnica*, Octaedro, Madrid, 2003.

## Los sistemas de IA recuerdan más y mejor que los humanos; identifican objetos, imágenes y patrones en la realidad fenoménica.

conducir (mas no de los riesgos de viajar en automóvil a velocidades que pueden resultar letales) es una buena metáfora del futuro dominado por la IA: no sabremos conducir, peor aún, ni siquiera podremos dirigir nuestras propias vidas. Solo hace falta pensar en todas las habilidades motrices, cognitivas y morales que se requieren para manejar un automóvil, y reflexionar acerca de su pérdida. ¿Las ocuparemos para realizar actividades más sofisticadas que ir al trabajo o a la escuela? No lo creo: no hay otras actividades similares con las que podamos evitar esas pérdidas cognitivas. La IA puede provocar que muchas habilidades neurocognitivas, necesarias para deliberar y decidir, se atrofien por falta de ejercicio y destreza. Las habilidades que hemos ganado con las tecnologías informáticas se concentran en la reacción rápida, en la percepción visual fina, en emitir juicios emocionales inmediatos, en las capacidades motoras entre los ojos, las manos y los dedos, pero no en una profundización de las capacidades reflexivas que requieren tiempo para meditar y razonar.

Imaginemos ahora la automatización total de la enseñanza universitaria, incluidas las humanidades: no se necesitarán profesores, quizá ni siquiera un ambiente dialógico y argumentativo. Podremos aumentar nuestros conocimientos y obtener de inmediato ideas y argumentos a través de medios automatizados más eficientes que el aprendizaje en las aulas; quizá mediante dispositivos incorporados al cerebro o prótesis inteligentes en el cuerpo, a los que se envíen directamente los datos que solicitemos. Sin embargo, ¿acaso las humanidades pueden reducirse al cúmulo de in-

formación resguardada en las bibliotecas del mundo? ¿Qué sucederá con el desarrollo de habilidades para discutir, reflexionar, enseñar y aprender?

Todo parece indicar que la IA, capaz ya de hablar lenguas naturales, podría leer y escribir (lo que a fin de cuentas hacemos en las humanidades) a partir de los insumos disponibles en las bibliotecas digitales, las cuales contendrán el saber escrito de casi toda la historia (pues muchas obras antiguas no están digitalizadas hoy), y que lograrán hacerlo como polímatas con una memoria y una capacidad de aprendizaje que ni un genio como Leibniz podría igualar. La IA humanística, tras haber leído y procesado (mas no comprendido) todo el saber de estas disciplinas, podría filosofar y escribir cualquier género de literatura. Así, la inteligencia humanística podría *dejar de ser humana*. Quizá no quedará ninguna persona capaz de *filosofar o poetizar* con originalidad o profundidad. Los humanos solo se dedicarían a las cosas útiles y necesarias; en cambio, las máquinas humanísticas ejercerían la libertad de dedicarse a lo inútil. Pero no vayamos tan lejos: ¿qué sucederá con la experiencia y las destrezas humanas indispensables para el desarrollo de estas disciplinas el día en que la IA supere con creces a cualquiera que se “queme las pestañas” leyendo y escribiendo?

Mi hipótesis principal, que alienta un temor fundamental sobre los sistemas de IA, es que se propagarán en un sentido inversamente proporcional al declive de nuestra capacidad de conocer, imaginar, crear, valorar información y datos, reflexionar, dialogar, deliberar y decidir; en suma, aquellos rasgos que *nos hacen humanos* y nos permiten ejercer nuestra autonomía, es decir: pensar por cuenta propia. Es preocupante la amenaza que se cierne sobre

la autonomía, la deliberación colectiva y la prudencia política, la cual ha logrado que la especie humana sobreviva pese a siglos de conflictos y guerras. La invasión rusa de Ucrania y la israelí en Gaza nos han hecho testigos del uso de drones que, guiados por la IA, bombardean poblaciones: esta tecnología militar hace un análisis previo de sus “objetivos” en el que los humanos no participan.

Mucha gente creerá que el temor de que la IA nos reemplace es infundado. *Que eso no sucederá nunca porque las máquinas no pueden pensar.* Pero hemos comenzado a definir el pensamiento como el procesamiento algorítmico de datos e información, por lo tanto, cualquier máquina ya *piensa* (calcula, delibera, elige) mejor que

muchos de nosotros. Por lo menos, los sistemas de IA recuerdan más y mejor que los humanos; identifican objetos, imágenes y patrones en la realidad fenoménica; y pueden descubrir lo que no se ve a simple vista: lo esencial que, escondido entre las apariencias, surge mediante el procesamiento masivo de datos. Si las máquinas inteligentes ya observan y escuchan mejor que nosotros, ¿eso las hace más *inteligentes*? Quizá no, pero sí son más aptas para realizar las labores que hoy los humanos hacemos a cambio de un salario, para ganarnos la vida.

Dos problemas graves se atisban a la vuelta de la esquina: la sustitución multitudinaria de trabajadores por sistemas de IA y la disminución progresiva de capacidades cognitivas hu-



Imagen generada con Imagine.art

manas relevantes para la sociedad. Por si fuera poco, emerge otro enorme riesgo que irá expandiéndose: la IA podría no ser *tan inteligente*. Es posible que falle por el mal diseño de sus algoritmos, por el contagio de nuestros errores y sesgos cognitivos, o por malevolencia de las personas que la saboteen o la usen con fines destructivos. Un peligro adicional es que ni siquiera nos percatemos de sus fallas, o bien, que no podamos identificar dónde se produjo el desperfecto y que este termine por replicarse en una reacción en cadena. El error y el accidente son inherentes a los sistemas tecnológicos, hoy imparables.

Leibniz escribió en su *Teodicea* (1710) que, en *el mejor de los mundos posibles*, la inteligencia humana podía confiar en que la racionalidad

divina sostiene el universo. En el mundo de la IA no hace falta creer en la sabiduría o en la bondad de ningún dios, ni en la armonía preestablecida; basta con encomendarnos a la efectividad de los algoritmos. ¿Bastará con creer que la IA resolverá los problemas ecológicos y sociopolíticos en los que nos hemos entrampado? Por desgracia, la IA no parece ser la expresión de la mente divina (o del *espíritu absoluto* hegeliano), y no es tan sabia ni benevolente. Quizá la IA al fin nos revele que nuestra idea de Dios estaba equivocada: él tampoco es omnipotente ni omnisciente. **U**

Óscar Santillán, *Antimundo 00Y*, 2023.  
Dibujo retrabajado en Blender en combinación  
con MidJourney y Stable Diffusion ►



Imagen generada con Imagine.art