

LA UNESCO Y EL MARCO ÉTICO MUNDIAL PARA LA IA

Dafna Feinholz

Desde hace algunos años, la inteligencia artificial (IA) empezó a suscitar promesas y preocupaciones, pero con la aparición de ChatGPT,¹ una de sus aplicaciones más populares, el tema se volvió ubicuo. Los gobiernos del mundo comprendieron, como nunca antes, la urgencia de regular esta tecnología.

No es para menos: hoy estamos viviendo la cuarta Revolución Industrial. Todo comenzó con el desarrollo exponencial y la convergencia de una serie de tecnologías que aumentaron la capacidad de recopilación, almacenamiento, automatización y procesamiento de datos. La fuerza detrás de la IA son los datos: es el insumo que alimenta sus sistemas y lo que les permite hacer análisis más amplios (GPT-3 fue entrenado con 175 mil millones de parámetros). La diferencia entre la Revolución Industrial del presente y las anteriores es que la IA es muy invasiva: con el tiempo, se integra a la vida cotidiana y deja de ser una herramienta que uno elige usar. También es disruptiva porque puede ocasionar, para bien o para mal, cambios drásticos.

No existe una definición consensuada de la IA,² como tampoco la hay en el caso de la inteligencia humana, de modo que debemos ser cautelosos al equipararlas (los robots con forma humana son un ejemplo de esta

¹ ChatGPT pertenece a los modelos de lenguaje grande o *Large Language Model* (LLM por sus siglas en inglés). Se trata de una forma de inteligencia artificial generativa.

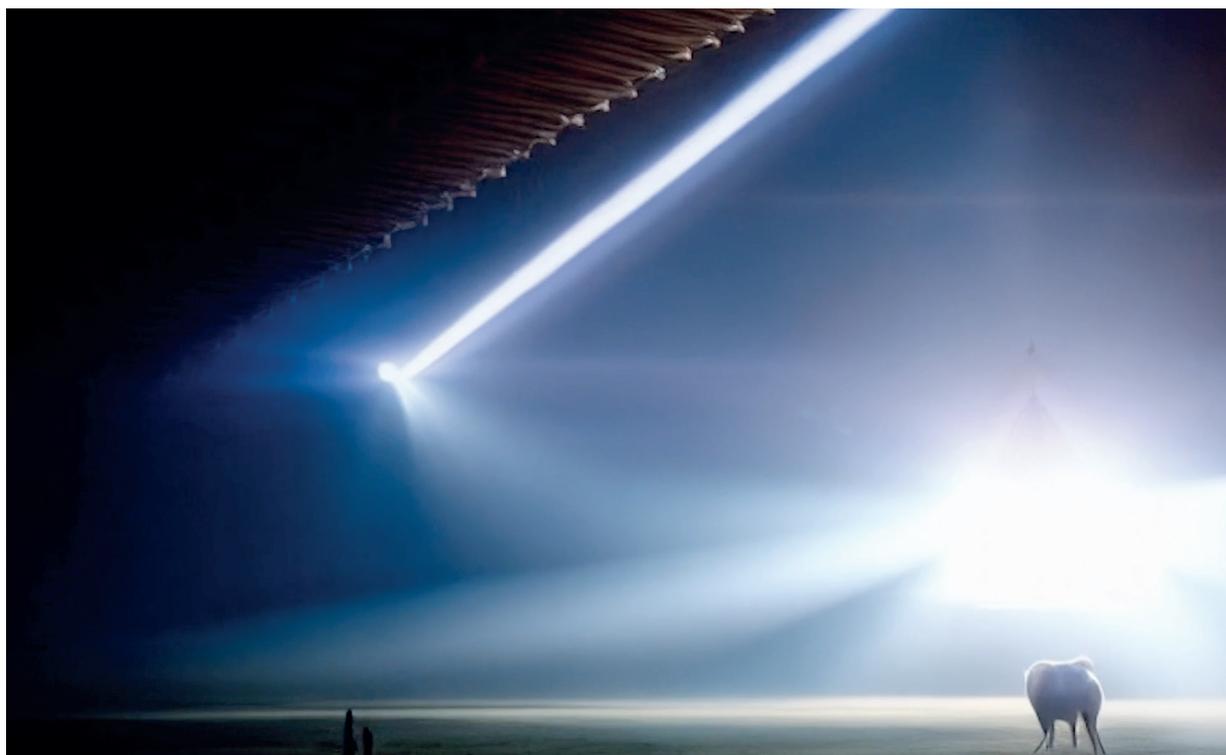
² Las definiciones de la IA dependen de aspectos técnicos e incluso de factores políticos.

confusión). Al acercarse por primera vez al tema, conviene recordar que las IA son algoritmos y programas computacionales. Es cierto que poseen una capacidad asombrosa para hacer algunas tareas infinitamente más rápido que los humanos y que, bajo ciertos parámetros, pueden realizarlas con mayor precisión. Pero al hablar de inteligencia artificial no hablamos de seres inteligentes ni de máquinas con una “gran inteligencia” sino de programas muy sofisticados. En vez de antropomorfizar a la IA, debemos conocer los verdaderos riesgos que se desprenden de su uso.

Uno de los mayores peligros que entraña es el mal uso de los datos personales y la violación de la privacidad. Basta con recordar un caso reciente. Durante la pandemia de covid-19, esta

tecnología generó algunas de las primeras alertas sobre el brote del virus; sin embargo, los sistemas de detección provocaron inquietudes. ¿Quién conservaría los datos y por cuánto tiempo? ¿Qué otros usos les podrían dar? ¿Se emplearán para mejorar los diagnósticos médicos o, por ejemplo, para medir la concentración de estudiantes y empleados, es decir, como una herramienta de vigilancia?

La desinformación y las cámaras de eco constituyen otra grave preocupación, debido a que refuerzan estigmas y propician actos de discriminación contra poblaciones vulnerables, además de que promueven la radicalización de ciertos grupos sociales. El *big data* y la IA proponen el consumo de noticias, opiniones y otros contenidos a sus usuarios con base en



Alonso Cedillo, *Panspermia*, 2023. Video generado con Gen-2 + Tortoise + MusicGEN



Alonso Cedillo, *Evidencia indiscutible de la existencia del Yeti*, 2023. Síntesis de texto a video

sus perfiles, contruidos a partir de sus compras en línea. Al usar los dispositivos digitales y aceptar las *cookies*, las tecnologías obtienen acceso a los datos que aprovechan para elaborar dichos perfiles. Así, la IA produce un efecto paradójico: por un lado, las opciones “personalizadas” reducen la oferta cultural y comercial; por el otro, su acceso inequitativo ensancha la brecha digital y reproduce la desigualdad.

El mal uso de los datos personales también representa un riesgo para la democracia, como ha ocurrido en elecciones recientes: la información seleccionada por la IA puede ser un instrumento para manipular la opinión pública con falsedades. Estas tecnologías también son una amenaza para los derechos humanos, pues han sido utilizadas para vigilar personas y comunidades.

Un par de ejemplos sirven para mostrar otro problema: los sesgos que reproduce la IA. Si un médico introduce los expedientes de sus pacientes en algún programa basado en esta tecnología para obtener un diagnóstico, la precisión del resultado no dependerá del número de expedientes sino de su diversidad. Hacer un pronóstico sobre la incidencia del paro cardíaco entre mujeres, a partir de expedientes de hombres, resultará en un análisis sesgado (los síntomas de un infarto en hombres y mujeres son muy diferentes). Y lo mismo sucede en muchos otros ámbitos.

La IA se usa cada vez más para tomar decisiones públicas y privadas, por ejemplo, para evaluar si una persona merece ser acreedora de un préstamo bancario, para contratar y ascender empleados y para dictar veredictos de inocencia o culpabilidad en los juzgados, pese a todos los sesgos que esto conlleva. Así que resulta crucial formular algunas preguntas sobre el diseño y la programación de los algoritmos. ¿Quién adquiere esta información y de dónde la obtiene? ¿Es suficientemente diverso el equipo de programadores como para elaborar algoritmos sin sesgos de género, edad, discapacidad, etnia y lugar de residencia (por ejemplo, buscan un equilibrio entre localidades urbanas y rurales)? ¿Los programas solo se nutren de información que ya es accesible? De ser así, esta contiene sesgos. De no ser el caso, ¿los datos compilados cumplen con estándares de rigor estadístico? Es bien sabido que carecemos de mucha información sobre varias poblaciones y sus lenguas; jamás se les representará de forma adecuada si esta industria no implementa medidas para remediarlo.

La incorporación de la IA a la toma de decisiones —y el alcance que esto tendrá en el futuro, considerando los sistemas de inteligencia

La Unesco se enfocó en los sistemas capaces de procesar datos e información de manera similar al comportamiento inteligente.

generativa— hace indispensable crear mecanismos que aseguren que el desarrollo y el uso de esta tecnología garantice el ejercicio de los derechos humanos.

LA ÉTICA COMO CIMIENTO DE LA REGULACIÓN

Aun antes de que la empresa OpenAI publicara, en noviembre de 2022, una versión de prueba de ChatGPT, algunos organismos intergubernamentales regionales (como la OECD y la Comisión Europea), varias instituciones académicas, numerosos países y la propia industria asumieron el deber de elaborar códigos para regular la IA. Sin embargo, cada uno lo hizo desde su perspectiva y con objetivos distintos: algunos priorizaron el desarrollo económico; otros, el fomento de la innovación tecnológica; otros más se enfocaron en la protección de los derechos humanos elementales. Los esfuerzos no pasaron de conformar una pedacería y, por si fuera poco, estas entidades tienen una membresía muy reducida.

Para el 2019 era evidente la necesidad de redactar normas globales que armonizaran el creciente número de estrategias nacionales y regionales. Entonces los representantes de los 193 países miembros de la Unesco (es decir, la Conferencia General) decidieron que esta organización era la plataforma idónea para establecer y promover un marco normativo ético de alcance mundial.

La Unesco fue elegida para hacerlo porque fue creada, en 1945, para asegurar que la ética sea el principio rector del progreso científico y tecnológico. El organismo fue fundado para fomentar la paz por medio de la cooperación internacional en la educación, las ciencias naturales y humanas, la cultura, y la comunicación y la información. Desde hace más de treinta

años, la Unesco, junto con otros órganos consultivos,³ lidera, al frente de la ONU, los programas que elaboran los únicos instrumentos normativos globales sobre ética y bioética en la ciencia y la tecnología.

Como mencioné antes, no existe una definición única de la IA. El documento sobre sus pautas normativas no debía fijar una definición, pues hay muchas y cambian conforme avanza la tecnología. En cambio, la Unesco se enfocó en los sistemas capaces de procesar datos e información de manera similar al comportamiento inteligente, lo que incluye operaciones de razonamiento, aprendizaje, percepción, predicción, planeación y control.

Ninguna tecnología es neutral y, por ello, su creación y sus usos deben basarse firmemente en un conjunto de valores fundamentales y compartidos (en particular, las declaraciones internacionales de derechos humanos). La disciplina de la ética hace explícitos los valores detrás de las decisiones, las políticas y la asignación de recursos, y los marcos éticos brindan respuestas cuando aún no existe una regulación o cuando es imposible que exista.

En este caso, la Unesco define la ética como la reflexión sistemática, dinámica, multidisciplinaria, multicultural y plural que acompaña el desarrollo tecnológico para crear la sociedad que deseamos. Al mismo tiempo, considera que los límites entre lo que es deseable y lo que no lo es deben evaluarse conforme se desarrollan las tecnologías y se manifiestan sus efectos benéficos y nocivos. La ambición

³ Por ejemplo, el Comité Internacional de Bioética (CIB) y la Comisión de Ética de la Ciencia y la Tecnología (Comest).

del documento fue abordar los aspectos de la IA que poseen relevancia ética: para qué necesitamos esta tecnología —en específico, quién la necesita—, por qué se decidió desarrollarla, a quién beneficia y a quién perjudica, cómo los beneficia y los perjudica.

Entre 2020 y 2021, la Unesco trabajó, con ayuda de redes de expertos, en la *Recomendación sobre la ética de la inteligencia artificial*. El documento establece cuatro valores que guían la redacción de leyes y el diseño de políticas.⁴ Estos se relacionan con diez principios que los desglosan y los traducen en conceptos precisos que los actores involucrados en la IA deben suscribir.⁵

La *Recomendación* no se limita a enlistar y describir valores y principios, también contiene varios capítulos sobre políticas orientadas a la acción. Por ejemplo, detalla las medidas necesarias para garantizar la rendición de cuentas, la responsabilidad y la transparencia, así como las regulaciones indispensables para garantizar el Estado de derecho.⁶

Hasta ahora, es lo más completo que se ha escrito sobre la ética de la IA. Incluso considera los riesgos de la IA generativa. Al poco tiempo demostró haber anticipado esta tecnología

y con ello evitó la obsolescencia de sus principios. La *Recomendación* es un documento excepcional porque es el único que realmente se ha redactado a partir de la ética, incorpora las perspectivas del sur global y tiene un alcance mundial.

Como una muestra de su amplitud, la *Recomendación* abarca once áreas de acción y políticas públicas. Propone, por ejemplo, una metodología de evaluación *ex ante* y *ex post* de las IA, y otra que asegure que estas tecnologías cumplan con todos los requisitos éticos y de seguridad antes de su comercialización.

Si se decide que una IA es necesaria para resolver ciertos problemas —y si en verdad no existe otra solución—, esta tecnología deberá someterse a una vigilancia ética durante todo su ciclo de vida (los procesos de investigación, desarrollo, comercialización y evaluación del impacto). Si al final se concluye que es más conveniente dejar de utilizar una IA, deberá considerarse cómo hacerlo.

Ningún otro documento normativo sobre la IA incorpora el tema de género de una manera tan contundente; por ejemplo, recomienda la asignación de fondos para asegurar la paridad entre hombres y mujeres. Acerca del medio ambiente, la *Recomendación* no solo describe cómo utilizar la IA para evitar catástrofes, sino que contiene indicaciones para reducir y prevenir los daños ocasionados por su uso.

Para conseguir que las personas sean responsables de la IA, y garantizar la reparación de daños, el documento insiste en que no se debe adjudicar personalidad jurídica a estas tecnologías. Los humanos siempre son responsables, más aún en decisiones que pueden resultar en la vida o la muerte de otros y en elecciones que tienen efectos irreversibles (un diagnóstico equivocado o una sentencia

⁴ Valores: 1. Respeto, protección y promoción de los derechos humanos, las libertades fundamentales y la dignidad humana. 2. Prosperidad del medio ambiente y los ecosistemas. 3. Garantizar la diversidad y la inclusión. 4. Vivir en sociedades pacíficas, justas e interconectadas.

⁵ Principios: 1. Proporcionalidad e inocuidad. 2. Seguridad y protección. 3. Equidad y no discriminación. 4. Sostenibilidad. 5. Derecho a la intimidad y protección de datos. 6. Supervisión y decisiones humanas. 7. Transparencia y explicabilidad. 8. Responsabilidad y rendición de cuentas. 9. Sensibilización y educación. 10. Gobernanza y colaboración adaptativas y de múltiples partes interesadas.

⁶ Las áreas que abarca son política de datos, género, desarrollo y cooperación internacional, medio ambiente y ecosistemas, salud y bienestar social, comunicación e información, educación e información, economía, trabajo y cultura.



Alonso Cedillo, *Accidente Ovni* (detalle), 2023. Imagen generada con DALL.E 2

injusta). Los responsables tienen la obligación de informar a las personas si una decisión que las afecte fue tomada por una IA, y los perjudicados deben acceder a un mecanismo de reparación de daños.

La *Recomendación* establece que se debe prohibir el uso de la IA para la vigilancia de poblaciones. También contempla la posibilidad de solicitar auditorías de los sistemas y los datos, sin obstaculizar la innovación. Los sistemas de IA deben cumplir con la trazabilidad y, en el sector privado, deben ser supervisados por oficiales éticos independientes. En cuanto a las herramientas de implementación y seguimiento, el documento propone un mecanismo de informes en el que participen la sociedad civil y los grupos marginados durante todo el ciclo de vida de estas tecnologías.

El propósito de la Unesco no es frenar el desarrollo tecnológico ni el económico, solo

quiere asegurar que ambos nos conduzcan a una sociedad en la que todos queramos vivir. Para lograrlo, es indispensable que la tecnología respete, en todo momento, las libertades fundamentales y los derechos humanos. Ahora bien, la decisión de cómo debe ser esa sociedad atañe a cada uno de los grupos y personas que la conforman, y esto requiere de un proceso de reflexión colectiva que sea muy incluyente. En este sentido, el corolario de los esfuerzos es relevante: en noviembre de 2021, la *Recomendación* fue adoptada (entre la aclamación general) por los 193 países miembros de la Unesco. En la actualidad se está implementando en más de cincuenta países.⁷ **U**

⁷ Para saber más sobre la *Recomendación* de la Unesco y el trabajo de este organismo para lograr su implementación, se puede consultar aquí: <https://www.unesco.org/es/artificial-intelligence/recommendation-ethics>.