

¿Podremos domar a la IA antes de que sea tarde?



Es el año 2045 y la inteligencia artificial (IA) es ubicua y se encuentra distribuida en miles de billones de agentes comunicados y coordinados entre sí por todo el planeta. La IA es una fuerza omnipresente y universal.

Los agentes de IA dirigen hospitales, realizan diagnósticos y cirugías; enseñan y dirigen instituciones educativas; operan aerolíneas y aviones, gestionan y conducen flotas de taxis aéreos y de automóviles autónomos compartidos; se enfrentan entre sí en los tribunales. Reinan en todos los ámbitos de la vida. Están, incluso, dentro de nuestro cuerpo e interactúan con trillones de agentes exteriores sin nuestro control.

La productividad ha alcanzado niveles sin precedentes y a diario aparecen nuevas innovaciones y productos; mientras la ciencia -y la propia tecnología- sigue acelerando a ritmos vertiginosos.

Al mismo tiempo, el mundo se vuelve más frágil e impredecible a medida que aparecen nuevas amenazas en forma de ciberarmas inteligentes y de virus modificados genéticamente. Los trabajadores ven caer su poder adquisitivo a niveles insospechados o pierden sus empleos en masa.

Hace apenas cinco años atrás, este *escenario uto-distópico -un híbrido entre la mejor utopía tecnológica y la peor distopía social-*, habría parecido un ejercicio de la ficción; hoy, parece algo menos imposible. Y si no ocurriera en veinte años, podría hacerlo en treinta o en cuarenta. ***Es irrelevante el plazo ya que la pregunta correcta no es "si esto podría suceder o no"; sino "a qué velocidad ocurrirá"; "qué grado de disrupción traerá"; "qué costos sociales acarreará"; y "qué medidas se habrán implementado" -o no- para mitigarlos.***

Amenaza a los Estado-Nación

A diferencia de transformaciones tecnológicas pasadas, la IA provocará un sismo en la estructura y en el equilibrio del poder global; al amenazar el estatus de los estados-nación como actores geopolíticos únicos. Lo admitan o no, los creadores y dueños de la IA son -hoy- actores geopolíticos relevantes, y la propiedad sobre las herramientas de IA profundizan un nuevo orden "*tecnopolar*" en el que las empresas tecnológicas podrían comenzar a ejercer el tipo de poder que antes estaba reservado sólo a una nación.

A su vez, la complejidad de la tecnología y la velocidad de su avance, hacen que sea algo casi imposible de regular para los gobiernos que se mueven en otra velocidad; *con una lentitud pasmosa. Y con cero comprensión del problema y de los riesgos.*

Por desgracia, gran parte del debate sobre la gobernanza de la IA quedó empantanada en un falso dilema: aprovechar la inteligencia artificial para expandir el poder nacional o restringirla para evitar sus riesgos.

Incluso aquellos que diagnostican de manera correcta el problema intentan resolverlo insertando a la IA en marcos de gobernanza tradicionales como, por ejemplo, los tratados de no proliferación nuclear. **Error: la IA no puede "governarse" como ninguna tecnología anterior. El desafío es otro: se debe diseñar un marco de gobernanza único para una tecnología que es -por su naturaleza intrínseca- única.**

Si la gobernanza global de la IA va a ser posible, el sistema internacional debe superar las ideas tradicionales de soberanía y sentar en la mesa de negociación a las empresas tecnológicas. Estos actores pueden no basar su legitimidad de un contrato social, el voto popular o la provisión de bienes públicos; pero sin ellos, la gobernanza de la IA no tendrá oportunidad. Este es un ejemplo de cómo la comunidad internacional necesita repensar las premisas básicas sobre el orden geopolítico.

Más rápido, más alto; más fuerte

La inteligencia artificial es diferente, tanto de otras tecnologías como en su efecto sobre el poder. No solo plantea desafíos éticos, morales, científicos, económicos, políticos y sociales; su naturaleza hiper-evolutiva también hace cada vez más difícil la resolución de cada nuevo desafío que plantea. *Esta es la paradoja del poder de la inteligencia artificial.* Poder que sigue un ritmo de progreso exponencial y que, hoy, muestra avances notables en meses; no en décadas como antes; dejándonos sin tiempo.

Los modelos de inteligencia artificial también hacen cada vez más con menos. Las capacidades de vanguardia de ayer se ejecutan hoy en sistemas más pequeños, más baratos y accesibles. Como cualquier software, los algoritmos de IA son fáciles de copiar y de compartir y, por ejemplo, sólo tres años después de que OpenAI lanzara GPT-3, equipos de código abierto crearon modelos gratis disponibles en Internet, capaces de lograr el mismo nivel de rendimiento en equipos sesenta veces más baratos.

Los futuros modelos de lenguaje a gran escala de seguro seguirán esta misma trayectoria de eficiencia, estando disponibles en forma de código abierto tan sólo dos o tres años después de que los principales laboratorios de inteligencia artificial gasten miles de millones de dólares en su desarrollo. Los riesgos son evidentes.

La IA es diferente a todo

El daño que la IA podría causar no tiene una forma específica; mientras que los incentivos para construirla (y los beneficios de hacerlo) continúan creciendo.

La IA difiere de las tecnologías anteriores en que casi toda ella puede caracterizarse como "*de doble uso*". La generalidad es el objetivo principal que muchas empresas de inteligencia artificial persiguen -adrede-, para que sus aplicaciones alcancen a tantas personas de tantas maneras como sea posible. Pero los mismos sistemas que conducen automóviles pueden conducir tanques. Una aplicación de inteligencia artificial diseñada para diagnosticar enfermedades es capaz de crear y militarizar nuevos virus y bacterias. Ya hay experiencias al respecto.

La IA podría usarse para generar y difundir información tóxica -erosionando la confianza social y la democracia-; para vigilar, manipular y someter a los ciudadanos socavando la libertad individual y colectiva; o para crear armas digitales o físicas que amenacen vidas humanas. La IA también podría destruir millones de empleos, empeorar las desigualdades existentes y crear otras nuevas inimaginables todavía; consolidar patrones discriminatorios y distorsionar la toma de decisiones al amplificar retroalimentaciones de información incorrecta; o desencadenar escaladas militares no deseadas e incontrolables que puedan conducir a la guerra. Nada de esto es ciencia ficción apocalíptica.

La inteligencia artificial no es la primera tecnología con algunas de estas características, pero sí es la primera en combinarlas todas. Los sistemas de IA no son como automóviles o aviones, que se construyen en hardware susceptible de mejoras incrementales y cuyos fallos más costosos se producen en forma de accidentes individuales. No son como armas químicas o nucleares, caras de desarrollar; difíciles de almacenar; imposibles de compartir y de desplegar en secreto. Y, al igual que Internet y los teléfonos inteligentes, la IA se propagará sin respetar fronteras de ningún tipo. Peor, todo ocurre en tiempo real a escala planetaria: una vez liberado, un modelo de IA estará en todas partes. Y solo se necesita uno "maligno o rompedor" para causar estragos. *Por esto, no se puede regular la IA de manera fragmentada. No tiene sentido regularla en algunos países y no en otros; su gobernanza no puede tener brechas.*

Tampoco está claro el marco de tiempo para estos riesgos. La desinformación en línea es una amenaza diaria; al igual que la guerra autónoma parece plausible a medio plazo. Más lejos en el horizonte se encuentra la promesa de una IA general; un punto incierto donde la inteligencia artificial supere el rendimiento humano en cualquier tarea dada, y el (admitido) peligro de que la IAG pueda volverse auto-replicante, auto-mejorante y autodirigida; más allá del control humano.

Juegos de suma cero

La competencia por la supremacía en inteligencia artificial será feroz. La IA no es solo otra herramienta o arma que puede traer prestigio, poder o riqueza. Tiene el potencial de permitir una ventaja militar, social, política y económica determinante sobre los adversarios. Así, los dos jugadores que más importan - China y Estados Unidos-; ambos ven el desarrollo de la IA como un juego de suma cero que dará al ganador una ventaja estratégica decisiva en las décadas venideras sobre el adversario. La guerra tecnológica es clara y reedita la carrera espacial del siglo pasado a una escala y velocidad nunca vistas.

Ningún sistema de gobernanza de la IA será fácil de implementar y, a pesar de todo el ruido y la charla provenientes de los líderes mundiales sobre la necesidad de regularla, no hay voluntad política para hacerlo. Es más, en este momento, todos los incentivos apuntan hacia una mayor inacción.

El siglo XXI presentará pocas oportunidades tan prometedoras y pocos desafíos tan intimidantes como los que plantea la IA. En el siglo pasado, los formuladores de políticas comenzaron a construir una arquitectura de gobernanza global que -esperaban-, estuviera a la altura de los desafíos de la época. Ahora, deben encarar la construcción de una nueva arquitectura de gobernanza para contener y aprovechar la fuerza más formidable y potencialmente definitoria de toda una nueva era. El año 2045 está a la vuelta de la esquina. No hay tiempo que perder.

