Observatorio de Humanismo Digital
AREANDINA
Fundación Universitaria del Área Andina

Los desafíos éticos de la Inteligencia Artificial

Alexander P. Springer[1]
Jefe de Misión Adjunto, Embajada de Austria - Colombia

Introducción.

La inteligencia artificial (IA) progresa a una velocidad inesperada, incluso para sus mismos creadores. Tal y como lo reporta Bill Gates, un sistema de IA como GTP de OpenAI puede responder un examen de biología de nivel universitario con preguntas para las cuales no ha sido entrenado específicamente, y obtener el grado más alto (1). La IA puede reconocer objetos en imágenes y videos, transcribir lo que se habla, traducir en simultánea, vencer a los humanos en juegos como Jeopardy, Go o el póquer, pintar al estilo de Picasso, escribir "nuevas canciones" de los Beatles, preparar documentos legales, comprar y vender acciones, conducir autos, volar drones, identificar el cáncer en los tejidos, y resolver el estado cuántico de muchas partículas a la vez. En los próximos años, se espera que la IA alcance y supere el rendimiento humano en muchas más tareas, cada vez más complejas (2).

A medida que las tecnologías de IA continúan avanzando, las preguntas éticas se vuelven más apremiantes. Muchas compañías de tecnología digital tienen información en sus páginas de internet sobre su posición ética y la justificación de sus acciones, como es el caso del famoso lema de Google "don't be evil" (3). Para nuestra sorpresa y aún mayor alarma, hace poco los ejecutivos de los principales desarrolladores, como OpenAI, Google DeepMind, Anthropic y otros laboratorios de Inteligencia Artificial, firmaron una carta abierta que advierte que los sistemas de IA podrían constituirse en riesgos tan letales como las pandemias y las armas nucleares (4).

La carta ha desatado toda suerte de especulaciones. No es para menos. Las mismas personas que hasta hace poco nos habían asegurado que estos sistemas eran perfectamente seguros, y que estaban a muchos años de adquirir la capacidad para la toma de decisiones autónoma, ahora, de repente, como atrapados por una epifanía, nos advierten sobre el apocalipsis a la vuelta de la esquina. ¿Han estado mintiendo todo el tiempo? ¿O son ellos los sucesores del físico estadounidense Robert Oppenheimer, quien participó

1. *Quiero aclarar que este artículo no representa el punto de vista oficial del gobierno de Austria, sino mi opinión privada, aunque presenta algunas políticas públicas y experiencias de nuestro país.*

con entusiasmo en el desarrollo de la bomba atómica, para luego renunciar a su cargo como director del Proyecto Manhattan, declarando que tenía las manos manchadas de sangre? (5)

Las preguntas éticas parecen relevantes solo ahora, cuando el genio ya salió de la botella. Sin embargo, cuesta creer que los más prominentes desarrolladores de las plataformas de IA, cuyos complejos problemas éticos eran bien conocidos por los especialistas, justo ahora encuentren su conciencia social y se conviertan en los denunciantes de un problema que ellos ayudaron a crear. ¿Los que ahora claman por el Estado y sus controles no son los mismos que por años se opusieron a cualquier regulación pública de sus actividades alegando que la industria podía vigilarse a sí misma? (6)

Me temo que sus intervenciones no son nada distinto que una admisión tardía de que la autorregulación de la industria ha fracasado completamente. Como en otros casos recientes, es otro intento por privatizar las ganancias y socializar las pérdidas.

Los desafíos éticos de la Inteligencia Artificial son numerosos y crecientes, y se presentan en muchos campos y aplicaciones tecnológicas. Solo por mencionar algunos ejemplos, el filósofo Byung-Chul Han ha mostrado que la digitalización y la creación de redes llevan a la minería masiva de datos y al análisis de estos datos mediante la inteligencia artificial, lo que sustituye a la esfera pública discursiva. En nombre del intercambio eficaz de la información, se está aboliendo la autonomía y la libertad individuales, y -lógicamente- la esencia misma de la democracia (7).

En este orden de ideas, la IA no será sino otra herramienta más para perfeccionar la vigilancia y el control de poderosas corporaciones anónimas sobre los individuos. Esta no es una mera especulación teórica. La empresa hiQ, especializada en la venta de productos de IA a los profesionales corporativos en recursos humanos, ha desarrollado pronósticos altamente fiables que predicen cuales son los empleados en riesgo de abandonar la compañía. Con esta información, los clientes de hiQ pueden intervenir preventivamente, tratando de retener al empleado, o despidiéndolo por considerarlo en "riesgo de huida" (8).

Las principales empresas tecnológicas están construyendo automóviles autónomos, que prometen aumentar la movilidad personal de las personas mayores y discapacitadas y salvar vidas, reduciendo los errores del conductor. Sin embargo, en una situación de emergencia, ¿un automóvil autónomo debe priorizar la vida de los pasajeros o la vida de los peatones? (9, 10). ¿Quién responde si un carro autónomo, con un pasajero humano distraído o incapaz de manejar, produce un accidente? (11) El CEO de Volvo ya ha declarado inequívocamente que "Volvo aceptará toda la responsabilidad siempre que uno de sus automóviles esté en modo autónomo" (12).

Otro desafío ético es el ya muy avanzado desarrollo de sistemas armados letales autónomos (LAWS) que potencialmente tendrán la capacidad de identificar, atacar y neutralizar objetivos sin intervención o autorización humana previa. Un LAWS suficientemente autónomo podría tomar sus propias decisiones, independientemente del diseño y las órdenes con las que haya sido programado, convirtiéndose en un arma con la capacidad de decidir por sí misma a quién matar. ¿Sería eso compatible con mantener las estructuras existentes de comando y control? (13) ¿Quién será responsable si el LAWS decide matar a civiles o a soldados que se han rendido? ¿Y si ataca un hospital o una escuela? ¿Podemos responsabilizar al diseñador o al programador (o al fabricante) por los crímenes de

guerra ejecutados por un LAWS? ¿O quizás tendremos que responsabilizar al robot mismo? ¿Cómo funcionaría esta imputación? (14). Hay expertos que advierten que, con el despliegue de armas letales autónomas, la probabilidad de conflictos armados aumentará (15). Dicho sea de paso, Austria está desde hace años trabajando, en conjunto con otros países interesados y con la CICR, en una convención internacional para prohibir completamente el desarrollo, el almacenamiento y el uso de estos sistemas (16).

La Inteligencia artificial depende integralmente de la calidad tanto de sus algoritmos como de los datos que esté manejando. Dado que los sistemas de IA esencialmente derivan su función de ellos, su disponibilidad y selección de datos de entrenamiento es la base para garantizar una toma de decisiones éticamente correcta ("garbage in, garbage out") (2). Por ejemplo, cuando utilizamos los datos disponibles en Google Translate como datos de entrenamiento para una aplicación de idioma, obtenemos resultados que discriminan contra las mujeres, ya que esencialmente todos los idiomas del mundo contienen un sesgo de género inherente (17).

El efecto discriminatorio de la IA es bastante conocido y no solamente puede perpetuar patrones históricos de segregación (genero, edad, raza), sino también crear nuevas formas de discriminación (18). El sesgo se ve claramente en los resultados obtenidos por los sistemas de reconocimiento facial basados en IA. Mientras los sistemas de reconocimiento facial ya logran una tasa de error de tan solo el 1% para rostros de hombres blancos, esta tasa de error aumenta al 7% para rostros femeninos blancos y al 12% para rostros masculinos de piel oscura. En rostros femeninos de piel oscura alcanza un 35 % (19, 20). Estos errores pueden tener efectos en el mundo real: las empresas, las aseguradoras, la policía, y los juzgados ya están usando sistemas de IA que predican riesgos y determinan resultados que afectan vidas humanas concretas, muchas veces reforzando la desigualdad social (21).

Finalmente, está el caso del preocupante incremento del uso de la IA en la fabricación de información falsa con fines estratégicos y políticos (22). Utilizando tecnologías avanzadas, estos operadores emplean la inteligencia artificial para generar propaganda política, manipular elecciones o crear historias o videos falsos sobre cosas que nunca sucedieron ("Deep fake"). Hace poco, el Comité Nacional Republicano usó inteligencia artificial para crear un anuncio de treinta segundos advirtiendo sobre cómo sería un segundo mandato del presidente Joe Biden. El video representa una serie de crisis ficticias, desde una invasión china de Taiwán hasta el cierre de la ciudad de San Francisco, ilustradas con imágenes y noticias falsas. Un pequeño descargo de responsabilidad en la parte superior izquierda advierte que el video fue "Creado con imágenes de IA" (23). Esta es la última de una sucesión de campañas de desinformación cada vez más profesionales y de alcance masivo (como en las elecciones presidenciales en EE.UU. y la campaña del Brexit en el Reino Unido), que puedan tener un efecto demoledor en los resultados (24).

Un estudio reciente del centro de pensamiento británico Chatham House muestra que las actuales políticas y prácticas sobre la IA están basados sobre suposiciones comunes pero arraigadas que no representan los intereses de todas las partes interesadas, y que contradicen un creciente volumen de evidencia. Entre estas asunciones están las convicciones que la IA es "inteligente", que se necesitan "más datos" para mejorar la IA, que el desarrollo de la IA es una "carrera" entre los Estados que invierten masivamente en el desarrollo de esta tecnología, y que la IA misma podría ser "ética". Todos estos supuestos resultan ser muy cuestionables, porque no son "neutrales" sino que responden a los intereses de algunos actores particulares -los que tienen más que perder- que al bien

común (25).

¿Qué hacemos ahora? Es evidente que, con la aceleración del mundo digitalizado, los correctivos éticos y sociales de antaño ya no funcionan, o lo hacen demasiado tarde. Lo que se necesita es una nueva comprensión básica de cómo lidiamos como sociedad con las nuevas posibilidades tecnológicas, y dónde trazar las líneas rojas de lo prohibido. Necesitamos partir de la base de que una explotación desinhibida de todas las posibilidades técnicas conducirá tarde o temprano al colapso social y ecológico con consecuencias difícilmente evaluables y ciertamente irreversibles. (7) Para recalibrar las políticas públicas sobre la IA, debemos cambiar nuestra perspectiva, reconocer a quién sirven estos supuestos incuestionables y considerar si son representativos de todas las partes interesadas, buscar nuevas suposiciones (ojalá comunes), planificar para los peores escenarios y evaluar si debe continuarse la exploración de ciertas aplicaciones de IA (auditorías ex ante) (25).

Principalmente desde la academia, se han presentado diferentes propuestas para un código de ética para la Inteligencia Artificial (26, 27). El Humanismo digital es una iniciativa apoyada por la diplomacia austriaca que propaga una visión de la sociedad en la que la digitalización se utiliza para fortalecer la autodeterminación, la autonomía y la dignidad humanas (28, 29). Está claro que necesitamos un nuevo contrato social para dar una respuesta coherente y consistente a los desafíos asociados con la masificación de los sistemas de IA. Los derechos humanos deben primar también en el ciberespacio. Si no logramos un acuerdo básico sobre la materia, estas mismas tecnologías se emplearán sin ninguna consideración ética, ni respeto para el individuo o para nuestras sociedades.

Referencias

(1) Bill Gates, "The Age of IA has begun: Artificial intelligence is as revolutionary as mobile phones and the Internet.", Blog GatesNotes, March 21st, 2023, URL: https://www.gatesnotes.com/The-Age-of-AI-Has-Begun
(2) S. Matthew Liao, "A Short Introduction to the Ethics of Artificial Intelligence", In: Ethics of Artificial Intelligence. Edited by: S. Matthew Liao, Oxford University Press 2020: 1-42.
(3) Paula Boddington, AI Ethics. A Textbook. Springer Nature, 2023, 76.
(4) Kevin Roose, "A.I. Poses 'Risk of Extinction,' Industry Leaders Warn", The New York Times, May 30th, 2023, URL: https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html
(5) Hector Rodriguez, "Robert Oppenheimer, el padre arrepentido de la bomba atómica" National Geografic Espana, April de 2023, URL: https://www.nationalgeographic.com.es/ciencia/robert-oppenheimer-padre-arrepentido-bomba-atomica_19743
(6) Alex Woodie, "Self-Regulation Is the Standard in AI, for Now", Datanami, April 17th, 2023, URL: https://www.datanami.com/2023/04/17/self-regulation-is-the-standard-in-ai-for-now/
(7) Byung-Chul Han, Infocracia: La digitalización y la crisis de la democracia, Taurus (Penguin-Random House), 2022, 66-69.
(8) Shoshana Zuboff, The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power, Profile Books, 2019, 173.
(9) Stefan H. Vieweg, "Preface", In: AI for the Good. Artificial Intelligence and Ethics, Edited by: Stefan H. Vieweg, Springer Nature, 2021, vii-x.
(10) Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan, "The Social Dilemma of Autonomous Vehicles," Science 352, no. 6293 (2016): 1573–76. URL: http://science.sciencemag.org/content/352/6293/1573
(11) Vikram Bhargava and Tae Wan Kim, "Autonomous Vehicles and Moral Uncertainty", In: Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence, Edited by Patrick Lin, Ryan Jenkins, and Keith Abney, Oxford University Press, 2017, 5-19.
(12) Citado por: Jeffrey K. Gurney, "Imputing Driverhood: Applying a Reasonable Driver Standard to Accidents Caused by Autonomous Vehicles", In: Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence, Edited by Patrick Lin, Ryan Jenkins, and Keith Abney, Oxford University Press, 2017, 57.
(13) Heather M. Roff, "The Strategic Robot Problem: Lethal Autonomous Weapons in War," Journal of Military Ethics, vol. 13, no. 3 (2014), 211-227, URL: https://www.tandfonline.com/doi/abs/10.1080/15027570.2014.975010
(14) Trevor N. White and Seth D. Baum, "Liability for Present and Future Robotics Technology," In: Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence, Edited by Patrick Lin, Ryan Jenkins, and Keith Abney, Oxford University Press, 2017, 66-79.
(15) Marcus Wagner, "The Dehumanization of International Humanitarian Law: Legal, Ethical, and Political Implications of Autonomous Weapon Systems" Vanderbilt Journal of Transnational Law, vol. 47 (2014), 1371-1424, URL: https://scholarship.law.vanderbilt.edu/vjtl/vol47/iss5/4
(16) Statement by Austria, Conference on Conventional Weapons, Group of Governmental Experts on lethal autonomous weapons systems, UN Geneva, Meeting on 6 March 2023, URL: https://conf.unog.ch/digitalrecordings/index.html?gui-

(18) Mateja Durovic and Jonathon Watson, "AI, Consumer Data Protection and Privacy", In: The Cambridge Handbook of Artificial Intelligence. Global Perspectives on Law and Ethics, Edited by: Larry A. DiMatteo, Cristina Poncibó and Michel Cannarsa. Cambridge University Press 2022: 273-287.
(19) Helmut Leopold, "Mastering Trustful Artificial Intelligence", In: Responsible Artificial Intelligence: Challenges for Sustainable Management, Edited by: René Schmidpeter and Reinhard Altenburger, Springer Nature, 2023, 133-157, en 146.
(20) J. Guynn, "Google photos labeled black people "gorillas." USA Today. July 1st, 2015, URL: www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identi fy-black-people-as-gorillas/29567465/
(21) Virginia Eubanks, Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor, New York: St. Martin's Press, 2018.
(22) John Villasenor, How to deal with AI-enabled disinformation, Washington: Brookings Institution Report, November 2020, URL: https://www.brookings.edu/research/how-to-deal-with-ai-enabled-disinformation/
(23) Shannon Bond, "AI-generated deepfakes are moving fast. Policymakers can't keep up", National Public Radio, Morning Edition, April 27, 2023, URL: https://www.npr.org/2023/04/27/1172387911/how-can-people-spot-fake-images-created-by-artificial-intelligence
(24) Berta García-Orosa, " Desinformación, redes sociales, bots y astroturfing: la cuarta ola de la democracia digital". Revista Profesional de la información, vol. 30, 6 (2021), 1-10. URL: https://revista.profesionaldelainformacion.com/index.php/EPI/article/download/86730/63051
(25) Arthur Holland Michel, Recalibrating assumptions on AI: Towards an evidence-based and inclusive AI policy discourse. Londres: The Royal Institute of International Affairs, (Research Paper of the Digital Society Initiative), April 2023. URL: https://www.chathamhouse.org/2023/04/recalibrating-assumptions-ai
(26) Paula Boddington, Towards a Code of Ethics for Artificial Intelligence. Springer International 2017.
(27) Amitai Etzioniand Oren Etzioni. "Incorporating ethics into artificial intelligence." The Journal of Ethics, vol. 21 (2017): 403-418.
(28) Julian Nida-Rümelin and Nathalie Weidenfeld, Digital Humanism: For a Humane Transformation of Democracy, Economy and Culture in the Digital Age, Springer Nature, 2022. URL: https://library.oapen.org/bitstream/handle/20.500.12657/58637/978-3-031-12482-2.pdf?sequence=1&isAllowed=y
(29) Federal Ministry of European and International Affairs, The Poysdorf Declaration on Digital Humanism: A Compass for Citizens During the Digital Transformation, June 21st, 2021, URL: https://www.bmeia.gv.at/fileadmin/user_upload/-Vertretungen/K-F_Bukarest/Blg._RE_Poysdorf_Declaration_v._30._Juni_2021__2_.pdf