Editorial

Ética en la inteligencia artificial para la investigación científica

Rab. Dr. Fishel Szlajen [1]

[1] Post Doctorado en Bioética (PUCRS), Doctor en Filosofía (UNEM), Maestría en Filosofía (Bar Ilan University), Jerusalem Fellow Graduated (Mandel Leadership Institute, Israel), Scholar Fellow on Religion and the Rule of Law (Oxford University); Rabino (Yeshivá Maalé Gilboa, Israel). Profesor titular en UBA, USAL y UNLaM.
Miembro Titular de la Pontificia Academia para la Vida, Vaticano; Miembro del Consejo Académico de Ética en Medicina, de la Academia Nacional de Medicina; Miembro del Consejo Argentino para la Libertad Religiosa; y Miembro del International Center for Law and Religion Studies, USA.
Su más reciente libro es "Ética y Políticas Públicas: biotecnología, religión, derecho y sociedad" (2025).

Fecha de publicación: 26/09/2025

Citación sugerida: Szlajen, F. Ética en la inteligencia artificial para la investigación científica. Anuario (Fund. Dr. J. R. Villavicencio) 2026;33. Disponible en: https://villavicencio.org.ar/anuario/33/editorial-etica-en.pdf. ARK: https://id.caicyt.gov.ar/ark:/s2796762x/kla0worxk

La irrupción de la Inteligencia Artificial (IA) en la investigación científica ha reformulado no sólo las técnicas de recolección y análisis de datos, sino también los paradigmas epistémicos de validación del conocimiento y la responsabilidad científica.

Entre los aspectos positivos, ya señalados por Max Tegmark, los algoritmos permiten escalar la capacidad cognitiva humana en términos de volumen y complejidad, clave por ejemplo en investigaciones epidemiológicas. Durante la pandemia COVID-19, el uso de IA en el sistema BlueDot permitió identificar brotes emergentes antes que la OMS.

Además, la inteligencia artificial permite realizar minería de datos sobre grandes volúmenes de literatura científica no supervisada, es decir, identificar patrones, correlaciones, temas emergentes o lagunas en el conocimiento. El Semantic Scholar no sólo organiza y clasifica millones de publicaciones científicas, sino que también aplica redes neuronales y técnicas de procesamiento de lenguaje natural para determinar qué artículos han sido influyentes en un campo determinado, o detectar áreas temáticas donde hay escasa producción académica pese a su relevancia. Esta capacidad optimiza la planificación de futuras líneas de investigación al señalar, por ejemplo, áreas donde hay preguntas sin respuesta o enfoques que aún no se han explorado en profundidad. Incluso permite evitar la redundancia científica o duplicación de estudios, crucial en contextos donde los recursos de investigación son limitados.

En biotecnología y farmacología, la IA también ha demostrado poder anticipar resultados experimentales. AtomNet, un sistema basado en redes neuronales convolucionales predijo qué moléculas podrían inhibir proteínas asociadas a enfermedades como el ébola o ciertos cánceres, incluso antes de que esas moléculas fueran sintetizadas químicamente. Esto representa un salto cuántico en la investigación, pudiendo ahora filtrar millones de compuestos virtuales mediante simulaciones reduciendo drásticamente el tiempo y costo necesarios para el desarrollo de nuevos medicamentos.

Sin embargo, estas aplicaciones plantean nuevas preguntas éticas sobre quién detenta la propiedad intelectual de descubrimientos mediados por IA, y cómo validar empíricamente hipótesis generadas por sistemas autónomos. Estas advertencias sobre el uso de IA en ciencia se agudizan por la opacidad de los algoritmos, fenómeno llamado por Nicholas Diakopoulos, cajas negras. Su funcionamiento interno no es accesible o comprensible, ni siquiera para los propios usuarios, debido a la complejidad de sus algoritmos o porque el código fuente y los datos de entrenamiento están protegidos por derechos de propiedad intelectual. Esta falta de transparencia plantea un problema epistemológico central en la ciencia, la imposibilidad de verificar, replicar o refutar los

resultados obtenidos por estos sistemas, lo que contraviene directamente el principio de falsabilidad de Karl Popper como criterio de demarcación científica. Un ejemplo es el IBM Watson for Oncology, sistema desarrollado para asistir en decisiones terapéuticas oncológicas que prometía personalizar tratamientos a partir del análisis de literatura médica y bases de datos clínicos. Pero sus recomendaciones no eran trazables ni reproducibles, no pudiendo entender por qué sugería ciertos tratamientos ni comprobar si los datos que usaba eran representativos o sesgados. En contextos clínicos, esta opacidad no sólo es científicamente problemática, sino también éticamente riesgosa, ya que decisiones que afectan vidas humanas no deberían tomarse sobre una base ininteligible o no auditable.

Cathy O'Neil ha advertido que muchos sistemas de IA, lejos de ser neutrales u objetivos, pueden reproducir e incluso amplificar desigualdades sociales estructurales. Los modelos de IA aprenden de datos históricos, y si estos reflejan prejuicios sistémicos, entonces los algoritmos tienden a perpetuarlos bajo una apariencia tecnocientífica de legitimidad. En sociología, diversos sistemas sobre representan características propias de grupos dominantes como la "norma" estadística, desplazando o invisibilizando los patrones propios de otros colectivos. PredPol, un software utilizado por departamentos de policía en ciudades como Los Ángeles y Atlanta, fue entrenado con datos históricos de arrestos, los cuales ya estaban sesgados contra comunidades afroamericanas y latinas. El algoritmo COMPAS para evaluar el riesgo de reincidencia en procesos judiciales, sesgaba sistemáticamente contra personas negras. El algoritmo de admisión internacional de la Universidad de Cambridge excluía sistemáticamente a postulantes de países con bajos ingresos, por haberse entrenado con datos históricos de rendimiento académico y tasas de admisión previas que, sin mediar correcciones socioeconómicas o contextuales, ni atender a méritos individuales o por alcanzar logros en contextos adversos, reflejaban sesgos acumulados por décadas.

Estos ejemplos, entre otros, traduciendo prejuicios y sesgos como regla permanente disfrazados de neutralidad matemática, demuestran que la automación que toma decisiones sobre personas no sólo debe ser técnicamente eficiente, sino también ecuánime, auditable y contextualizado culturalmente.

En este respecto, uno de los factores más críticos, aunque frecuentemente ignorados, en la aplicación de IA a la investigación científica y formulación de políticas es que los modelos de "machine learning" no distinguen entre correlación y causalidad. Es decir, estos sistemas son eficaces para detectar patrones estadísticos en grandes volúmenes de datos, pero carecen a priori de una comprensión estructural de las relaciones causales subyacentes. Esta limitación, destacada por Judea Pearl, es la incapacidad de los algoritmos para responder preguntas del tipo "¿qué pasaría si…?" o para distinguir entre meras asociaciones y causas. Por ello, Pearl propone un marco teórico de "cálculo causal" que permite formalizar inferencias causales de manera lógica y computacional evitando las conclusiones espurias. Por ejemplo, un sistema podría concluir que llevar paraguas causa lluvia, simplemente por ser ambos eventos concurrentes, sin captar que la causa común es la presencia de nubes. Esta falacia puede tener consecuencias graves en medicina donde un algoritmo que relacione la presencia de determinados medicamentos con la recuperación de pacientes podría erróneamente inferir que dicho fármaco "cura", cuando en realidad se debe a otra razón. En políticas públicas, se podría errar al asumir que ciertas medidas generan progreso económico sin haber descartado variables de confusión ni haber controlado los factores contextuales.

Aquí, el punto epistemológico es la diferencia entre predicción y explicación. Por eso, si los resultados producidos por IA se utilizan como base para intervenciones sociales, sanitarias, educativas, jurídico-legales o en materia de seguridad, sin un marco causal explícito y validado, se corre el serio riesgo de tomar decisiones ineficaces, injustas o incluso dañinas.

Luego, mientras que estas tecnologías amplían exponencialmente las capacidades humanas para detectar patrones, formular hipótesis y acelerar descubrimientos, también introducen nuevos riesgos en la producción, validación y legitimación del conocimiento. Éticamente, surge el problema de delegar funciones cognitivas críticas como la evaluación de evidencia o la toma de decisiones clínicas, a sistemas que no pueden rendir cuentas, carecen de intencionalidad moral y de cálculo causal, y operan bajo lógicas estadísticas opacas. Desde la epistemología si el conocimiento científico comienza a depender de modelos cuya lógica interna es inaccesible, se debilita el ideal ilustrado de ciencia como empresa pública, racional y verificable. Políticamente, se corre el riesgo de privatizar el conocimiento dado que los modelos de IA más avanzados son desarrollados por corporaciones que no necesariamente comparten sus algoritmos, datasets ni criterios de evaluación, creando asimetrías de poder y posiciones dominantes de quienes poseen la infraestructura algorítmica frente a quie-

nes dependen de sus resultados. Este escenario exige, por tanto, la formulación de nuevas normativas y principios ético-legales, como la no delegación, la trazabilidad algorítmica, el acceso abierto a datos científicos, y mecanismos institucionales de auditoría de IA en ciencia, para asegurar que el progreso tecnológico no desplace los fundamentos de la práctica científica ni los derechos, obligaciones y responsabilidades de quienes la protagonizan o la reciben.

Referencias

1. Angwin, J., Larson, J., Mattu, S., y Kirchner, L. (2016). Machine Bias: there's software used across the country to predict future criminals. And it's biased against blacks. ProPublica.

2. Cheong, B. (2024). "Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making". En Frontiers in Human Dynamics 6, 1-11.

3. Diakopoulos, N. (2016). "Accountability in algorithmic decision making". En Communications of the ACM 59, 56-62.

4. O'Neil, C. (2017). Armas de Destrucción Matemática: cómo el big data aumenta la desigualdad y amenaza la democracia. Capitán Swing.

5. Pearl, J. (2009). Causality: models, reasoning and inference. Cambridge University Press.

6. Popper, K. (1980). La Lógica de la Investigación Científica. Tecnos.

7. Tegmark, M. (2017). Life 3.0: being human in the age of artificial intelligence. Knopf.

8. Topol, E. (2019). Deep Medicine: how artificial intelligence can make healthcare human again. Basic Books.

9. Zhavoronkov, A. (2019). "Artificial intelligence for drug discovery, biomarker development, and generation of novel chemistry". En Molecular Pharmaceutics 15(10), 4311–4313.