Universidad de Antioquia

Facultad de Ingeniería

"El último dinosaurio no supo que se había extinguido."

(Reflexión del autor)

LOS RIESGOS OCULTOS DE LA INTELIGENCIA ARTIFICIAL Y LOS MODELOS DE LENGUAJE

Autor: Omar Alberto Torres

Docente: Julio Eduardo Cañón Barriga

Asignatura: Creatividad para la investigación

Lugar: Caucasia, Colombia

Fecha: 28 noviembre de 2025

Nota de autor

Este artículo fue elaborado por Omar Alberto Torres**,** estudiante de la Facultad de Ingeniería de la Universidad de Antioquia, como requisito académico para la asignatura *Creatividad para la investigación* orientada por el docente Julio Cañón.

Para correspondencia: omara.torres@udea.edu.co.

Resumen

El devenir histórico de nuestro mundo nos recuerda constantemente que la humanidad no ha dejado de evolucionar y que su creatividad colectiva parece no tener límites. Desde la prehistoria, el ser humano ha logrado grandes desarrollos y descubrimientos, como el dominio del fuego, la agricultura y la fundición de metales, de modo que cada innovación trajo consigo no solo un salto evolutivo, sino también la consolidación del hombre como especie dominante en la tierra y quizás en el universo. Con cada nuevo avance se ha multiplicado la creatividad humana, generando progresos a velocidades cada vez mayores. Sin embargo, este mismo impulso innovador trajo implícitamente oportunidades, mejoras y también riesgos. En este contexto, la inteligencia artificial emerge como el más reciente reflejo de la capacidad creativa del ser humano, pero también como una fuente de riesgos invisibles derivados de sus propias estructuras de diseño, sus sesgos y su complejidad interna. A través de un enfoque teórico y reflexivo, este artículo examina como los riesgos fundamentales de la inteligencia artificial surgen de la opacidad de sus modelos, la arquitectura de caja negra, la dependencia tecnológica y el control por parte de potencias extranjeras. Se advierte además sobre las amenazas que estas tecnologías representan para las libertades ciudadanas y los derechos humanos. Por tales razones puedo afirmar que, sin una comprensión ética y regulatoria adecuada, la inteligencia artificial puede transformarse en un instrumento de dominación más que de progreso. Este artículo examina como los riegos fundamentales de la inteligencia artificial surgen

por la opacidad de sus modelos, los sesgos presentes en los datos y la posibilidad de un uso malicioso de sus sistemas.

Abstract

The historical evolution of our world constantly reminds us that humanity has never ceased to evolve and that its collective creativity seems to have no limits. Since prehistory times, human beings have achieved great developments and discoveries, such as the mastery of fire, agriculture, and metal smelting. Each innovation has brought not only an evolutionary leap but also the consolidation of humankind as the dominant species on Earth. With every new advance, human creativity multiplied, generating progress at increasingly faster rate. However, this same innovative drive implicitly brought both opportunities and risk. In this context, artificial intelligence emerges as the latest reflection of human creativity capacity, but also as a source of invisible risks derived from its design structures, biases, and internal complexity. Trought a theoretical and reflective approach, this article examines how the fundamental risks of artificial intelligence arise from the opacity of its models, their black box architecture, technological dependence, and control by foreign powers. It also warns about the threats these technologies pose to civil liberties and human rights. The article concludes that, without proper ethical and regulatory understanding, artificial intelligence may become an instrument of domination rather than progress.

Keywords: artificial intelligence, technological risks, algorithmic bias, digital ethics, human rights, technological dependence, black box.

Los riesgos ocultos de la inteligencia artificial y los modelos de Lenguaje

Introducción.

En la última década, la inteligencia artificial (IA) ha pasado de ser un campo de investigación especializado a convertirse en una tecnología presente e influyente en la vida cotidiana de las personas y

en el funcionamiento de las empresas. Entre los desarrollos relevantes se encuentran los modelos de lenguaje (Large Language Models, LLM), que han despertado un interés sin precedentes en la sociedad y sectores como la educación, la comunicación corporativa, el empleo y los organismos internacionales. Según Liang et al. (2025), el uso de los LLM se ha extendido rápidamente en ámbitos diversos como la atención a las quejas de los consumidores, las ofertas laborales y los comunicados de prensa de la organización de las Naciones Unidas (ONU).

El uso masivo de la Inteligencia artificial y los modelos lingüísticos en diversos contextos, desde los entornos académicos y productivos hasta las entidades gubernamentales y de seguridad, plantea un conjunto de interrogantes fundamentales como: ¿qué herramienta estamos utilizando?, ¿ qué beneficios puede aportar?, ¿Qué riesgos se derivan de su uso masivo y sin regulación?, ¿Cómo fueron entrenados?, ¿qué oculta la llamada caja negra?

Diversos autores debaten sobre la influencia negativa de los modelos de lenguaje en el campo de la educación, el empleo, el fomento del plagio y la disminución de la creatividad. Sin negar la relevancia de estos cuestionamientos, considero que tales preocupaciones no representan el riesgo más profundo asociado a las llamadas “cajas negras”. En realidad, el mayor peligro radica en que no sabemos cómo han sido entrenados estos modelos, cual es la calidad de los datos utilizados durante el entrenamiento, ni como se controla la presencia de información intencionalmente sesgada o manipulada. No tener una tecnología propia es el más grave de los riesgos al que se enfrentan países en desarrollo.

Faiza Patel (2025) revisa la tesis central de Ashley Deeks sobre el fenómeno de la “doble caja negra” en los ámbitos de la seguridad nacional, describiendo como los niveles de opacidad se combinan para dificultar la rendición de cuentas. El primer nivel corresponde a la propia secretividad institucional del poder ejecutivo en asuntos de seguridad nacional, donde muchas decisiones se toman en contextos clasificados, mientras que el segundo nivel es la opacidad inherente a los sistemas de inteligencia

artificial, cuyos procesos internos resultan difíciles o imposibles de comprender incluso para sus diseñadores.

Las preocupaciones planteadas por Patel y otros autores sobre los riesgos que la inteligencia artificial representa para los derechos civiles, los derechos humanos y el control de los arsenales nucleares abren una reflexión inquietante: ¿qué ocurriría en una sociedad como la estadounidense si la IA tuviera capacidad de intervenir en decisiones críticas, como "controlar el botón rojo" o ejercer funciones de supervisión civil? Estas preguntas son legítimas dentro de su contexto, pero mi reflexión va más allá: si las IA están encapsuladas en una doble caja negra, ¿cuáles son los riesgos derivados de una tecnología que podría ser entrenada con sesgos maliciosos? ¿cómo se vería afectada la seguridad de los estados que dependen de una tecnología que no es propia? ¿Qué niveles de manipulación o coerción podrían conducir al control absoluto de una nación sobre otra? Y finalmente ¿cómo podría lograrse un acuerdo internacional que permita "abrir la caja" y establecer relaciones políticas justas y transparentes.

Por estas razones, resulta ineludible analizar los riesgos estructurales de la inteligencia artificial mas allá de su impacto inmediato en la educación o el empleo. Este artículo propone un abordaje reflexivo que examine las implicaciones éticas, sociales y políticas derivadas de la opacidad de los modelos de lenguaje y de los sesgos presentes en sus procesos de entrenamiento. Se parte de la hipótesis de que los riesgos más profundos de la IA no surgen solo de su uso cotidiano, sino de la falta de transparencia y control sobre los datos y arquitecturas que la conforman. Desde esta visión, el trabajo busca aportar a la discusión de la necesidad de una regulación internacional, de auditorías abiertas y de un desarrollo tecnológico más justo y responsable.

De esta manera, la comprensión de la opacidad y los sesgos algorítmicos se convierten en un punto de partida esencial para examinar las implicaciones éticas y geopolíticas de la inteligencia artificial.

En la siguiente sección se exponen los antecedentes teóricos más relevantes sobre la relación entre la opacidad, sesgos y poder algorítmico.

Antecedentes

En los últimos años los modelos de lenguaje, en particular, han sido objeto de debate por su opacidad y carácter de caja negra, con un potencial considerable para amplificar los sesgos sociales y dificultar la auditoría de sus procesos de entrenamiento (Liang, 2025).

Una de las preguntas que muchos investigadores se formulan hoy es cómo se entrenan estos modelos, como se construyen los conjuntos de datos y como se etiquetan las muestras que los alimentan. También se cuestionan cuáles son los criterios políticos, morales o ideológicos que orientan la clasificación de los datos y hasta qué punto estos pueden introducir sesgos, de forma intencional o no. El trabajo de investigación Crawford y Paglen (2021), desarrollado junto a Trevor Paglen, resulta especialmente revelador en este sentido. En su artículo *Excavating AI: The Politics of Image in Machine Learning Training Sets*, la autora analiza los procesos de etiquetado de imágenes utilizado para el entrenamiento de sistemas de inteligencia artificial. Crawford y Paglen (2021) examinan miles de fotografías y descubren casos inquietantes: por ejemplo, una imagen de una mujer sonriendo en bikini aparece clasificada como "holgada, zorra, mujer desaliñada, ramera", Un joven bebiendo cerveza está categorizado como "alcohólico, borracho o dipsómano".

Estos ejemplos evidencian como los conjuntos de entrenamiento pueden incorporar y reproducir prejuicios culturales y morales, reflejando una visión distorsionada del mundo que, al ser integrada en sistemas automáticos, se transforman en un riesgo social en gran escala.

Los hallazgos encontrados en los datos ponen en evidencia que los sesgos en la inteligencia artificial no son simples errores técnicos ni anomalías accidentales, sino el reflejo de estructuras culturales, ideológicas y económicas que influyen en la forma en que se recopilan, seleccionan y etiquetan

los datos. En algunos casos, estos sesgos pueden ser incluso introducidos de manera deliberada, con el propósito de reforzar estereotipos, manipular narrativas o favorecer determinados intereses políticos y comerciales. Cuando estas distorsiones se integran en modelos de lenguaje de gran escala, su impacto se multiplica, las decisiones algorítmicas comienzan a moldear percepciones, conductas y políticas públicas. Así, el riesgo deja de estar únicamente en el funcionamiento interno del modelo y se traslada al terreno de lo social y lo geopolítico, donde la manipulación informativa y el control ideológico pueden ejercerse con una eficacia sin precedentes.

Como advierte Crawford (2021) en *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, la inteligencia artificial debe entenderse no solo como una tecnología abstracta, sino como una infraestructura material y política que extrae recursos, trabajo y datos de distintas regiones del mundo. Esta genera nuevas formas de dependencia para los países consumidores del Sur global, que utilizan herramientas que no comprenden ni controlan plenamente. Zuboff (2019) denomina a este fenómeno *capitalismo de la vigilancia*, un sistema en el que la información y los comportamientos humanos se convierten en materia prima para el control social y económico.

Bostrom (2021), por su parte, advierte que la desigualdad en el desarrollo de la inteligencia artificial puede conducir a escenarios de dominancia estratégica donde los estados con mayor capacidad algorítmica impongan sus valores, intereses y marcos regulatorios sobre el resto del mundo. En consecuencia, los riesgos de la inteligencia artificial no solo amenazan la privacidad o la creatividad individual, sino también la autonomía política de las naciones. De ahí la urgencia de promover una gobernanza internacional que garantice la transparencia, la cooperación y un equilibrio tecnológico justo entre países desarrolladores y los países dependientes.

Riesgos Principales

Los riesgos asociados a la inteligencia artificial y a los modelos de lenguaje no provienen del uso mediático, ni del empleo inocente o malicioso por parte de los usuarios finales. Son originados en la propia arquitectura de estas herramientas, en su carácter de caja negra, que implica el desconocimiento de los datos con los cuales fueron entrenadas. Esto genera la imposibilidad de cuantificar los sesgos derivados de la información dispersa o deliberadamente seleccionada, así como de conocer los criterios utilizados durante el proceso de etiquetado. Además, la falta de métricas confiables para evaluar las etiquetas y la complejidad de sus arquitecturas, que incluso sus diseñadores no logran comprender plenamente, profundizan la incertidumbre. Finalmente, la ausencia de mecanismos de validación independientes y de herramientas políticas de control constituye uno de los riesgos más significativos en el desarrollo actual de la inteligencia artificial.

Riesgo 1: Opacidad y caja negra

Uno de los riesgos más profundos de la inteligencia artificial es la opacidad de sus sistemas. La mayoría de los modelos de lenguaje y algoritmos de aprendizaje funcionan como verdaderas cajas negras, en las que resulta imposible rastrear como se toman las decisiones o qué factores influyen en los resultados. Esta falta de transparencia impide auditar sus procesos, identificar errores y establecer responsabilidades cuando una decisión automatizada genera consecuencias éticas o sociales.

Patel F. (2025) advierte que esta opacidad se agrava con el fenómeno de “doble caja negra”. Por un lado, la confiabilidad institucional de gobiernos y corporaciones, y, por otro lado, la complejidad técnica de los propios sistemas de IA. Cuando ambos niveles se combinan, se diluye la rendición de cuentas y se establece un escenario en el que ni los ciudadanos ni los diseñadores pueden explicar cómo ni por que una máquina decide. Esta imposibilidad de comprensión representa un desafío ético central de la inteligencia artificial en estos tiempos.

Riesgos 2: Sesgos algorítmicos

Otro riesgo estructural inherente a la inteligencia artificial es la presencia de sesgos algorítmicos. Los sistemas de aprendizaje automático aprenden a partir de datos humanos y, por tanto, reproducen prejuicios, estereotipos y desigualdades que los datos contienen. Crawford y Paglen (2021) demostraron cómo los conjuntos de entrenamiento pueden incluir categorías ofensivas, sexistas y racistas, lo que convierte a la IA en un espejo amplificador de los sesgos sociales. Como advierte Zuboff (2019), estos errores no son simples fallas técnicas, sino el reflejo de estructuras políticas y culturales que definen que información se considera valida y como se representa el mundo.

Cuando los modelos de lenguaje se entrenan con datos sesgados, terminan consolidando visiones discriminatorias que afectan a grupos minoritarios, a mujeres, o a comunidades históricamente marginadas. La magnitud de este riesgo se multiplica en los modelos de lenguajes de gran escala que hoy llegan a grandes segmentos de la población y que son capaces de **diseminar y naturalizar prejuicios a nivel global**, bajo la apariencia de una presunta neutralidad tecnológica.

Riesgo 3: Dependencia tecnológica y control geopolítico.

Un tercer riesgo estructural de la inteligencia artificial se relaciona con la dependencia tecnológica y el control geopolítico que ejercen las potencias desarrolladoras de estas herramientas. La mayoría de los modelos de lenguaje y sistemas de IA pertenecen a corporaciones privadas radicadas en países del Norte global, lo que otorga a estas naciones un poder estratégico sobre el acceso, a la información y la infraestructura digital. Zuboff (2019) advierte que la desigualdad en el desarrollo de la inteligencia artificial puede conducir al escenario de la dominancia algorítmica donde los estados con mayor capacidad tecnológica imponen sus valores y marcos regulatorios sobre el resto del mundo. En este contexto, los países que dependen de tecnologías externas quedan vulnerables a la manipulación informativa y a la perdida de soberanía digital. Este riesgo trasciende lo técnico: compromete la

autonomía política de las naciones y refuerza la urgencia de promover una gobernanza internacional que garantice la transparencia, la cooperación y un "equilibrio" tecnológico justo.

Desafíos éticos y sociales

"*Talvez el último dinosaurio levantó la mirada hacia un cielo que parecía el mismo de siempre, sin comprender que una luz distante anunciaba el final de su era*." Ninguna criatura de aquel tiempo tenía la capacidad de entender lo que estaba ocurriendo: el cambio era demasiado grande, demasiado rápido, demasiado ajeno a su conciencia. De forma semejante la humanidad podría hallarse ante un desafío similar. Si no logramos comprender los riesgos estructurales, la opacidad y el potencial de la inteligencia artificial, como herramienta de dominio, podríamos presenciar transformaciones irreversibles sin advertirlo, como aquellos seres que no supieron que se habían extinguido. La historia no se repite, pero advierte: La ignorancia ante lo desconocido siempre ha sido el preludio del colapso.

La opacidad de los sistemas algorítmicos plantea un dilema ético esencial: si no comprendemos como decide la máquina, tampoco podemos asignar responsabilidad moral ni legal a sus resultados. Este artículo fusiona lo técnico con lo humano e invita a reflexionar sobre: la necesidad de transparencia algorítmica, la protección de los derechos humanos y el dilema de la creatividad frente a la automatización constituyen los principales desafíos éticos de esta nueva era tecnológica. Estas cuestiones invitan a reflexionar sobre la responsabilidad moral del ser humano ante los sistemas que ha creado y sobre los límites del control que aún puede ejercer sobre ellos.

Como señala Patel (Patel, 2025), el problema moral de la "doble caja negra" en gobiernos y empresas, evidencia la urgencia de establecer marcos éticos y normativos que permitan auditar los sistemas automatizados y garantizar que la tecnología permanezca al servicio de la humanidad y no al contrario.

Perspectivas y Propuestas

Los riesgos estructurales de inteligencia artificial, opacidad, sesgos y dependencia tecnológica, solo pueden afrontarse mediante una acción conjunta que combine educación crítica, transparencia y soberanía digital.

En primer lugar, es urgente impulsar una alfabetización crítica en inteligencia artificial, que no se limite a enseñar su uso, sino que fomente la comprensión ética, política y social de los algoritmos. Formar ciudadanos capaces de cuestionar la tecnología es el primer paso para evitar una dependencia ciega y absoluta de los sistemas tecnológicos como la inteligencia artificial.

En segundo lugar, se deben establecer políticas de transparencia y auditoría algorítmica, que obliguen a las instituciones y empresas a documentar los datos y modelos utilizados, abrir sus procesos de entrenamiento y permitir la evaluación independiente de sus decisiones. La trazabilidad es el fundamento de la confianza tecnológica.

Finalmente, los países del sur global deben pasar de ser consumidores a productores de inteligencia artificia propia. El primer paso es construir y controlar conjuntos de datos, así como desarrollar modelos de lenguajes basados en inteligencia artificial, desarrollar herramientas locales capaces de responder a sus realidades culturales, sociales y lingüísticas. Esta capacidad tecnológica propia es la única vía para reducir el creciente distanciamiento con las potencias desarrolladas y fortalecer la posición negociadora de los países en desarrollo frente al dominio de las grandes corporaciones y centros tecnológicos del Norte.

Conclusiones

La inteligencia artificial no es una tecnología neutral ni transparente: refleja las estructuras de poder, los valores y los sesgos de las sociedades que la crean. Sus riesgos mas complejos no se encuentran

en el uso cotidiano, sino en su diseño estructural y en la opacidad de los modelos que concentran el control del conocimiento y de los datos.

Frente a este panorama, requerimos de una regulación ética, cooperación internacional y alfabetización digital crítica, capaces de equilibrar innovación y responsabilidad. Solo a través de la educación, la transparencia y la participación ciudadana será posible mantener la tecnología al servicio de la humanidad.

Los países del Sur global enfrentan el desafío de quedar una vez más rezagados en esta nueva revolución. Debemos desarrollar nuestras propias herramientas y modelos de inteligencia artificial, fortaleciendo así la soberanía tecnológica y la capacidad de negociación frente a las potencias desarrolladas.

En última instancia, la inteligencia artificial refleja la creatividad humana, pero también sus sombras: la opacidad, el sesgo y la desigualdad. El reto no es que detengamos su avance, sino abrir la caja negra para comprender, regular y humanizar sus sistemas antes de que ellos definan los limites de nuestra propia autonomía.

Bibliografía

Bostrom, N. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Obtenido de https://yalebooks.yale.edu/book/9780300252392/atlas-of-ai/

Crawford, K. &. (2021). Excavating AI: The politics of images in machine learning training sets. AI and Society. *36*, 1105-1116. doi:https://www.lsba.org/documents/CLE/Diversity/excavatingai.pdf

Crawford, K. (. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Obtenido de https://yalebooks.yale.edu/book/9780300252392/atlas-of-ai/

Liang, W. Z. (17 de Febrero de 2025). Widespread Adoption of Large Language Model-Assisted Writing Across Society. doi:arXiv:2502.09747 / DOI:10.48550/arXiv.2502.09747

Patel, F. (18 de Julio de 2025). Peering into the ‘Double Black Box’ of National Security and AI. Lawfare. Obtenido de https:/www.lawfaremedia.org/article/peering-into-the--double-black-box--of-national-security-and-ai?utm_source=chatgpt.com

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. Obtenido de https://www.hachettebookgroup.com/titles/shoshana-zuboff/the-age-of-surveillance-capitalism/9781610395700/?lens=publicaffairs