Alexandra María
Silva Monsalve*

LOS SESGOS EN LA INTELIGENCIA ARTIFICIAL: RETOS Y OPORTUNIDADES

La ética es esencial para garantizar que el uso de la inteligencia artificial (IA) esté alineado con los valores humanos y sociales

CALEIDOSCOPIO

Introducción

La Cuarta Revolución Industrial ha intensificado las desigualdades a nivel global debido a la rápida transformación tecnológica que ha impulsado. Si bien avances como la inteligencia artificial, la automatización y la digitalización han favorecido el crecimiento económico y la innovación, también han exacerbado las disparidades preexistentes (Foro Económico Mundial, 2016). Las oportunidades generadas por estas tecnologías no se han distribuido de manera equitativa, dejando a sectores vulnerables de la sociedad rezagados. Un ejemplo claro es la automatización, que ha sustituido empleos tradicionales en industrias menos tecnológicas, impactando desproporcionadamente a las comunidades de bajos ingresos. Además, el acceso desigual a la educación y a las habilidades digitales necesarias para prosperar en este nuevo contexto ha ampliado aún más las brechas entre quienes tienen acceso a los recursos tecnológicos y quienes no.

* Docente investigadora de los programas de Licenciatura en Educación Infantil, y Maestría en Tecnología e Innovación Educativa. Ingeniera de Sistemas, Esp. en Nuevas Tecnologías de Desarrollo. Maestría en Gestión y Aplicación SW, PhD en educación. Estudios posdoctorales en Ciencias Sociales, Educación e Interculturalidad. ORCID: https://orcid.org/0000-0001-7554-0237; alexandrasilva@usta.edu.co

La inteligencia artificial (IA) es uno de los pilares de la Cuarta Revolución Industrial y ha cambiado muchas industrias a nivel mundial. La automatización, el análisis de grandes volúmenes de datos y la toma de decisiones en tiempo real son algunos de los campos en los que la IA está contribuyendo a cambiar la forma en que se fabrican bienes y servicios (Flórez *et al.*, 2019). Sin embargo, también está generando desafíos importantes, como la sustitución de empleos en sectores tradicionales y la necesidad de adquirir nuevas habilidades para adaptarse a este cambio. Además, la IA está generando nuevas industrias y oportunidades, pero su implementación desigual está creando más brechas socioeconómicas entre países y dentro de las mismas sociedades, porque no todos tienen acceso a la misma tecnología (Alonso *et al.*, 2020).

La inteligencia artificial (IA) enfrenta críticas debido a los sesgos que pueden introducirse en diversos ámbitos. Estos sesgos surgen principalmente de los datos con los que se entrenan los sistemas de IA, los cuales pueden contener prejuicios históricos, sociales o culturales (Organización Internacional del Trabajo, 2024). Así, cuando los algoritmos aprenden de estos datos, terminan replicando patrones discriminatorios y amplificando las desigualdades preexistentes. Este fenómeno se observa en sectores como la justicia penal, la contratación laboral y la atención médica, donde los sistemas automatizados pueden tomar decisiones injustas que afectan a grupos vulnerables.

Durante la Cumbre Ministerial sobre IA en Montevideo (Unesco, 2020), se discutió ampliamente sobre la necesidad de construir IA más ética e inclusiva, abordando los riesgos de perpetuar estos sesgos sistémicos. Los expertos llamaron la atención sobre la urgencia de revisar las bases de datos y los modelos utilizados para entrenar las IA, con el fin de garantizar que no se perpetúen patrones de discriminación históricos y promover una implementación justa de estas tecnologías.

Sin embargo, ¿qué entendemos por un sesgo?

Un sesgo es una inclinación o predisposición que afecta la imparcialidad de juicios, decisiones o análisis. En el contexto de la inteligencia artificial y otras áreas, los sesgos ocurren cuando los datos o los algoritmos utilizados contienen patrones o errores sistemáticos que reflejan prejuicios, ya sean conscientes o inconscientes. Estos prejuicios pueden estar relacionados con el género, la raza, la clase social, la etnia u otras características demográficas o contextuales (O'Neil, 2016). Como resultado, las decisiones basadas en sistemas con sesgo pueden reproducir o incluso amplificar discriminaciones ya presentes en la sociedad.

En términos generales, un sesgo puede ser cognitivo (cuando afecta el razonamiento humano), estadístico (cuando surge de muestras no representativas o errores de medición) o algorítmico (cuando un sistema automatizado refleja patrones desiguales presentes en los datos). Es crucial reconocer y mitigar los sesgos, especialmente en tecnologías como la IA, para evitar la perpetuación de desigualdades y asegurar la equidad en la toma de decisiones.

Es crucial reconocer y mitigar los sesgos, especialmente en tecnologías como la ia, para evitar la perpetuación de desigualdades y asegurar la equidad en la toma de decisiones.

El escrito, a manera de reflexión, se orienta en tres preguntas clave. Primera, ¿cuáles son los tipos de sesgos en la inteligencia artificial? Aquí se discutirán los sesgos algorítmicos, de género, raciales, entre otros, que surgen en los sistemas de IA. Segunda, ¿qué políticas éticas existen actualmente sobre la inteligencia artificial? Esta sección analizará normativas y marcos internacionales que buscan regular el uso responsable de la IA. Finalmente, se reflexionará sobre cómo la ética contribuye al desarrollo y aplicación de la inteligencia artificial, enfatizando la importancia de construir sistemas equitativos y transparentes que minimicen el riesgo de discriminación.

Tipos de sesgos en la IA

Los sesgos en la inteligencia artificial (IA) generan problemas que afectan la equidad en las decisiones automatizadas, particularmente en sectores críticos como la salud, la justicia y la contratación laboral. Uno de los principales problemas proviene del *sesgo en los datos*, ya que los conjuntos de datos utilizados para entrenar los sistemas de IA son incompletos o no reflejan la diversidad de la población. Esto provoca que los algoritmos favorezcan ciertos grupos sobre otros, como ocurre en el reconocimiento facial, donde las personas de color y las mujeres tienden a ser identificadas incorrectamente debido a su subrepresentación en los datos de entrenamiento.

Los sesgos en la inteligencia artificial (IA) generan problemas que afectan la equidad en las decisiones automatizadas, particularmente en sectores críticos como la salud, la justicia y la contratación laboral

Además, *el sesgo algorítmico* surge cuando los algoritmos priorizan características que refuerzan estereotipos sociales existentes, como los relacionados con el género o la etnia, creando sistemas técnicamente eficientes, pero socialmente injustos (Barocas *et al.*, 2019). Por otra parte, *el sesgo de interpretación* ocurre cuando se confía excesivamente en los resultados generados por la IA sin tener en cuenta sus limitaciones, lo que puede tener graves consecuencias en áreas como la justicia o la atención médica.

Los sesgos en la inteligencia artificial no se limitan solo a los datos y algoritmos. También existen *sesgos de género*, como se ha observado en asistentes virtuales como *Alexa o Siri*, los cuales utilizan predominantemente voces femeninas, perpetuando estereotipos de sumisión y servicio (Crawford, 2021). En el ámbito laboral, los algoritmos de contratación han mostrado una discriminación significativa, como en el caso de *Amazon*, donde un sistema penalizaba automáticamente a las mujeres al aprender de patrones históricos que favorecían la contratación de hombres en roles técnicos.

Por otra parte, los *sesgos en la IA generativa* también son preocupantes. Los sistemas de generación de imágenes, como *DALL-E*, tienden a representar estereotipos de género y a subrepresentar a ciertos grupos demográficos. Estos generadores de imágenes pueden reforzar prejuicios visuales al asociar ciertos trabajos con un solo género o excluir completamente a personas de color en sus representaciones, como se puede identificar en la siguiente figura.

Figura 1

Sesgos de IA generativa

Nota: estereotipos de empleos por género. Imagen creada con IA generativa DALL-E

Legislación y normatividad en el uso ético de la IA

Actualmente, existen diversas políticas éticas a nivel global que buscan regular el desarrollo y uso de la inteligencia artificial (IA). Estas normativas tienen como objetivo garantizar que la IA sea implementada de manera segura, transparente y justa. Entre las más destacadas se encuentran las Directrices de la Unesco sobre la ética de la inteligencia artificial, aprobadas en 2021, las cuales enfatizan la protección de los derechos humanos y la igualdad, instando a los Estados miembros a desarrollar marcos legales que promuevan la transparencia en los sistemas de IA y garanticen que su uso no perpetúe la discriminación ni las desigualdades. Este marco internacional es fundamental para guiar a los países en la creación de políticas éticas que aseguren un uso responsable de la IA a nivel global (Unesco, 2020).

Por otro lado, la Unión Europea ha adoptado un enfoque regulatorio con su "Reglamento de Inteligencia Artificial", que propone una clasificación

de riesgos para los sistemas de IA y prohíbe aquellos considerados de alto riesgo, como los que podrían manipular el comportamiento humano o utilizarse para vigilancia masiva sin supervisión. La legislación también establece normas estrictas de transparencia, exigencias de control humano y responsabilidad para los desarrolladores y usuarios de IA. Estas políticas buscan equilibrar la innovación tecnológica con la protección de los derechos fundamentales (European Commission, 2021).

En otros contextos internacionales, como en la Cumbre Mundial sobre la IA para el Bien Común, se ha discutido la necesidad de políticas globales que promuevan el uso responsable de la IA en beneficio de la humanidad. En estas cumbres, se han destacado iniciativas como la creación de comités de ética en IA y el desarrollo de guías que instan a los países a implementar marcos normativos que aseguren la justicia y la equidad en la implementación de estas tecnologías. Estos esfuerzos internacionales se enfocan en evitar que la IA perpetúe sesgos, discrimine a poblaciones vulnerables o intensifique las desigualdades económicas (AI for Good, 2020).

En el marco de las políticas éticas sobre inteligencia artificial en América Latina, la reciente Cumbre de Inteligencia Artificial de Montevideo ha sido un evento clave para abordar los desafíos y oportunidades de la IA en la región. En esta cumbre, se discutieron los avances en la implementación de IA en sectores como la salud, la educación y la gobernanza, además de enfatizar la necesidad de adoptar marcos éticos que respondan a las realidades socioeconómicas de América Latina. Los líderes y expertos presentes hicieron un llamado a los gobiernos para que se comprometan a crear normativas nacionales que aseguren la transparencia y la rendición de cuentas en el uso de la IA, especialmente en países en desarrollo que enfrentan riesgos de ampliar las desigualdades sociales si no se regulan adecuadamente estas tecnologías. Asimismo, se resaltó la urgencia de implementar políticas inclusivas que tengan en cuenta la diversidad cultural y las brechas digitales que aún persisten en la región (Unesco, 2024).

Además, la cumbre destacó la creación de comités de ética regionales, encargados de supervisar el desarrollo e implementación de la IA para asegurar que su uso respete los derechos humanos y promueva el bienestar social. La discusión incluyó cómo abordar los sesgos inherentes en los algoritmos y garantizar que las comunidades más vulnerables no sean excluidas de los beneficios que la IA puede ofrecer. Los participantes subrayaron la importancia de la cooperación regional para establecer estándares éticos compartidos y para impulsar la capacitación tecnológica que permita a la región competir a nivel global sin comprometer los principios éticos fundamentales (Unesco, 2024).

Contribución de la ética en el uso y desarrollo de la IA

La ética es esencial para garantizar que el uso de la inteligencia artificial (IA) esté alineado con los valores humanos y sociales, y no solo con los avances tecnológicos. Adela Cortina (2021) destaca la importancia de la ética aplicada en la toma de decisiones tecnológicas, afirmando que las innovaciones, como la IA, deben estar al servicio del bien común y no convertirse en herramientas que perpetúen desigualdades. En este sentido, la ética no solo previene abusos en la toma de decisiones automatizadas, sino que también fomenta la creación de sistemas justos y equitativos que respeten los derechos fundamentales. Incorporar principios éticos en el desarrollo de la IA ayuda a mitigar los riesgos relacionados con los sesgos algorítmicos y protege a las comunidades más vulnerables, lo que permite que la IA trabaje de manera inclusiva y justa (Floridi y Cowls, 2019).

Además, la ética tiene un rol crucial en el diseño de políticas regulatorias que orienten el uso responsable de la IA, protegiendo tanto la privacidad como la dignidad humana. Estas políticas deben abordar los posibles impactos negativos de la IA, como la automatización desmedida o la discriminación, y enfocarse en garantizar una distribución equitativa de los beneficios de estas tecnologías. Cortina subraya que las decisiones tecnológicas deben promover una sociedad más justa, sin dejar de lado los principios éticos que sustentan la convivencia social (Cortina, 2021). En este contexto, la ética se convierte en una guía que no solo evita los peligros de la IA, sino que también maximiza sus beneficios para todos.

A modo de reflexión

Como bien señaló Stephen Hawking, "El desarrollo de la inteligencia artificial podría significar el fin de la raza humana si no se controla adecuadamente" (Hawking, 2018). Esta advertencia nos invita a reflexionar profundamente sobre el papel que desempeña la ética en el desarrollo de la inteligencia artificial (IA). A medida que la IA avanza y se integra en todos los aspectos de nuestras vidas, desde la medicina hasta la toma de decisiones judiciales, es crucial que mantengamos una vigilancia constante sobre sus aplicaciones. La advertencia de Hawking no solo es una llamada de atención sobre los riesgos existenciales de la IA, sino también sobre su impacto a nivel social, ético y humano.

Si bien la IA tiene el potencial de transformar nuestra sociedad para bien, sin una adecuada regulación y un marco ético sólido, sus beneficios pueden convertirse en amenazas. Al gestionar estos avances de manera responsable, con principios de equidad y justicia, podemos aprovechar su enorme potencial sin comprometer nuestra supervivencia ni los valores que nos definen como especie. Como sociedad, debemos priorizar

el bienestar común y asegurarnos de que la inteligencia artificial esté guiada por el humanismo y la responsabilidad moral, tal como la llamada *Sociedad 5.0* nos instó a considerar.

REFERENCIAS

AI FOR GOOD. (2020). *AI for Good Global Summit 2020 Report*. https://aiforgood.itu.int/

ALONSO, C., KOTHARI, S., & REHMAN, S. (2020). *Cómo la inteligencia artificial podría ampliar la brecha entre países ricos y pobres*. Fondo Monetario Internacional. Recuperado de https://bit.ly/3YpShje

BAROCAS, S., HARDT, M., & NARAYANAN, A. (2019). *Fairness and machine learning: Limitations and opportunities*. https://fairmlbook.org/

CORTINA, A. (2021). *Ética cosmopolita: Una apuesta por la cordura en tiempos de pandemia*. Paidós.

CRAWFORD, K. (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press.

EUROPEAN COMMISSION. (2021). *Proposal for a regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

FLORIDI, L., & COWLS, J. (2019). *A unified framework of five principles for AI in society. Harvard Data Science Review*. https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/8

FLÓREZ, J., AGUILERA, M., & SALCEDO, O. (2019). *Industria 4.0: Tendencias de la literatura académica reciente. Espacios, 40*(30), 2-17. http://es.revistaespacios.com/a19v40n30/a19v40n30p27.pdf

FORO ECONÓMICO MUNDIAL. (2016). *La Cuarta Revolución Industrial* [Video]. YouTube. https://www.youtube.com/watch?v=-OiaE6l8ysg

HAWKING, S. (2018). *Brief answers to the big questions*. Bantam Books. https://ia903405.us.archive.org/5/items/brief-answers-to-the-big-questions/Brief%20Answers%20to%20the%20Big%20Questions.pdf

O'NEIL, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Grupo Editorial Crown. https://dl.acm.org/doi/10.5555/3002861

ORGANIZACIÓN INTERNACIONAL DEL TRABAJO. (2024). *Cuidado con la brecha: Cerrar la brecha de la IA garantizará un futuro equitativo para todos*. https://bit.ly/3YpLnKO

UNESCO. (2020). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000373434

UNESCO. (2024). *Cumbre ministerial sobre IA en Montevideo destaca avances regionales y presenta el Informe RAM de Uruguay*. https://www.unesco.org/es/articles/cumbre-ministerial-sobre-ia-en-montevideo-destaca-avances-regionales-y-presenta-informe-ram-de