

MITIGACIÓN DEL SESGO ALGORÍTMICO EN LA IA: PROPUESTAS DESDE LA BIOÉTICA

PABLO DE ROBINA DUHART - DOCTORADO EN BIOÉTICA

JUSTIFICACIÓN

El sesgo algorítmico puede agravar las desigualdades existentes en la sociedad, afectando negativamente la equidad, socavando la confianza en la tecnología, y pudiendo tener impactos severos como tratos injustos y discriminatorios en sistemas de justicia penal o en la calidad de cuidados médicos ofrecidos a diversas poblaciones. Abordar estos desafíos es fundamental para asegurar un futuro en el que la IA beneficie a toda la humanidad de manera equitativa.



INTRODUCCIÓN

En la actualidad, la Inteligencia Artificial (IA) ha emergido como una herramienta omnipresente que impacta significativamente en varios aspectos de nuestra vida cotidiana, desde decisiones empresariales hasta diagnósticos médicos, prometiendo una mejora en la calidad de vida y eficacia de los sistemas adoptados (Diakopoulos, 2016). Sin embargo, este avance tecnológico integra un fenómeno preocupante conocido como sesgo algorítmico, el cual puede perpetuar injusticias o discriminaciones preexistentes a través de los algoritmos de IA, planteando importantes desafíos éticos y sociales (Floridi, 2019; Terrones, 2023).



METODOLOGÍA

Se adoptó un enfoque multidisciplinario, bioético, para el análisis del sesgo algorítmico, involucrando una revisión de literatura y una evaluación crítica de ejemplos específicos en los que el sesgo se manifiesta.

Se hizo una comparación de la literatura, de 2015 a 2024, y el uso de la IA, identificando así los sesgos algorítmicos y la búsqueda de cómo eliminarlos desde los principios de la bioética.

OBJETIVOS

1. Describir el fenómeno del sesgo algorítmico y su impacto en la sociedad.
2. Evaluar las implicaciones éticas y filosóficas del sesgo algorítmico en la aplicación de IA.
3. Proponer medidas preventivas y soluciones para mitigar los riesgos asociados al sesgo algorítmico, promoviendo un desarrollo y uso responsable de la IA.



RESULTADOS

A partir de la revisión literaria y el uso de IA para demostrar los sesgos se encontró lo siguiente:

1. Se reconoce como sesgo algorítmico a la presencia de errores en el diseño o la programación de sistemas de IA que sistemáticamente generan resultados favorables o desfavorables hacia algún grupo o individuos en particular.
2. Se encontró que existen múltiples sesgos que van desde la programación, hasta los resultados que presenta la IA, pasando por el entrenamiento de la IA.
3. Los tipos de sesgo encontrados fueron: selección, confirmación, exclusión, inclusión y representación.
4. Los principales sesgos se representan según: género, raza, condición socio-económica y cultural.
5. Finalmente, los sesgos muestran falta de equidad y justicia en la aplicación de procedimientos y en el seguimiento de información, lo que genera desigualdades.

CONCLUSIONES

Derivado de lo anterior, debemos considerar lo siguiente:

1. Es crucial priorizar la **transparencia**, promover la diversidad disciplinar y moral en los equipos de desarrollo (que haya ingenieros y filósofos de diversas corrientes) y realizar auditorías éticas y regulaciones rigurosas (Diakopoulos, 2016), para así evitar los sesgos de origen.
2. Es necesario reconocer y confrontar los **prejuicios** (sociales y de programación) inherentes en los procesos de diseño, programación y entrenamiento de la IA, para así hacer una programación más equitativa y ética (Barocas, S., Hardt, M., & Narayanan, A., 2019).



REFERENCIAS BIBLIOGRÁFICAS

1. Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and Machine Learning: Limitations and Opportunities*. Online: fairmlbook.org.
2. Diakopoulos, N. (2016). Accountability in Algorithmic Decision Making. *Communications of the ACM*, 59(2), 56-62. DOI:10.1145/2844110.
3. Floridi, L. (2019). *The Logic of Information: A Theory of Philosophy as Conceptual Design*. Oxford: Oxford University Press.
4. Terrones, A. L. (2023). Inteligencia Artificial fiable y vulnerabilidad: una mirada ética sobre los sesgos algorítmicos. En R. Suárez-Álvarez, M. Á. Martín-Cárdaba, & L. M. Fernández-Martínez, *Vulnerabilidad digital: desafíos y amenazas de la Sociedad hiperconectada*. (pp. 263-273). Madrid: Dykinson.