



Sesgo algorítmico en la IA: conceptos clave, implicaciones y soluciones

Este OER examina la IA y el sesgo algorítmico, sus causas y sus repercusiones en el mundo real. Destaca la importancia de la supervisión humana y las estrategias para mitigar el sesgo y conseguir sistemas de IA más justos.

La inteligencia artificial (IA): se refiere a los sistemas que muestran un comportamiento inteligente analizando su entorno y emprendiendo acciones -con cierto grado de autonomía- para alcanzar objetivos específicos.

El sesgo algorítmico: es un error sistemático o un resultado injusto producido por algoritmos de aprendizaje automático, que a menudo refleja o refuerza sesgos sociales, raciales o de género existentes, causados por suposiciones prejuiciosas en su diseño o por el entrenamiento con datos incompletos o distorsionados.

Causas:



Datos de entrenamiento defectuosos

La IA predice en función de los datos recogidos; si no hay datos, puede alucinar

La IA no sabe, recopila y pronostica

Prejuicios de los desarrolladores

La IA es un programa que sigue órdenes (su código y luego las indicaciones)



Los desarrolladores son personas con prejuicios humanos

Prejuicios en acción: Ejemplos e impactos

Sanidad: Los clasificadores de lesiones cutáneas de la CNN, entrenados principalmente en pacientes blancos, muestran una precisión -50% menor en pacientes negros.

Ámbito jurídico: Las herramientas de riesgo como COMPAS suelen etiquetar a determinadas etnias como de alto riesgo.

Recursos humanos: Los conjuntos de datos sesgados amplifican las diferencias salariales o excluyen a candidatos. Ejemplo: El algoritmo de Amazon de 2018 rebajó los currículos con «de mujer».

Vías de mitigación

- Datos diversos y representativos.
- Auditorías periódicas de exactitud e imparcialidad.
- Directrices reglamentarias y éticas.
- Transparencia y explicabilidad de las decisiones sobre IA.
- Uso de la IA para identificar posibles sesgos y disparidades.

Cosas a tener en cuenta:

La IA no percibe el lenguaje gramaticalmente, sino que predice estadísticamente

Puede confundir temas similares repitiendo patrones o mezclando conceptos relacionados

Piense en ello como una predicción de texto altamente cualificada con una ilusión de creación de conocimiento y aportación de significado.

TÚ creas conocimiento y das sentido cuando TÚ lees críticamente

La IA no sigue la lógica humana, por lo que sus resultados requieren una revisión humana

Audiencias destinatarias: 📖 Personal de bibliotecas, 👨‍🏫 Profesorado, 🎓 Estudiantado

Referencias

Bandara, R. J., K. Biswas, S. Akter, S. Shafique, and M. Rahman. "Addressing Algorithmic Bias in AI-Driven HRM Systems: Implications for Strategic HRM Effectiveness." Human Resource Management Journal, May 2025. <https://doi.org/10.1111/1748-8583.12609>

Jonker, Alexandra, and Julie Rogers. "What Is Algorithmic Bias?" IBM Think, September 20, 2024. <https://www.ibm.com/think/topics/algorithmic-bias>

Kalantzis, Mary, and Bill Cope. "Literacy in the Time of Artificial Intelligence." Reading Research Quarterly 60, no. 1 (November 23, 2024). <https://doi.org/10.1002/rrq.591>

Norori, Natalia, Qiyang Hu, Florence Marcelle Aellen, Francesca Dalia Faraci, and Athina Tzovara. "Addressing Bias in Big Data and AI for Health Care: A Call for Open Science." Patterns 2, no. 10 (2021): 100347. <https://doi.org/10.1016/j.patter.2021.100347>

Sheikh, Haroon, Corien Prins, and Erik Schrijvers. "Artificial Intelligence: Definition and Background." Research for Policy, January 2023, 15–41. https://doi.org/10.1007/978-3-031-21448-6_2

Contexto del proyecto GEDIS

Este recurso forma parte del proyecto europeo GEDIS (Gender Diversity in Information Science), que promueve herramientas educativas abiertas para abordar las desigualdades de género en la educación superior, con énfasis en las disciplinas relacionadas con la Documentación y la Biblioteconomía. Este material educativo se desarrolló en el marco de la Summer School de Barcelona, como parte del proyecto GEDIS.

GEDIS - Gender Diversity in Information Science: Challenges in Higher Education. <https://ub.edu/gedis>

Citación: Karmil, Kenza, Luka Boskovic and Denisa Kartašová. 2025. *Algorithmic Bias in AI: Ke Concepts, Implications, and Solutions*. DOI: [10.5281/zenodo.17067091](https://doi.org/10.5281/zenodo.17067091). Traducido por Claudia San José y Juan-José Boté-Vericad. *Sesgo algorítmico en la IA: conceptos clave, implicaciones y soluciones*. DOI: [10.5281/zenodo.17296079](https://doi.org/10.5281/zenodo.17296079).

Cofinanciado por la Unión Europea. Las opiniones y puntos de vista expresados son exclusivamente de los autores y no reflejan necesariamente los de la Unión Europea ni los del Servicio Español para la Internacionalización de la Educación (SEPIE). Ni la UE ni la autoridad que concede la subvención se hacen responsables.



Cofinanciado por la Unión Europea

