



Comprender el sesgo algorítmico

Identificar y mitigar los sesgos en los datos y algoritmos de la IA para promover la imparcialidad y la equidad



Objetivos del aprendizaje En pocas palabras



- Comprender qué es el sesgo algorítmico y cómo se origina.
- Identificar los diferentes tipos de sesgo en los datos y los sistemas de IA.
- Reconocer los riesgos asociados al uso de la IA en contextos educativos.
- Analizar el impacto de sesgo en la equidad y la inclusión.
- Aplicar estrategias básicas para mitigar el sesgo en su práctica docente.



Conceptos clave y definiciones

CONCEPTOS FUNDAMENTALES PARA IDENTIFICAR Y ABORDAR LOS SESGOS EN LOS SISTEMAS DE IA

Sesgo algorítmico

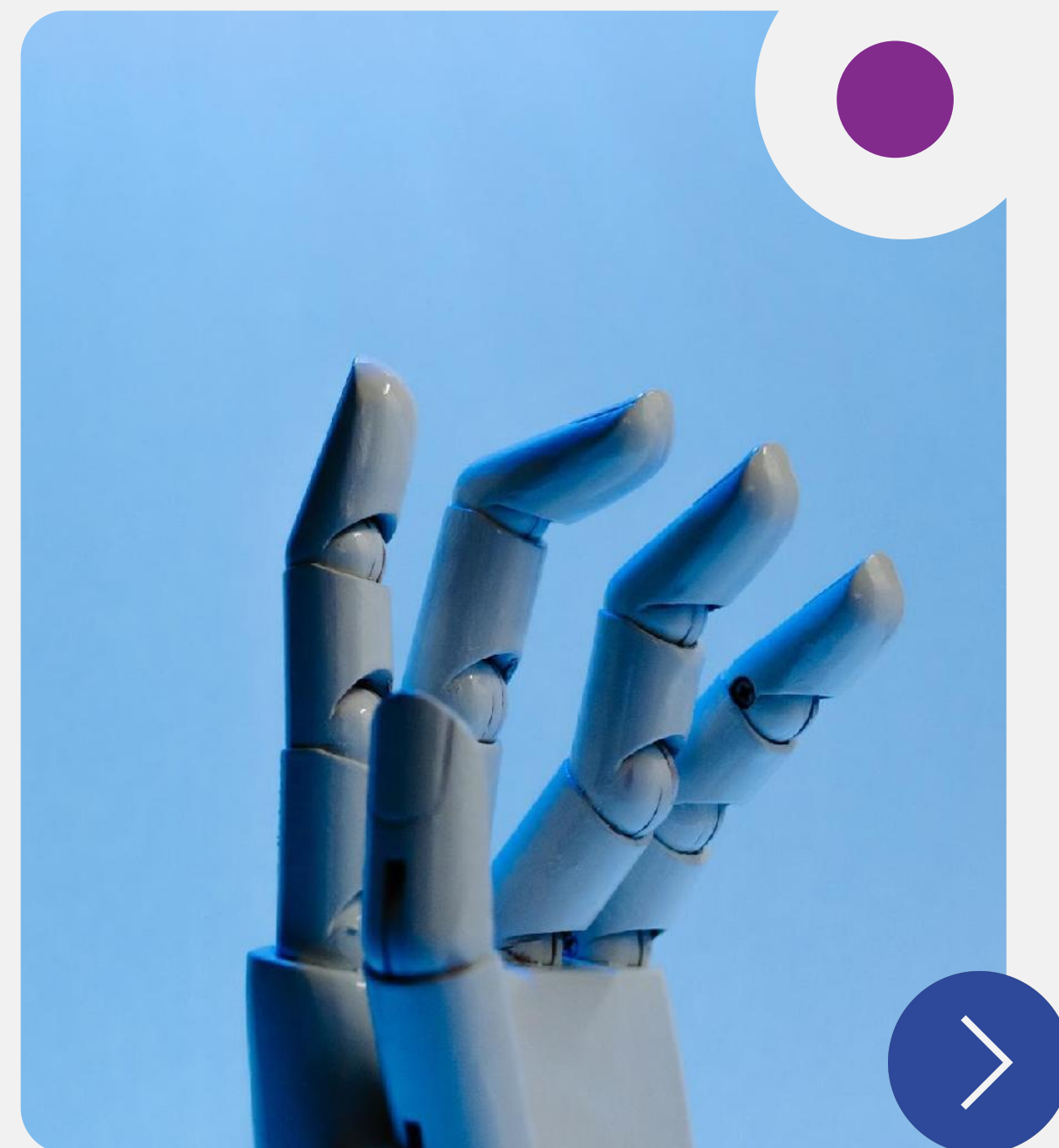
El sesgo algorítmico se produce cuando un sistema de inteligencia artificial genera resultados sistemáticamente injustos debido a datos sesgados, decisiones de diseño o supuestos implícitos

Sesgo de representación

El sesgo de representación se produce cuando determinados grupos están infrarrepresentados o sobrerrepresentados en los datos de entrenamiento.

Sesgo histórico

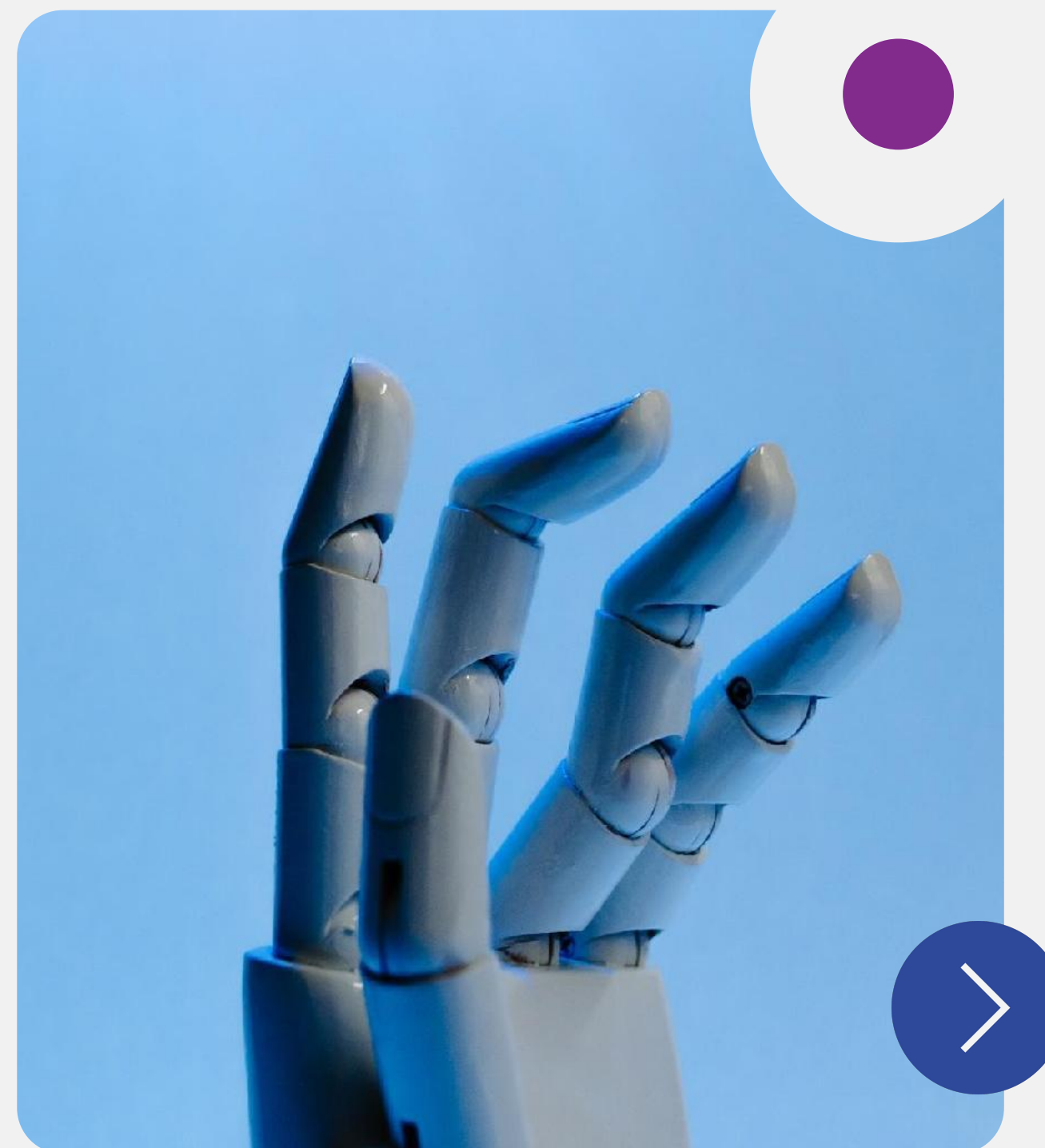
El sesgo histórico se produce cuando los conjuntos de datos reflejan desigualdades del pasado que los sistemas de IA reproducen o amplifican



¿Qué es el sesgo algorítmico?

Cómo surgen las distorsiones en los datos y los algoritmos de la IA.

- El sesgo algorítmico se produce cuando un sistema de IA genera resultados sistemáticamente injustos o desiguales debido a los datos o a la forma en que interpreta la información.
- No siempre es intencionado, pero puede reproducir o amplificar las desigualdades existentes.
- El sesgo algorítmico no es solo una cuestión técnica, sino también un reflejo de las desigualdades sociales más amplias arraigadas en los datos y las prácticas institucionales.



Principales tipos de sesgo



Sesgo histórico

Cuando los datos de entrenamiento reflejan desigualdades del pasado que quedan integradas en el sistema de inteligencia artificial.

Sesgo de medición

Cuando los datos utilizados no reflejan con precisión lo que pretenden evaluar.

Sesgo de agregación

Cuando se aplica un único modelo a grupos diversos sin tener en cuenta las diferencias.

Sesgo de diseño o implementación

Introducidos por decisiones humanas a la hora de definir variables, categorías o criterios de evaluación.

Sesgo de representación

Cuando ciertos grupos están infrarrepresentados o sobrerrepresentados en los datos de entrenamiento.

Sesgo de confirmación

Cuando los desarrolladores o diseñadores eligen datos o características que refuerzan las suposiciones existentes...



Riesgos relacionados con los datos derivados del sesgo algorítmico

Cómo los datos pueden dar lugar a resultados injustos en los sistemas de IA aplicados a la educación.



El sesgo algorítmico suele tener su origen en los propios datos.

Cuando los conjuntos de datos son incompletos, desequilibrados o están mal estructurados, los sistemas de IA pueden generar resultados distorsionados o injustos.

Riesgos clave relacionados con los datos:

- **Infra-representación de determinados grupos de estudiantes** → es posible que las experiencias y necesidades de algunos alumnos no se reflejen adecuadamente.
- **Conjuntos de datos desequilibrados** → el sistema puede funcionar mejor para los grupos mayoritarios que para las minorías.
- **Uso de variables proxy (indicadores indirectos)** → Variables como el Código postal o el tipo de centro educativo pueden reflejar involuntariamente diferencias socioeconómicas o demográficas.
- **Datos de baja Calidad o incoherentes** → la información inexacta u obsoleta puede distorsionar las predicciones y recomendaciones.
- **Dependencia excesiva de los datos cuantitativos** → las complejas realidades educativas pueden reducirse a métricas simplificadas

Si los datos están sesgados, incluso un algoritmo técnicamente sólido puede arrojar resultados sesgados.

El impacto del sesgo algorítmico en la equidad educativa



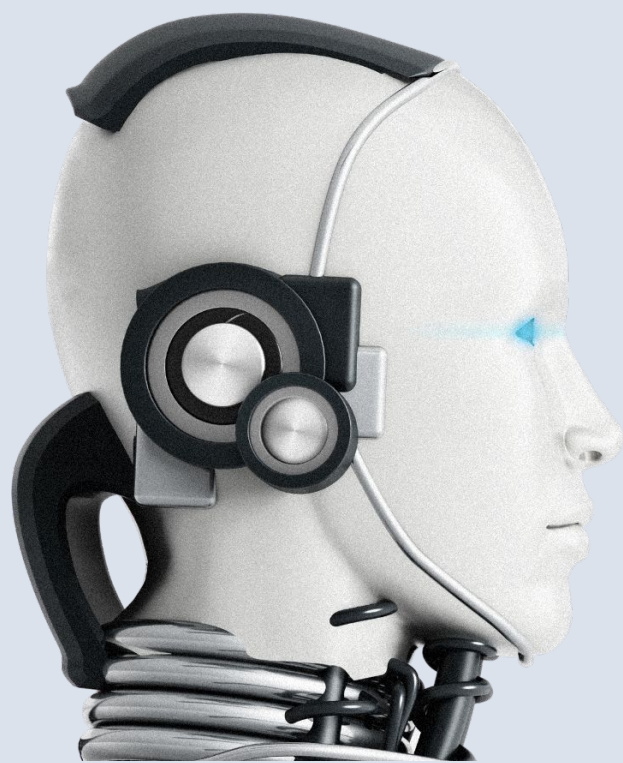
Cómo pueden influir los sistemas de IA en la igualdad de oportunidades en la educación superior

Cuando los sistemas de IA incorporan sesgos, pueden afectar de forma desigual a distintos grupos de estudiantes, incluso sin una intención explícita. Las posibles repercusiones en el ámbito educativo son:

- **Resultados de evaluación injustos o desproporcionados.** Ciertos perfiles de estudiantes pueden recibir sistemáticamente calificaciones o recomendaciones menos favorables.
- **Refuerzo de las desigualdades preexistentes.** Los patrones históricos pueden reproducirse en los análisis predictivos o en las recomendaciones sobre itinerarios académicos.
- **Restricción de las oportunidades académicas.** Los sistemas de orientación o seguimiento pueden influir en las decisiones futuras y la progression de los estudiantes.
- **Pérdida de confianza en la tecnología educativa.** La percepción de injusticia puede minar la confianza en las herramientas digitales y las prácticas insitucionales.
- **Repercusión en la inclusión y la diversidad.** Ciertos grupos pueden verse sistemáticamente en desventaja

Un sistema predictivo diseñado para identificar a los «estudiantes en situación de riesgo» puede señalar de manera desproporcionada a los estudiantes procedentes de entornos desfavorecidos si los datos históricos sobre el abandono escolar reflejan las desigualdades estructurales existentes.



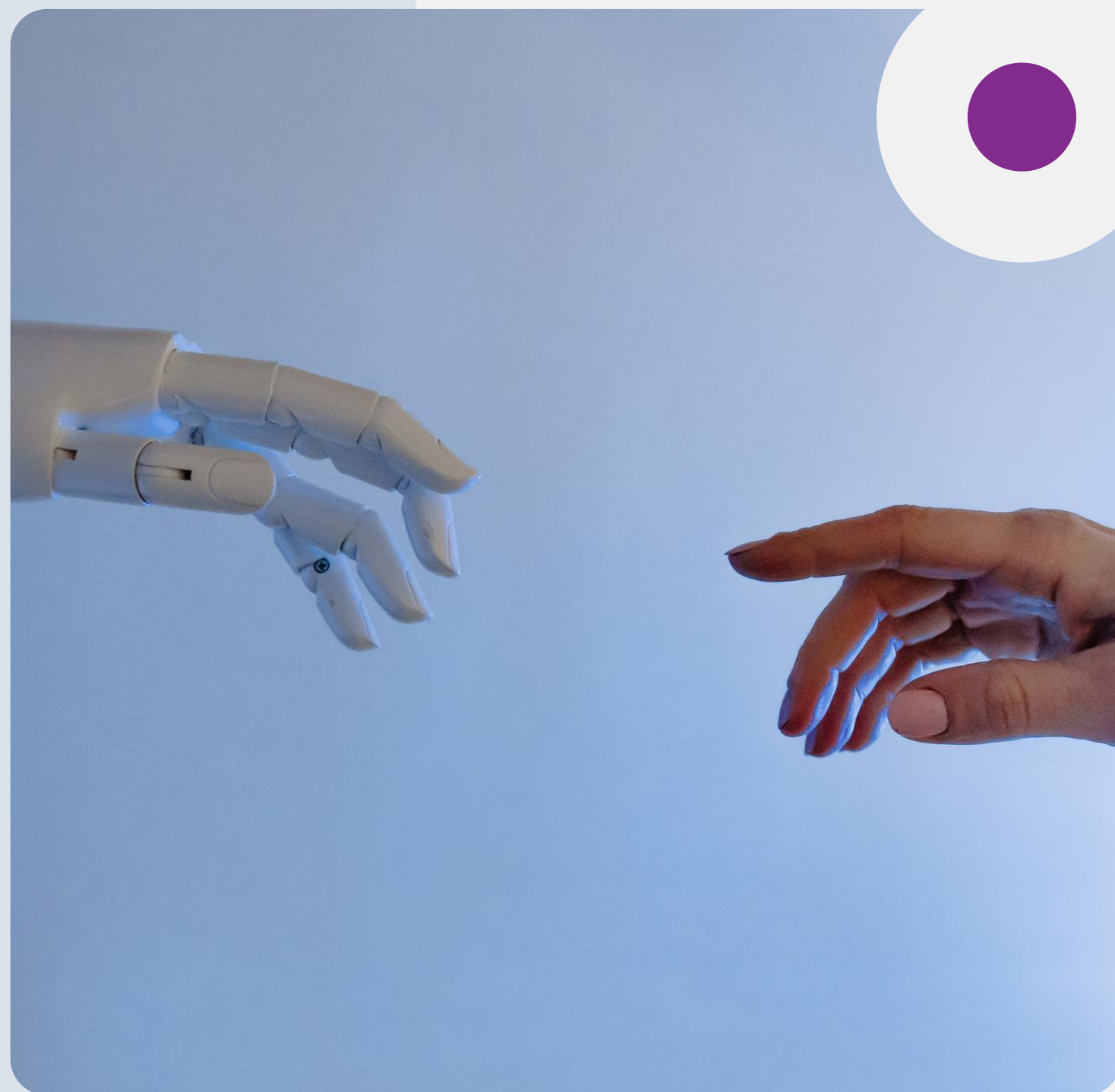


Medidas prácticas para mitigar el sesgo algorítmico y promover la equidad en la IA aplicada a la educación

Aunque no siempre es posible eliminar por completo los sesgos, estos pueden identificarse y reducirse mediante decisiones conscientes en la práctica docente.

- Cuestionar si los datos utilizados representan adecuadamente todos los perfiles de los estudiantes .
- Mantener una supervisión humana activa y revisar las recomendaciones generadas por la IA antes de tomar decisiones académicas.
- Promover la transparencia explicando a los estudiantes cómo se utilizan las herramientas de IA y cuáles son sus limitaciones..
- Observar si determinados grupos se ven afectados de forma sistemática.
- Revisar la representatividad de los datos valorando si el sistema se ha entrenado con conjuntos de datos diversos o si algunos grupos pueden estar infrarrepresentados.
- Comparar si el Sistema produce resultados sistemáticamente diferentes en función del género, el origen, el idioma u otros factores relevantes..
- Escuchar las perspectivas de los estudiantes para identificar si algún grupo percibe un trato desigual.





¿Puedes identificar el sesgo?

Aplica lo que has aprendido

El sesgo algorítmico no siempre es evidente. A menudo se manifiesta de forma sutil en las herramientas educativas y los sistemas de toma de decisiones. En esta actividad, analizarás situaciones educativas reales e identificarás el tipo de sesgo presente.

Pon en práctica tus conocimientos y pon a prueba tu sentido crítico

Comienza la actividad



Referencias

- Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities*.
<https://fairmlbook.org>
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1–35.
<https://doi.org/10.1145/3457607>
- UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.
<https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- European Commission. (2022). *Ethical guidelines on the use of artificial intelligence and data in teaching and learning for educators*.
<https://education.ec.europa.eu/news/ethical-guidelines-on-the-use-of-artificial-intelligence-and-data-in-teaching-and-learning-for-educators>



Fomentar la equidad en la educación no solo requiere buenas intenciones, sino también una conciencia crítica de cómo la tecnología influye en las oportunidades.

El Consorcio:



<https://trustai.unibit.bg/>

 Trust AI project

Descargo de responsabilidad: Financiado por la Unión Europea. No obstante, las opiniones y puntos de vista expresados son exclusivamente los de los autores y no reflejan necesariamente los de la Unión Europea ni los de la Agencia Nacional (AN). Ni la Unión Europea ni la AN se hacen responsables de los mismos.

Nº de ref. del proyecto: 2024-1-BG01-KA220-HED-000249185



Co-funded by
the European Union

