

Ethical AI

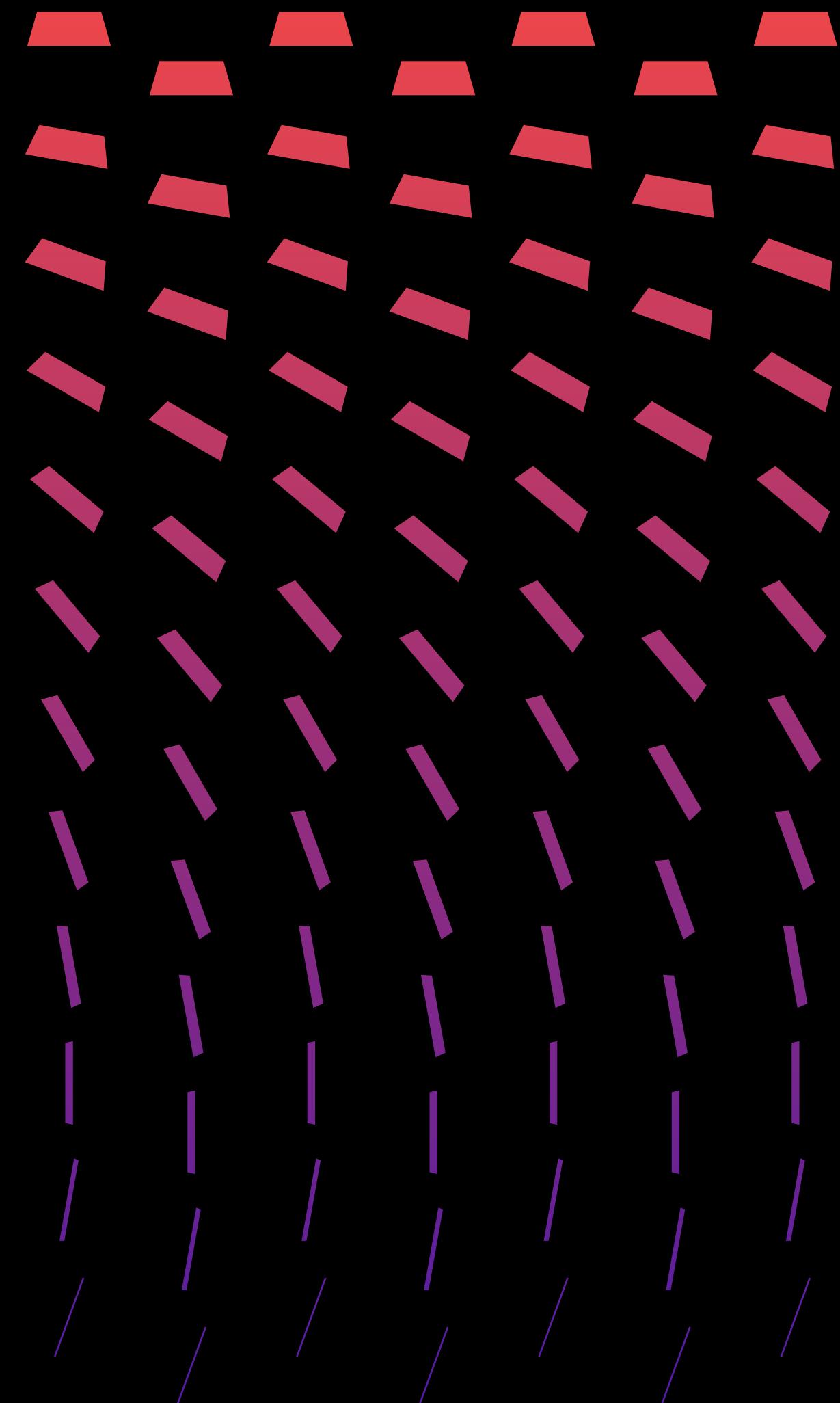
Equipo 5

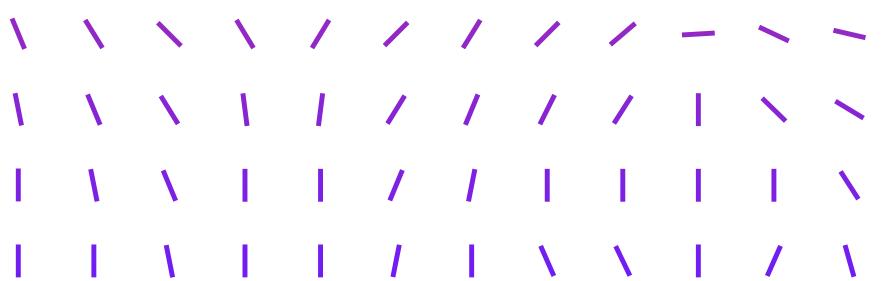
Daniela Jiménez Téllez

Lautaro Gabriel Coteja

Andrés Villarreal González

Héctor Hibran Tapia Fernández





Inteligencia Artificial

- ↳ La inteligencia artificial, o IA, es la tecnología que permite que las computadoras **simulen la inteligencia humana** y las capacidades humanas de **resolución de problemas**.

Propósitos

- ↳ La inteligencia artificial debe servir como una **herramienta para potenciar nuestras capacidades**, en lugar de sustituir el ingenio y creatividad humanos.
- ↳ Los datos y los conocimientos derivados de esos **datos pertenecen únicamente al creador**.
- ↳ El desarrollo y uso de la inteligencia artificial debe alinearse con **principios éticos que prioricen el bienestar** y los valores **humanos** en todas sus etapas.



¿Qué es la ética de la IA?

La ética es un conjunto de principios morales que nos ayudan a **discernir entre el bien y el mal**.

La ética de la IA es un campo multidisciplinario que estudia cómo optimizar el **impacto beneficioso** de la IA al tiempo que se **reducen los riesgos** y los resultados adversos

Principios Fundamentales

Respeto a las personas:

Reconoce y protege la autonomía individual, especialmente de quienes tienen capacidad disminuida. Enfatiza el consentimiento informado, permitiendo que las personas conozcan los riesgos y beneficios y elijan participar o retirarse en cualquier momento.



Beneficencia:

Basado en el principio de "no hacer daño", busca evitar que la IA cause perjuicios. Reconoce que los algoritmos pueden amplificar sesgos, por lo que es esencial diseñarlos para promover el bienestar y minimizar riesgos.



Justicia:

Se enfoca en la equidad y la igualdad en la distribución de beneficios y cargas de la IA. Propone cinco formas de distribuirlos: igualdad, necesidad individual, esfuerzo individual, contribución social y mérito, para asegurar una distribución justa.



Principales Preocupaciones

Modelos fundacionales e IA generativa

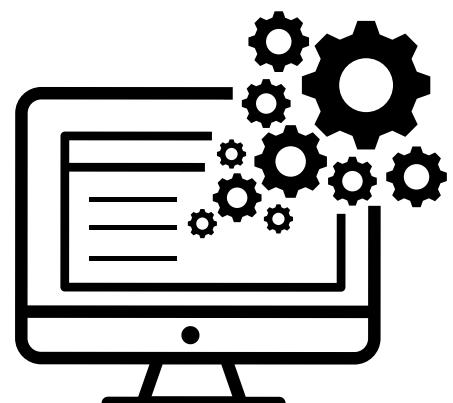
La ética de la IA busca maximizar beneficios y minimizar riesgos en inteligencia artificial, abordando temas como privacidad, imparcialidad y transparencia. Debido a diseños deficientes y datos sesgados, han surgido consecuencias imprevistas. Principios como respeto, beneficencia y justicia guían el desarrollo ético de algoritmos. La IA generativa, como ChatGPT, plantea nuevas preocupaciones éticas sobre sesgos y uso indebido.

Singularidad tecnológica

La singularidad tecnológica, donde la IA supera a la inteligencia humana, no es una preocupación cercana para los investigadores. Sin embargo, el uso de sistemas autónomos, como autos sin conductor, plantea debates éticos sobre la responsabilidad en caso de accidentes y si es mejor optar por tecnología semiautónoma para mayor seguridad.

Impacto de la IA en los empleos

Como con cualquier tecnología disruptiva, la IA cambiará la demanda de ciertos roles laborales en lugar de eliminar empleos por completo. Similar a la transición de la industria automotriz hacia los vehículos eléctricos, la IA creará nuevas oportunidades laborales, como gestionar sistemas y abordar problemas complejos en áreas como el servicio al cliente. Lo crucial será ayudar a las personas a adaptarse a estas nuevas demandas del mercado laboral.



Principales Preocupaciones

Privacidad

La privacidad se ha centrado en la protección de datos, impulsando políticas como el GDPR en Europa (2016) y la Ley de Privacidad del Consumidor de California (CCPA) en EE. UU., que otorgan más control a los usuarios sobre sus datos. Estas leyes han obligado a las empresas a replantear cómo gestionan los datos personales, priorizando la inversión en seguridad para prevenir vulnerabilidades, vigilancia y ciberataques.

Prejuicios y discriminación

El sesgo en los datos de entrenamiento puede afectar la imparcialidad de la IA, como ocurrió con Amazon, que tuvo que abandonar un proyecto de contratación automatizada debido a sesgos de género. Este problema también se extiende a otras áreas, como el reconocimiento facial y los algoritmos de redes sociales. Empresas como IBM, conscientes de estos riesgos, están tomando medidas, como eliminar sus productos de reconocimiento facial para evitar su uso en prácticas que violen derechos humanos o libertades fundamentales.

Responsabilidad

Aún no hay una legislación universal que regule la inteligencia artificial (IA), pero muchos países y estados están trabajando en desarrollar normas a nivel local. Aunque existen algunas regulaciones y otras están en camino, los marcos éticos han surgido para orientar la creación y uso de modelos de IA. Sin embargo, estos marcos solo ofrecen guías y no garantizan la prevención de daños, ya que la responsabilidad distribuida y la falta de previsión sobre las posibles consecuencias dificultan evitar impactos negativos en la sociedad.

Sesgos en la IA

El sesgo en la IA ocurre cuando un sistema de inteligencia artificial refleja o amplifica prejuicios existentes en los datos en los que se entrenó. Esto puede llevar a resultados injustos o discriminatorios.

Sesgo de datos

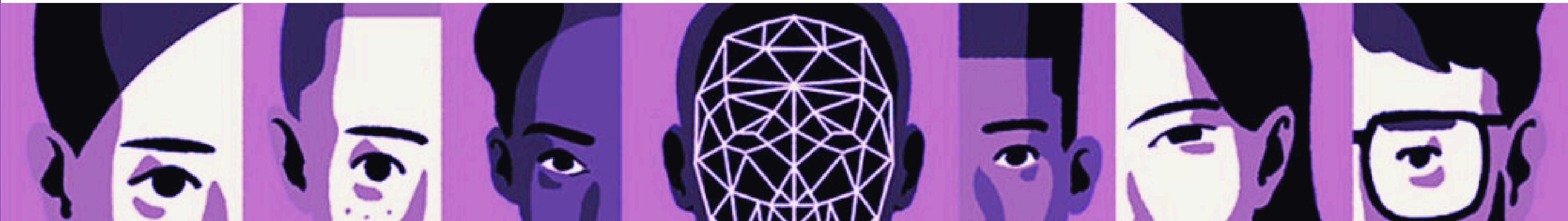
Si los datos utilizados para entrenar el modelo no son representativos, pueden introducir sesgos. Ejemplo: un conjunto de datos mayormente masculino puede sesgar decisiones hacia ese grupo.

Sesgo de algoritmo

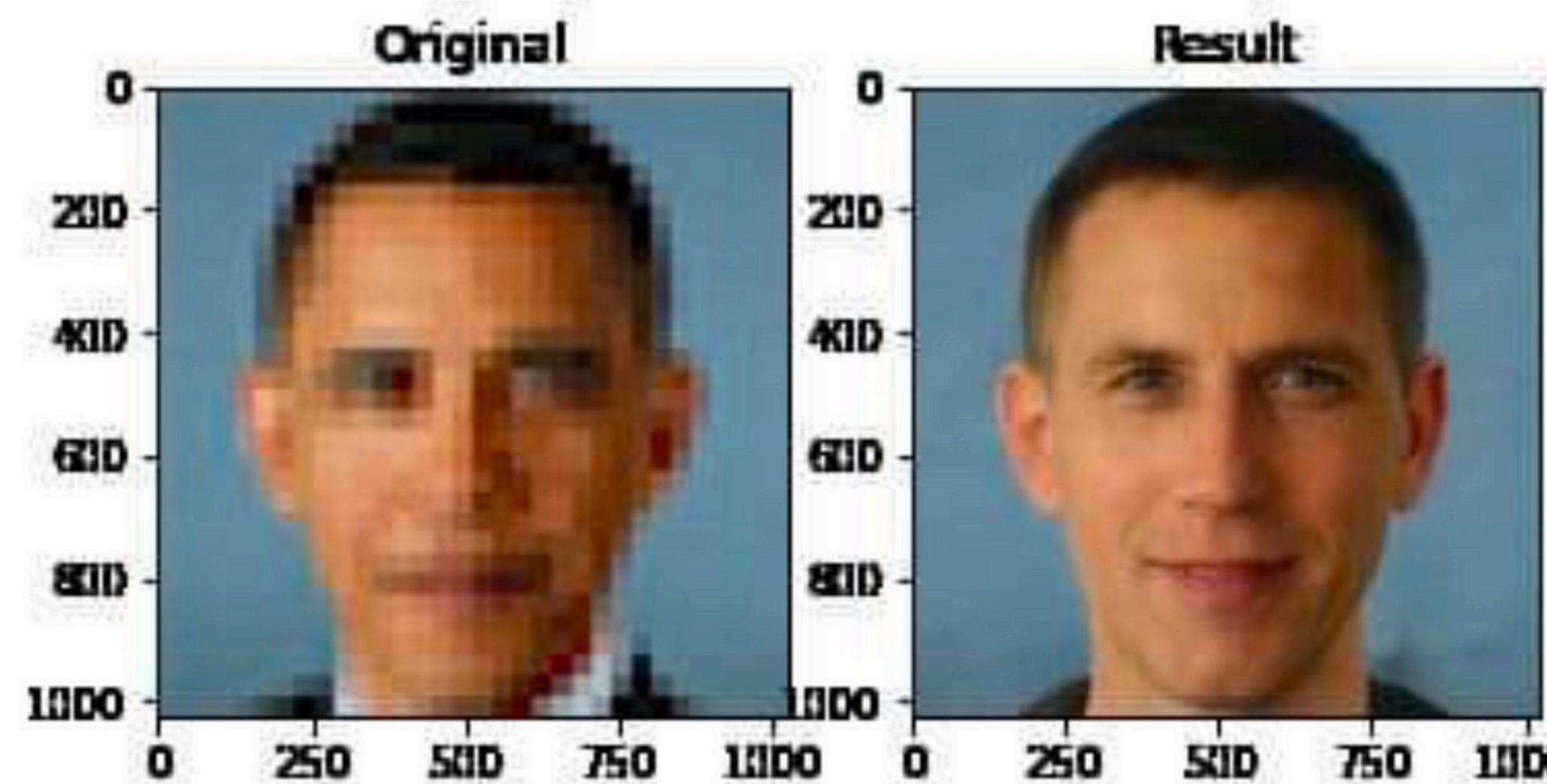
Ocurre cuando los algoritmos priorizan ciertos resultados sobre otros. Esto puede ser involuntario o debido a errores en su diseño.

Sesgo de interpretación

Los modelos de IA pueden ser malinterpretados, llevando a decisiones erróneas o interpretaciones sesgadas.

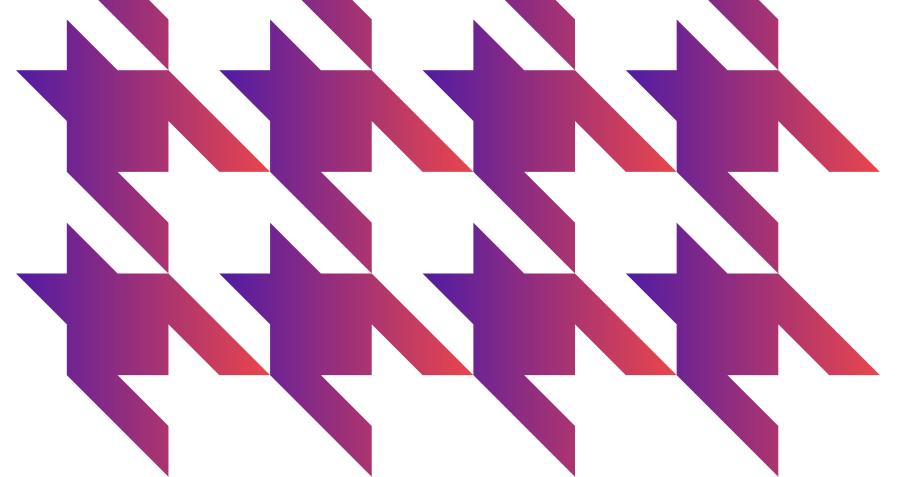


Sesgos en la IA



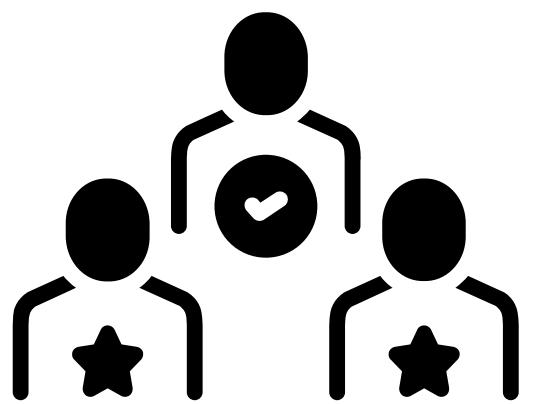
Ejemplo de Sesgo Racial en la IA

Normas y organizaciones



Normas

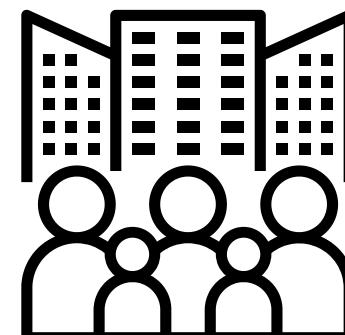
- Transparencia: Exigir que los modelos sean explicables y que se entienda como toman decisiones.
- Responsabilidad: Definir quién es responsable cuando la IA toma decisiones incorrectas o produce daños.
- Privacidad: Protección de los datos personales utilizados en el entrenamiento de los modelos, respetando regulaciones como el GDPR.



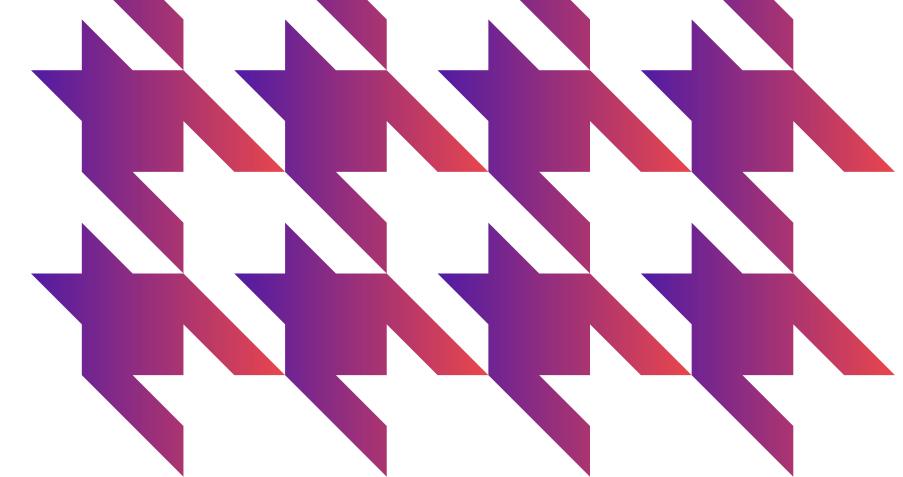
Organizaciones

Dado que los estándares éticos no son la principal preocupación de los ingenieros y científicos de datos en el sector privado, surgieron varias organizaciones para promover la conducta ética en el campo de la inteligencia artificial:

- AlgorithmWatch
- DARPA: La Agencia de Proyectos de Investigación Avanzada de Defensa
- NASCAI: La Comisión de Seguridad Nacional sobre Inteligencia Artificial



Importancia de que las IAs sean transparentes



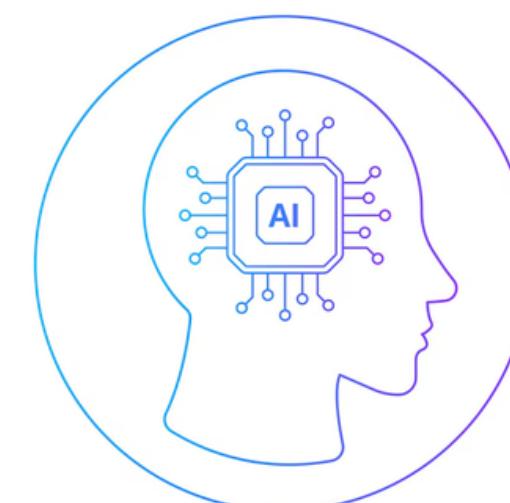
Transparencia

La transparencia en los sistemas de inteligencia artificial (IA) se refiere a la capacidad de entender cómo se toman las decisiones. Es fundamental que los algoritmos no sean "cajas negras", sino que los desarrolladores expliquen cómo se entrena los modelos, cómo procesan los datos y qué criterios siguen para llegar a los resultados.

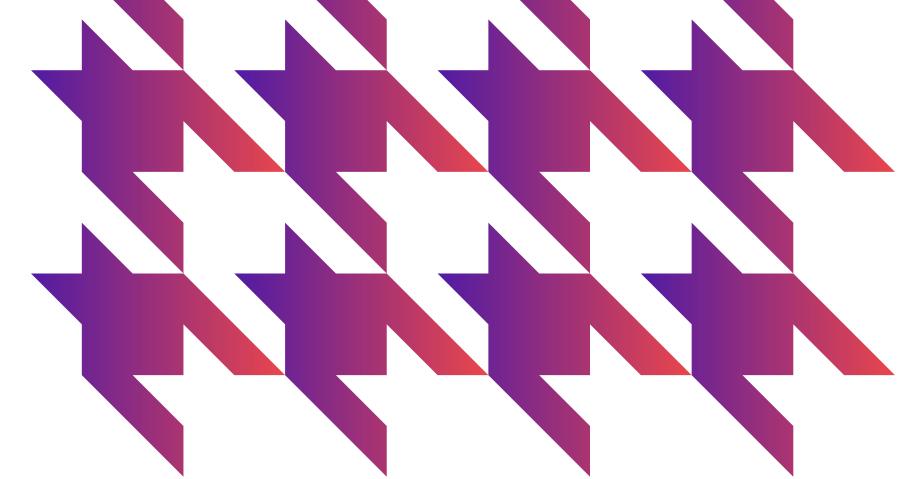


Explicabilidad

La explicabilidad es clave en contextos donde los errores en la IA pueden tener consecuencias graves, como en el sector salud o en el ámbito jurídico. Las organizaciones deben proporcionar herramientas y documentación que permitan comprender las decisiones de los modelos de IA, tanto para expertos como para usuarios finales.

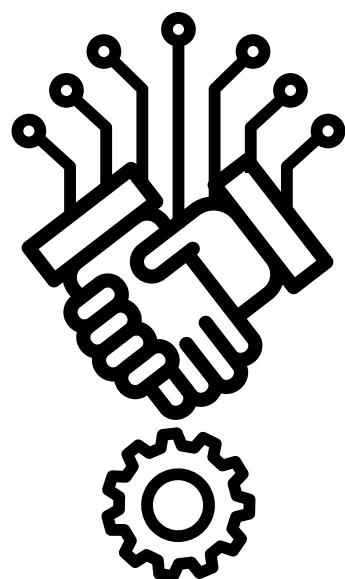


Responsabilidad en la toma de decisiones automática



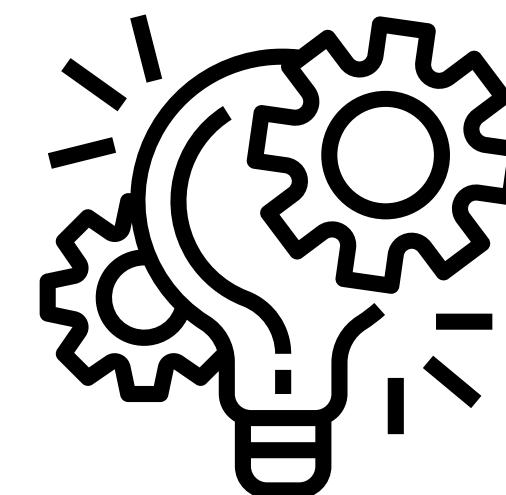
El responsable

- Cuando un sistema de IA toma una decisión errónea, ¿quién es responsable? La falta de claridad en la responsabilidad puede llevar a graves problemas, especialmente en sectores como la salud, la justicia y la contratación.



Estrategia

Definir claramente quién responde cuando una IA falla. Esto puede implicar la creación de equipos de auditoría, la documentación de las decisiones algorítmicas y la implementación de políticas para corregir errores.



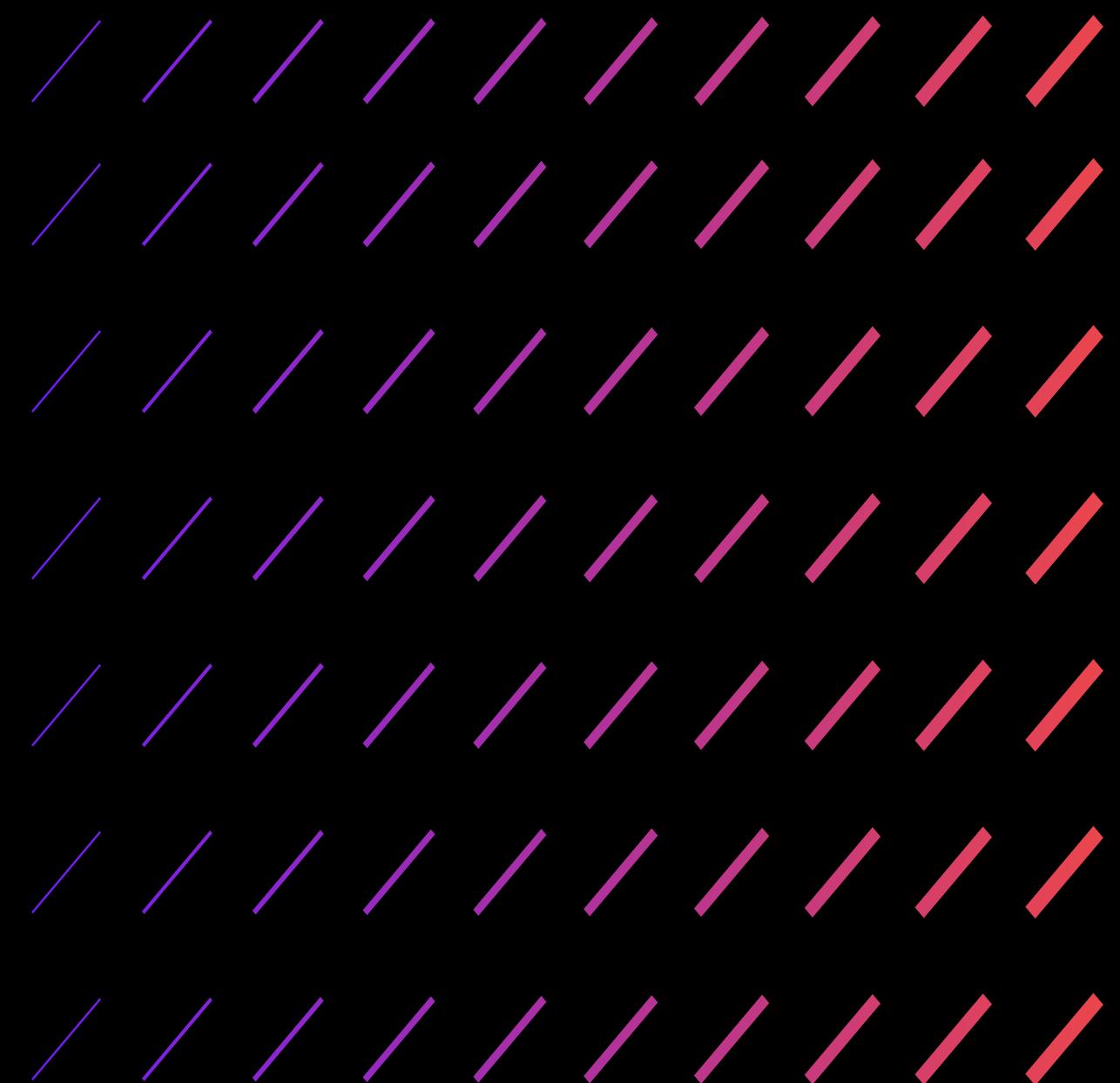
Conclusiones



¡Gracias!

“Una computadora nunca puede ser responsable. Por lo tanto,
una computadora nunca debe tomar una decisión.”

- IBM





Referencias

IBM What is AI Ethics?

What is AI Ethics?

Watch on YouTube

IBM

¿Qué es la ética de la IA?

La ética de la IA es un marco que guía a los científicos de datos e investigadores para crear sistemas de IA de manera ética que beneficien a l...

ibm.com / Sep 11

UNESCO

Ethics of Artificial Intelligence

The Recommendation

unesco.org