



SESGOS Y DISCRIMINACIÓN EN LA IA



Comité de
Participación Social
JALISCO



INTRODUCCIÓN

- El cerebro humano es un órgano complejo que funciona a través de conexiones conscientes e inconscientes.
- El sesgo se define como "simplificaciones o distorsiones generalizadas en el juicio y el razonamiento que afectan sistemáticamente la toma de decisiones human.
- El sesgo de la IA (o sesgo algorítmico) ocurre cuando un sistema de inteligencia artificial produce resultados sistemáticamente distorsionados, injustos o discriminatorios.



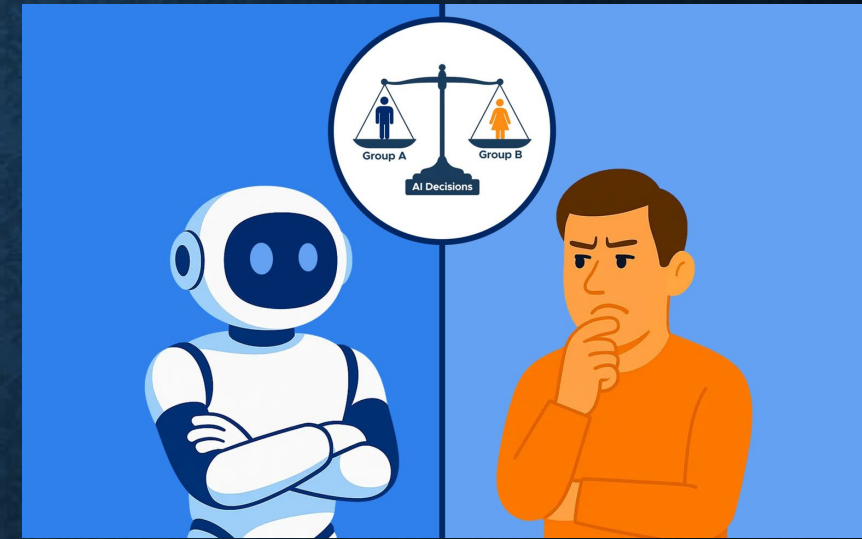
INTRODUCCIÓN

- Los modelos de IA aprenden a partir de datos generados por humanos y, a menudo, heredan, reflejan o amplifican los prejuicios históricos.
- El sesgo en la IA no es meramente un problema técnico, sino un desafío social.
- Los sistemas de IA se integran cada vez más en los procesos de toma de decisiones en la salud, la contratación, el cumplimiento de la ley y otras áreas críticas.



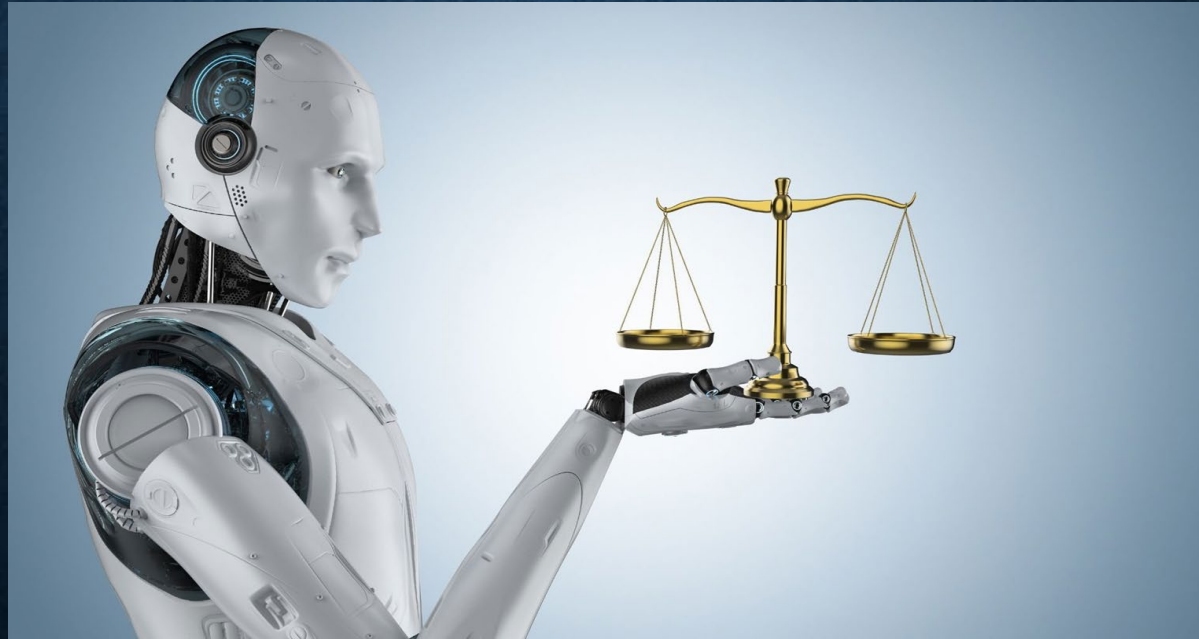
CONSECUENCIAS

- Los sistemas de IA absorben los sesgos de la sociedad que pueden quedar incrustados silenciosamente en las montañas de datos con los que son entrenados.
- Los sesgos llevan discriminación, injusticia y estigma.
- Los sesgos en términos de raza, sexo, género, edad, estatus socioeconómico y capacitismo socavan los principios de justicia y equidad.



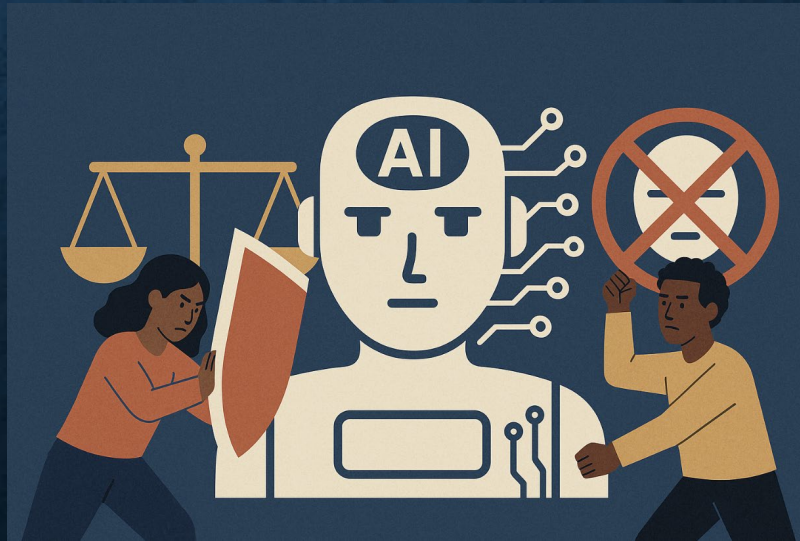
CONSECUENCIAS

- Existe un riesgo elevado de que el sesgo lleve a un aumento en la capacidad de generar un daño a terceros.
- Por ejemplo, en temas de salud puede dar lugar a resultados deficientes o erróneos por parte del sistema de IA.



SESGO IA

- **El sesgo de la IA, también llamado sesgo de aprendizaje automático o sesgo algorítmico, se refiere a la ocurrencia de resultados sesgados debido a sesgos humanos que distorsionan los datos de entrenamiento originales o el algoritmo de IA, lo que conduce a resultados distorsionados y un resultado potencialmente dañino.**



CASO: DEPARTMENT FOR WORK AND PENSIONS (DWP)

- En el Reino Unido un sistema de aprendizaje automático usado para detectar posibles fraudes en el crédito mostró que la IA seleccionaba incorrectamente a ciertos grupos con mayor frecuencia para investigación.



CASO: COMPAS

- El COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) usado en varios estados para evaluar el riesgo de reincidencia de acusados y ayudar en decisiones de fianza, sentencia o libertad condicional.
- Una investigación demostró que el algoritmo clasificaba erróneamente a personas afroamericanas con falsos positivos comparados con las personas blancas.



SESGOS Y OPACIDAD

- **Los sesgos dan lugar a una falta de trazabilidad, es decir, o responsabilidades poco claras y dificulta la rendición de cuentas.**
- **Suman al problema de la caja negra.**
- **Datos que reflejan la inequidad social.**



TIPOS DE SESGOS EN LA IA

- Se dividen en tres categorías principales:

1. Sesgo de Entrada (input bias)

2. Sesgo del Sistema (system bias)

3. Sesgo de Aplicación (application bias)



SESGO DE ENTRADA

- Se refiere a los sesgos presentes en los datos de entrada utilizados para el entrenamiento de los algoritmos.



SESGOS DEL SISTEMA

- Sesgo en el diseño y desarrollo de los algoritmos. Estos sesgos pueden originarse en el procesamiento y validación de los algoritmos.



SESGO DE APLICACIÓN

- Surge del uso de los sistemas de IA en la práctica y es propenso a una amplia gama de sesgos humanos.



GOBERNANZA DE LA IA

- La gobernanza de la IA crea un conjunto de políticas, prácticas y marcos para guiar el desarrollo y uso responsable de las tecnologías de IA.
- Igual de importante es un sistema de 'humano en el circuito' (human-in-the-loop) para ofrecer opciones o hacer recomendaciones.



MITIGAR LOS SESGOS

1. Seleccionar el modelo de aprendizaje correcto.

1. Entrenar con los datos correctos.

1. Elegir un equipo equilibrado.



MITIGAR LOS SESGOS

4. Realizar el procesamiento de datos con conciencia.

5. Monitoreo y seguimiento.



MÉXICO

- Los principios de Chapultepec señalan:

“Antes de automatizar, hay que comprender a quién y a qué afecta”.



MARCO NORMATIVO

- Constitución Política de los Estados Unidos Mexicanos.
- Ley Federal para Prevenir y Eliminar la Discriminación.
- Ley General de Protección de Datos Personales en Posesión de Sujetos Obligados (Arts 3, fr. XV y 10).



CONCLUSIONES

- La IA puede expandir derechos y eficientar servicios, también puede reproducir y profundizar las desigualdades existentes si se construye desde datos sesgados.
- Los gobiernos tienen la obligación garantizar equidad y proteger a la población de los riesgos de los sesgo a la hora de implementar algún sistema de IA.





EDUCACIÓN

Comisión de la RED NACIONAL CPC

¡MUCHAS GRACIAS!



RED NACIONAL
DE COMITÉS DE PARTICIPACIÓN
CIUDADANA



Comité de
Participación Social
JALISCO



INKLUDERING