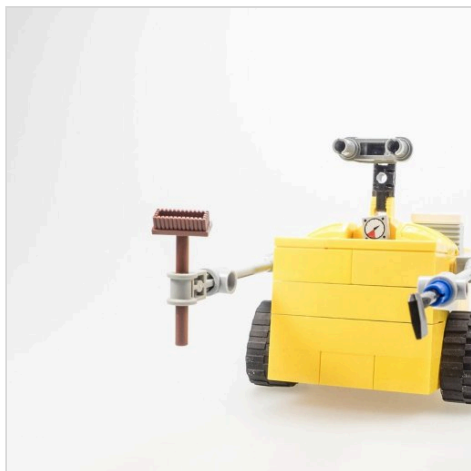


L U M B R E



Sesgo y equidad en modelos

Materia: Ética e IA Responsable · Responsabilidad Profesional

11/10/2026

Antes de empezar

Un modelo aprende patrones, incluidos los injustos. El sesgo se cuela por los datos, por la formulación y por el uso, y puede replicar discriminaciones. Medir la equidad es el primer paso para no automatizar la injusticia.

QUÉ APRENDERÁS

- Identificar fuentes de sesgo en datos y modelos y aplicar criterios de equidad para medir y mitigar sus efectos.

Sesgo y equidad en modelos

El sesgo se hereda y escala

Un modelo de crédito que aprueba menos a un grupo, un reconocedor facial que falla más con ciertos rostros: el sesgo no es un defecto teórico, es daño real automatizado. Reconocerlo y medirlo es parte del trabajo.

VIDEO

La justicia y la equidad en la inteligencia artificial

Cómo el sesgo se cuela en los modelos y qué criterios permiten medir y mitigar la injusticia.



Escanea para verlo en Lumbre

<https://www.youtube.com/watch?v=nfNcNM6d6o8>

De dónde viene el sesgo

- Datos históricos que reflejan decisiones sesgadas.
- Subrepresentación de ciertos grupos en el entrenamiento.
- Variables proxy que codifican atributos protegidos.

Sesgo en los datos

Si el pasado discriminó, los datos lo documentan. Un modelo que prediga «quién recibió el puesto» aprenderá a excluir a quien históricamente fue excluido.

Sesgo de medición

La propia etiqueta puede estar sesgada: «detenido» no equivale a «delincuente». Optimizar por una mala variable perpetúa el error con aire de objetividad.

Variables proxy

Aunque se quite el atributo protegido, otros correlacionados lo sustituyen: código postal como sustituto de etnia. La discriminación indirecta es sutil.

Qué es equidad

No hay una definición única. Equidad puede significar igualdad de error entre grupos, de oportunidades o de tratamiento. Algunas son matemáticamente incompatibles.

Métricas de disparidad

- Diferencia de tasas de resultado positivo entre grupos.
- Brecha en falsos positivos y falsos negativos.
- Paridad estadística frente a igualdad de oportunidad.

El trade-off de la equidad

Hacer un modelo justo por un criterio puede empeorarlo por otro, y a veces reducir precisión global. Elegir el criterio es un juicio de valores, no solo técnico.

Mitigación en datos

Rebalancear, aumentar o limpiar los datos reduce desequilibrios antes de entrenar. Ataca la raíz, aunque no borra el sesgo histórico por completo.

Mitigación en el modelo

- Reponderar la pérdida por grupo.
- Restricciones de equidad como parte de la optimización.
- Post-procesar umbrales distintos por grupo.



El sesgo inclina la balanza: medir la disparidad por grupo es el primer paso para no automatizar la injusticia.

Auditoría y contextos sensibles

Antes de desplegar sistemas con impacto —crédito, sanidad, justicia— conviene auditar su comportamiento por grupo y documentar el riesgo asumido.

El mito de la neutralidad

«El algoritmo es neutro» es falso: decidió qué predecir, con qué datos y qué umbral. La técnica nunca está libre de valores; hay que hacerlos explícitos.

Errores frecuentes

- Borrar el atributo protegido y creer que con eso desaparece el sesgo.
- Optimizar por una etiqueta viciada que confunde el síntoma con la realidad.
- Dar por neutro un modelo sin medir su comportamiento por grupo.

Ejemplo resuelto: un cribado que penaliza un barrio

Un modelo de crédito usa el código postal como señal de riesgo y aprueba menos a un barrio históricamente desatendido. Al medir la disparidad se ve la brecha; el equipo

identifica el proxy, lo retira, rebalancea los datos y vuelve a auditar por grupo antes de desplegar.

1. medir las tasas de resultado positivo por grupo
2. diagnosticar la fuente: datos, etiqueta o variables proxy
3. mitigar rebalanceando datos o reponderando la pérdida
4. elegir y documentar el criterio de equidad asumido
5. auditar el impacto residual antes de desplegar

¿Para qué sirve en la realidad?

Sistemas de crédito, contratación, sanidad o justicia tienen impacto directo sobre personas; medir y mitigar el sesgo evita reproducir discriminaciones históricas a escala y es exigencia creciente de la normativa.

1. Una variable proxy genera discriminación indirecta porque:

- ☐ contiene directamente el atributo protegido
- ☐ está correlacionada con el atributo protegido aunque no lo sea
- ☐ aumenta la precisión
- ☐ reduce el sesgo de medición

2. Existen definiciones de equidad que son matemáticamente incompatibles entre sí.

- ☐ Verdadero
- ☐ Falso

ACTIVIDAD INTERACTIVA DISPONIBLE EN LUMBRE

3. Equidad



4. Ordena el trabajo para abordar el sesgo:

Elementos:

- calcular disparidades por grupo
- identificar la fuente del sesgo
- aplicar mitigación en datos o modelo
- reverificar el impacto residual

Medir	. . .
Diagnosticar	. . .
Mitigar	. . .
Auditar	. . .

Equidad en aprendizaje automático [https://en.wikipedia.org/wiki/Fairness_\(machine_learning\)](https://en.wikipedia.org/wiki/Fairness_(machine_learning)) Criterios para medir y mitigar sesgos en modelos.

5. Une cada concepto de equidad con su definición:

5.	sesgo de datos	-----
5.	variable proxy	-----
5.	disparidad	-----
5.	mitigación	-----

Opciones: historial que codifica decisiones injustas · atributo que sustituye al protegido · diferencia de resultados entre grupos · reponderar o rebalancear para reducir sesgo

Equidad

Sesgo y equidad

Origen

datos históricos

subrepresentación

variables proxy

Medición

disparidad de tasas

falsos positivos

paridad estadística

Mitigación

rebalancear datos

reponderar pérdida

ajustar umbrales

Gobierno

auditoría

contextos sensibles

valores explícitos

El mapa mental interactivo es interactivo en la versión online.



Sesgo algorítmico https://en.wikipedia.org/wiki/Algorithmic_bias Sesgos sistemáticos en los resultados de un sistema informático.

6. Tu modelo de selección de personal reproduce el sesgo histórico. ¿Lo corriges en datos, en modelo o cambiando el objetivo? Razona qué criterio de equidad priorizarías.

Referencia curricular

Materia:

Ética e IA Responsable · Responsabilidad Profesional

Versión interactiva completa (con actividades, videos y autoevaluación):



<https://univ.lumbre.lat/object/cdia-etic-001-sesgo-y-equidad-en-modelos>

Material educativo generado desde Lumbre · lumbre.lat