

Transcripción de la voz en off

DÉCLIC - Los sesgos algorítmicos





En teoría, una inteligencia artificial es neutral. ¿en la práctica? No siempre. Si aprende a partir de datos sesgados, una IA puede reproducir —o incluso amplificar— estos sesgos, a veces incluso sin que nadie lo note. A esto se le llama sesgo algorítmico.

Por ejemplo, si un modelo de reclutamiento aprende a partir de currículums de hombres blancos de una misma escuela, el modelo podría favorecer este perfil. Los algoritmos de recomendación también pueden favorecer ciertos tipos de contenidos tóxicos.

Los sesgos pueden venir de archivos discriminatorios, de estereotipos, o de datos demasiado homogéneos. Otros pueden venir del propio código, de los criterios usados para entrenar el modelo, o de la técnica algorítmica elegida.

¿La buena noticia? Existen soluciones. Diversificar los datos de entrenamiento, probar los modelos usando auditorías de sesgos que analizan el algoritmo con muestreo inteligente. O también, incluir desde el principio a expertos en ética y en ciencias sociales desde la concepción de los algoritmos.

Muchos investigadores en ciencias digitales combaten a diario estos sesgos algorítmicos, trabajando sobre la IA confiable, sobre todo a través de auditorías.

La IA no tiene prejuicios... a menos que se los demos.

Conocer y entender los sesgos algorítmicos es el primer paso para prevenirlos.

Encuentra todos los episodios de DÉCLIC en nuestro canal de YouTube, en Inria.cl y en nuestras redes sociales.

Encuentra todos los episodios de DÉCLIC en nuestro canal de YouTube, en Inria.cl y en nuestras redes sociales.