¿Por qué una plataforma nos muestra cada vez más contenido parecido?

Porque los algoritmos aprenden de nuestros datos y de nuestros comportamientos.

IDEA CLAVE: La IA no es completamente neutral: puede reproducir y amplificar sesgos humanos.

IA y el SESGO

Un algoritmo presenta sesgo cuando favorece una opción frente a otra de manera sistemática.

Una forma sencilla de entenderlo: **predicen el futuro a partir del pasado.**

PROBLEMA: Si el pasado contiene desigualdades o errores, la IA puede aprenderlos.

EJEMPLO 1:

Buscas información sobre un tema en una red social.

El algoritmo aprende de tus intereses.

Empieza a mostrarte cada vez más contenido parecido.

IA y el SESGO

Resultado: puedes descubrir menos opiniones y contenidos diferentes.

IDEA CLAVE: La personalización puede crear una “burbuja” de contenidos.

IA y el SESGO

EJEMPLO 2 CONTRATACIÓN:

Un sistema de selección se entrena con muchos más currículums de hombres que de mujeres.

Aunque el género no aparezca como dato, otros datos pueden estar relacionados con ese patrón.

El sistema puede aprender que los candidatos masculinos son la opción “preferida”.

IDEA CLAVE: Un algoritmo puede reproducir desigualdades presentes en los datos.

EJEMPLO 3 RECONOCIMIENTO FACIAL:

Si un sistema se entrena principalmente con rostros de un grupo de población, puede funcionar mejor para ese grupo.

Con menos diversidad de imágenes, puede aumentar el número de errores al reconocer otros rostros.

IA y el SESGO

IDEA CLAVE: Los datos de entrenamiento deben representar la diversidad de las personas que utilizarán el sistema.

La IA puede reforzar creencias o tendencias que ya tenemos.

EJEMPLO: buscas contenido sobre que la tierra es plana y el sistema te recomienda cada vez más contenido que lo apoya.

IA y el SESGO

IDEA CLAVE: Si solo vemos información que confirma nuestras ideas, podemos terminar creyendo que todas las personas piensan igual.

EL SESGO ALGORÍTMICO EN LA IA/ Sesgo de recuerdo

Se produce cuando las etiquetas o clasificaciones se aplican de manera inconsistente.

EJEMPLO: diferentes personas clasifican películas de un festival siguiendo criterios distintos.

IA y el SESGO

El sistema aprende de esas etiquetas y puede ignorar o clasificar mal determinados títulos.

IDEA CLAVE: Si las categorías de los datos no son coherentes, los resultados también pueden serlo.

Prof.: Carmen Pérez González

EL SESGO ALGORÍTMICO EN LA IA/ Sesgo de exclusión

Aparece cuando se omite información durante la recogida o introducción de datos.

Lo que **NO está en los datos también importa.**

IA y el SESGO

EJEMPLO: si un conjunto de datos deja fuera a un grupo de personas, el modelo puede funcionar peor para ese grupo.

IDEA CLAVE: Una IA no puede aprender correctamente de información que nunca recibió.

Prof.: Carmen Pérez González

EL SESGO ALGORÍTMICO EN LA IA/ Sesgo de prejuicio

Los datos pueden contener estereotipos o valoraciones humanas.

Si estos patrones llegan al entrenamiento, el modelo puede reproducirlos.

IA y el SESGO

ESTEREOTIPO: atribuir determinadas características a un grupo.

PREJUICIO: valorar positiva o negativamente a una persona o grupo antes de disponer de información suficiente.

IDEA CLAVE: Los prejuicios humanos pueden quedar incorporados en los sistemas.

Prof.: Carmen Pérez González

IA y el SESGO

El sesgo de muestra o selección se produce cuando los datos utilizados **NO representan adecuadamente a la población.**

EJEMPLO: un estudio sobre salud mental que solo incluye personas mayores y no jóvenes. Las conclusiones serán incompletas porque falta una parte importante de la población.

IDEA CLAVE: Una muestra poco representativa produce resultados poco fiables.

EL SESGO ALGORÍTMICO EN LA IA/ Sesgo de medición

Los datos recogidos pueden ser defectuosos, incompletos o estar mal medidos.

EJEMPLO: estudiar la violencia en los medios analizando solo medios nacionales e ignorando los locales.

IA y el SESGO

El resultado no representa toda la realidad.

IDEA CLAVE: Si medimos mal una realidad, la IA también aprenderá una realidad distorsionada.

Prof.: Carmen Pérez González

IA y el SESGO

Puede aparecer cuando los equipos que diseñan un sistema son poco diversos.

Un equipo con experiencias muy similares puede detectar menos problemas que afectan a otros grupos.

IDEA CLAVE: La diversidad de las personas que diseñan y revisan la IA ayuda a detectar sesgos

EL SESGO ALGORÍTMICO EN LA IA/ Sesgo cognitivo

Las personas no siempre interpretamos la información de manera completamente objetiva.

Los **SESGOS COGNITIVOS** son tendencias humanas que pueden influir en nuestras decisiones y juicios.

IA y el SESGO

Si nuestras decisiones influyen en los datos, las etiquetas o el diseño, esos sesgos pueden llegar al sistema.

IDEA CLAVE: Antes de corregir el sesgo de una IA, hay que preguntarse de dónde proceden sus datos y decisiones.

Prof.: Carmen Pérez González

IA y el SESGO

TRES SESGOS COGNITIVOS IMPORTANTES

DISPONIBILIDAD: damos más importancia a la información fácil de recordar o encontrar.

ANCLAJE: damos demasiado peso a la primera información que recibimos.

ATRIBUCIÓN: relacionamos características o comportamientos con determinadas personas o grupos.

IDEA CLAVE: Nuestros propios atajos mentales pueden influir en los sistemas que construimos.

Prof.: Carmen Pérez González

Los datos más recientes pueden recibir demasiado peso frente a información anterior.

EJEMPLO: una plataforma musical recomienda canciones basándose casi exclusivamente en tus últimas escuchas.

IA y el SESGO

IDEA CLAVE: Lo último que hemos hecho no siempre representa todos nuestros intereses.

El sistema puede reforzar las creencias o patrones que ya tiene el usuario.

Puede mostrar principalmente contenido que coincide con lo que ya pensamos.

IA y el SESGO

IDEA CLAVE: La personalización puede reducir la exposición a ideas diferentes.

Puede aparecer cuando los resultados favorecen a la empresa que controla la plataforma.

IA y el SESGO

EJEMPLO: una plataforma de comercio electrónico puede priorizar productos de marcas con las que tiene acuerdos.

IDEA CLAVE: Hay que preguntarse QUIÉN diseña el algoritmo, QUIÉN se beneficia y QUÉ criterios utiliza.

EJEMPLO: CASO AMAZON · 2018

Amazon descubrió que un sistema de contratación discriminaba a las mujeres.

El sistema había aprendido de datos históricos de la empresa.

IA y el SESGO

Como la plantilla histórica estaba formada mayoritariamente por hombres, el algoritmo aprendió ese patrón.

IDEA CLAVE: Los datos históricos pueden reproducir desigualdades del pasado.

EJEMPLO: En 2016 se denunciaron problemas relacionados con sistemas utilizados para evaluar el riesgo de reincidencia delictiva.

IA y el SESGO

El debate mostró que determinados sistemas podían producir resultados diferentes según características relacionadas con el perfil racial.

IDEA CLAVE: Cuando una IA participa en decisiones importantes, los sesgos pueden tener consecuencias reales sobre las personas.

EL SESGO ALGORÍTMICO EN LA IA/ ¿De dónde vienen?

IA y el SESGO

DATOS: pueden ser incompletos o poco representativos.

PERSONAS: introducen valores, prejuicios y decisiones.

ETIQUETAS: pueden aplicarse de forma desigual.

DISEÑO: los criterios elegidos pueden favorecer unas opciones.

USO: las plataformas pueden utilizar los resultados con determinados intereses.

IDEA CLAVE: El sesgo puede aparecer en muchas etapas, no solo en el algoritmo.

Prof.: Carmen Pérez González

IA y el SESGO

NO completamente.

Pero sí podemos **DETECTARLOS, MEDIRLOS y REDUCIRLOS.**

Para ello necesitamos:

- Datos diversos y representativos.
- Métodos de verificación y evaluación del sesgo.
- Equipos diversos.
- Revisión de los resultados.
- Transparencia sobre cómo se utilizan los sistemas.
- Formación de las personas que utilizan IA.

Prof.: Carmen Pérez González

IA y el SESGO

ANTES DE USAR UNA IA:

¿De dónde vienen los datos?

¿A quién representan y a quién dejan fuera?

¿Quién ha diseñado el sistema?

¿Qué criterios utiliza?

¿Quién puede verse perjudicado?

¿Quién se beneficia?

¿Podemos revisar o cuestionar el resultado?

Prof.: Carmen Pérez González

IA y el SESGO

LA IA APRENDE DE DATOS.

SI LOS DATOS TIENEN SESGOS: LA IA PUEDE REPRODUCIRLOS.

SI LOS AMPLIFICA; EL PROBLEMA PUEDE CRECER.

NO ES SOLO UN PROBLEMA TÉCNICO: también es social, cultural y ético.

OBJETIVO: diseñar y utilizar una IA más diversa, inclusiva, transparente y responsable.

No podemos eliminar los sesgos por completo, pero sí aprender a **reconocerlos y a minimizarlos.**

Prof.: Carmen Pérez González

IA y el SESGO

MUCHAS GRACIAS

Prof.: Carmen Pérez González