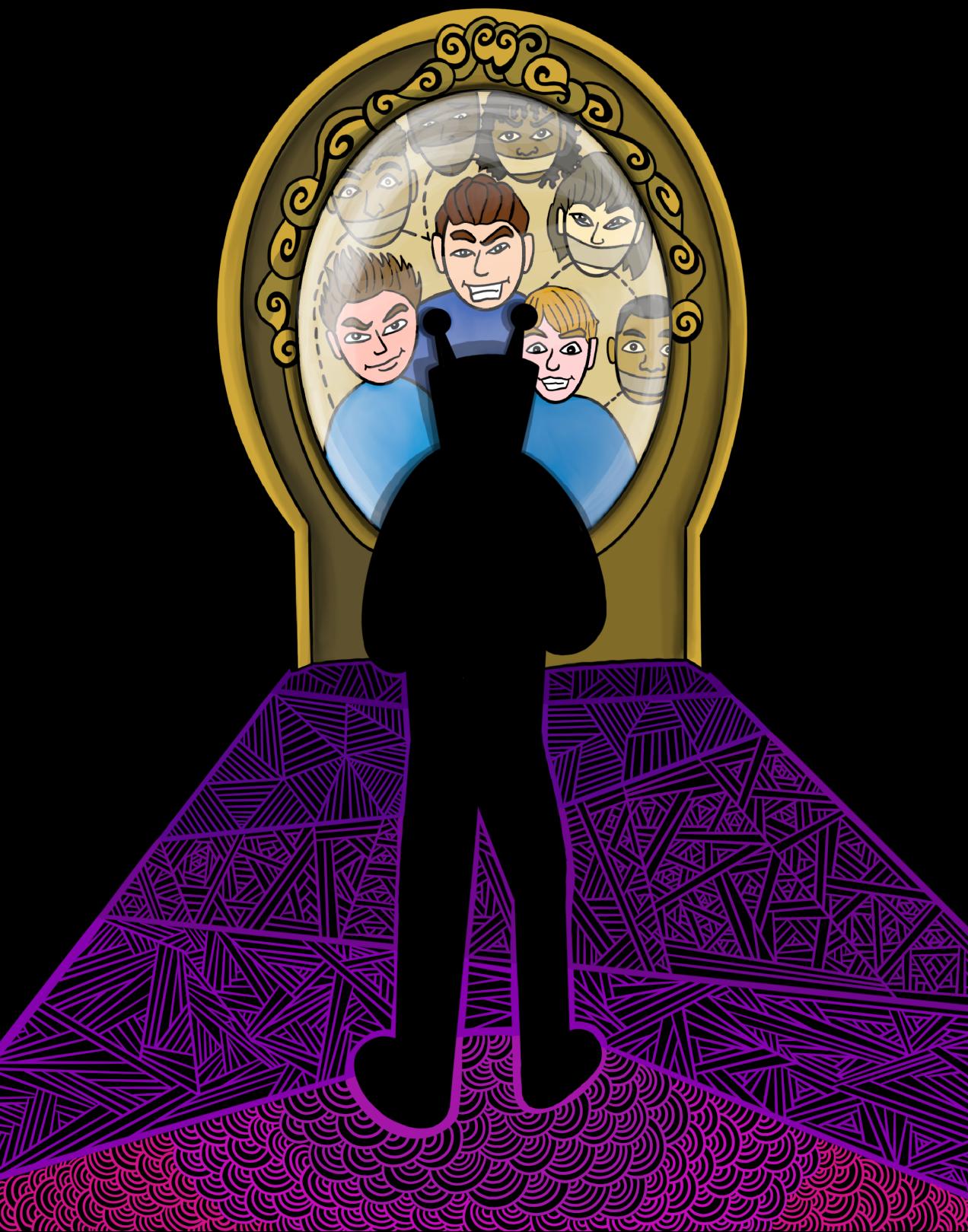


Somos IA #4: TODO SOBRE ESE SESGO



TERMINOS DE USO

Todos los paneles de este libro de historietas tienen licencia CC BY-NC-ND 4.0. Consulte la página de licencias para obtener detalles sobre cómo puede usar este material gráfico.

TL;DR: Se puede usar paneles/grupos de paneles en presentaciones/artículos, siempre y cuando

1. Se proporcione la cita adecuada
2. No se realicen modificaciones a los paneles individuales

Citar como:

Julia Stoyanovich y Falaah Arif Khan. “Todo sobre el sesgo”. Somos IA, Vol. 4 (2022) https://dataresponsiblemente.github.io/we-are-ai/comics/vol4_es.pdf

Contacto:

Dirija cualquier consulta sobre el uso de elementos de este cómic a themachinelearnist@gmail.com, con copia a stoyanovich@nyu.edu

HABLEMOS DE LO QUE ENTENDEMOS POR "SESGO" EN LA IA Y CÓMO SURGE.

DECIMOS QUE UNA IA ESTÁ SESGADA SI SU USO PUEDE CONDUCIR A UNA DISCRIMINACIÓN SISTEMÁTICA E INJUSTA CONTRA ALGUNOS INDIVIDUOS O GRUPOS EN FAVOR DE OTROS.

EL SESGO PUEDE PROVENIR DE PATRONES DAÑINOS RECOGIDOS DE LOS PROPIOS DATOS,

O DE CÓMO ESTÁ DISEÑADO EL ALGORITMO,

O DE LOS OBJETIVOS QUE LE ESPECIFICAMOS,

O DE CÓMO LO USAMOS.

EN SU ARTÍCULO SEMINAL DE 1996, BATYA FRIEDMAN Y HELEN NISSENBAUM IDENTIFICARON TRES TIPOS DE SESGOS QUE PUEDEN SURGIR EN LOS SISTEMAS INFORMÁTICOS,

QUE SE REPRESENTAN AQUÍ COMO UN DRAGÓN DE TRES CABEZAS:

PREEEXISTENTE

TÉCNICO

EMERGENTE

RECORDEMOS AHORA LA METÁFORA DEL HORNEADO QUE USAMOS PARA ENTENDER LOS ALGORITMOS BASADOS EN DATOS EN EL VOLUMEN 1.

¡USEMOS AHORA LA MISMA METÁFORA PARA ENTENDER EL SESGO!



EL SESGO PREEEXISTENTE EXISTE INDEPENDIENTEMENTE DEL ALGORITMO Y TIENE SU ORIGEN EN LA SOCIEDAD.

SESCO PREEEXISTENTE
(EN LOS DATOS)

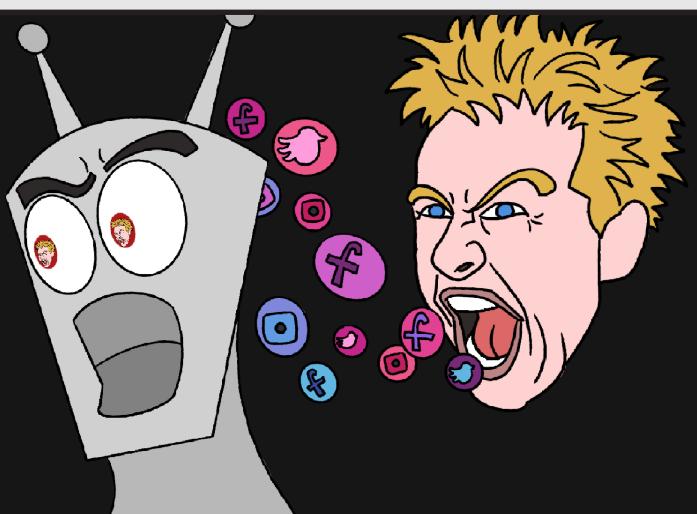
ESTAS SERÍAN LAS NOTAS DE SABOR QUE SE FILTRARÁN EN NUESTRO PAN SI NO PRIORIZAMOS LA PUREZA/FRESCURA DE LOS INGREDIENTES,

O SI DECIDIMOS USAR MASA PREMEZCLADA LISTA PARA USAR.

ESTOS SESGOS EXISTEN EN LA SOCIEDAD Y VIENEN "PRECOCINADOS" EN EL ALGORITMO

DEL SISTEMA DISCRIMINATORIO SUBYACENTE DEL QUE SE RECOLPILARON LOS DATOS:

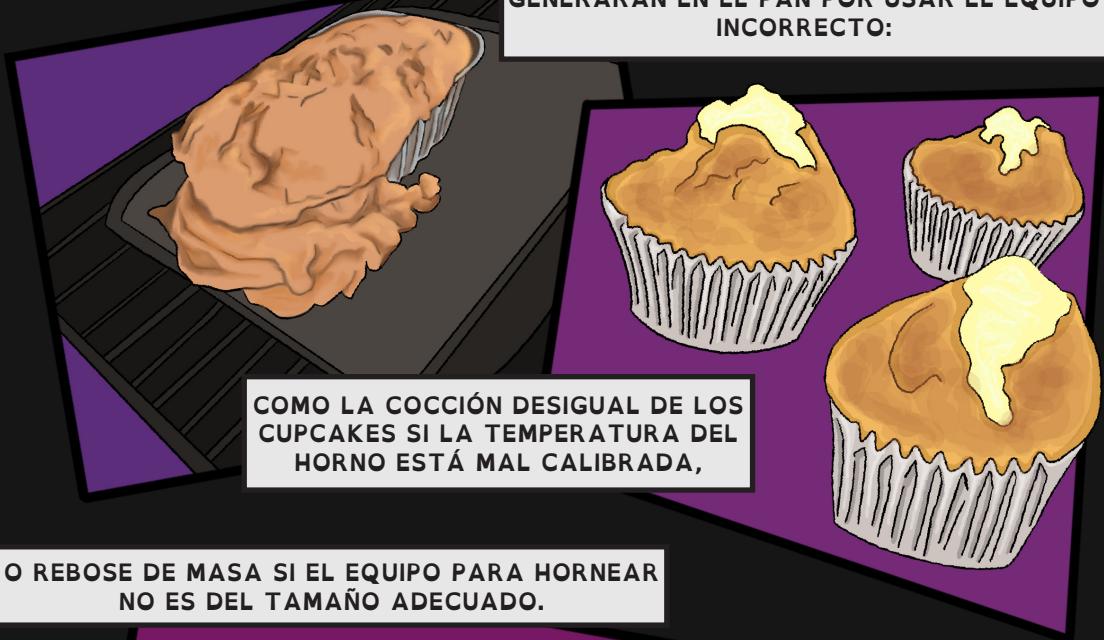
COMO LOS ESTEREOTIPOS RACIALES Y DE GÉNERO QUE LOS MODELOS LINGÜÍSTICOS RECOGEN CUANDO SE ENTRENAN CON DATOS DE LAS REDES SOCIALES.



SESGO TÉCNICO (EN EL SISTEMA TÉCNICO)

EL SESGO TÉCNICO LO INTRODUCE EL PROPIO SISTEMA,
DEBIDO A LA FORMA EN QUE ESTÁ DISEÑADO O FUNCIONA.

ESTAS SERÍAN LAS IMPERFECCIONES QUE SE
GENERARÁN EN EL PAN POR USAR EL EQUIPO
INCORRECTO:



Y VOLVIENDO A LOS SISTEMAS
INFORMÁTICOS:

UN EJEMPLO DESTACADO SON LAS
PLATAFORMAS DE REDES SOCIALES



SESGO EMERGENTE (POR DECISIONES)

EL SESGO EMERGENTE SURGE CON EL TIEMPO,
PORQUE LAS DECISIONES TOMADAS CON LA
AYUDA DEL SISTEMA CAMBIAN EL MUNDO,

LO QUE A SU VEZ AFECTA EL
FUNCIONAMIENTO DEL SISTEMA EN EL FUTURO.

PENSEMOS EN LOS CAMBIOS DE COMPORTAMIENTO QUE
SURGIRÁN COMO RESULTADO DE NUESTRO HORNEADO:

¿QUÉ PASARÍA SI NOS CONVIRTIÉRAMOS EN
MAESTROS REPOSTEROS E INCORPORÁRAMOS,
SIN DARNOS CUENTA, EL PAN COMO UNA
CONSTANTE EN NUESTRA DIETA?



¿O SI HICIÉRAMOS PAN CON TANTA FRECUENCIA
QUE NADIE A NUESTRO ALREDEDOR QUISIERA
PROBAR UNA REBANADA MÁS?

O REFLEXIONEMOS EN EL HECHO DE QUE NUESTRA
IDEA DE "A QUÉ DEBE SABER EL PAN" ESTÁ
MOLDEADA POR LA POPULARIDAD DE PRODUCTOS
COMO "WONDER BREAD".



DEL MISMO MODO, PENSEMOS EN CÓMO
NUESTRA EXPOSICIÓN A LAS NOTICIAS, Y
A LA INFORMACIÓN EN GENERAL,

ESTÁ FORMADA POR ALGORITMOS QUE
SELECCIONAN NOTICIAS DE REDES SOCIALES CON
PUBLICACIONES POPULARES Y "DE TENDENCIA".



PARA CONCRETAR UN POCO MÁS, VEAMOS EJEMPLOS DEL MUNDO REAL DE SESGO ALGORÍTMICO.

TOMEMOS EL SECTOR “CONTRATACIÓN” COMO UN DOMINIO REPRESENTATIVO EN EL QUE LOS ALGORITMOS SE UTILIZAN CADA VEZ MÁS PARA TOMAR DECISIONES CRÍTICAS DE MANERA MÁS “EFICIENTE”.



UNO DE LOS PRIMEROS INDICIOS DE QUE EXISTE UN MOTIVO DE PREOCUPACIÓN SE PRODUJO EN 2015, CON LOS RESULTADOS DEL ESTUDIO ADFISHER DE LA UNIVERSIDAD CARNEGIE MELLON [2].

LOS INVESTIGADORES REALIZARON UN EXPERIMENTO, EN EL QUE CREARON DOS CONJUNTOS DE PERFILES SINTÉTICOS DE USUARIOS DE LA WEB QUE ERAN IGUALES EN TODOS LOS ASPECTOS.

—EN TÉRMINOS DE DATOS DEMOGRÁFICOS, INTERESES DECLARADOS Y PATRONES DE NAVEGACIÓN—

CON UNA SOLA EXCEPCIÓN: SU GÉNERO DECLARADO, MASCULINO O FEMENINO.

LOS INVESTIGADORES DEMOSTRARON QUE GOOGLE MOSTRABA ANUNCIOS DE UN SERVICIO DE ORIENTACIÓN PROFESIONAL PARA PUESTOS EJECUTIVOS BIEN REMUNERADOS CON MUCHA MÁS FRECUENCIA AL GRUPO DE HOMBRES QUE AL GRUPO DE MUJERES.

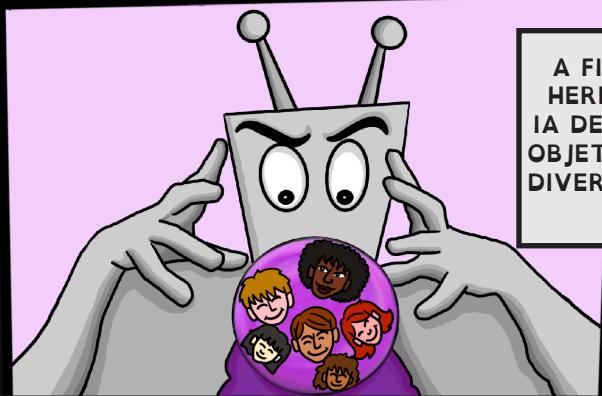
ESTO TRAE RECUERDOS DE LA ÉPOCA EN QUE ERA LEGAL ANUNCIAR PUESTOS DE TRABAJO POR GÉNERO EN LOS PERIÓDICOS. ESTA PRÁCTICA FUE PROHIBIDA EN LOS EE. UU. EN 1964, PERO PERSISTE EN EL ENTORNO PUBLICITARIO EN LÍNEA.

MÁS TARDE SE DEMOSTRÓ QUE PARTE DE LA RAZÓN POR LA QUE ESTO ESTABA SUCEDIENDO ES LA MECÁNICA DEL PROPIO SISTEMA DE ORIENTACIÓN DE ANUNCIOS, COMO UN ARTEFACTO DEL PROCESO DE LICITACIÓN.

ESTO ES SESGO TÉCNICO EN ACCIÓN.

[2] Women less likely to be shown ads for high-paid jobs on Google, study shows. Guardian (2015).

AVANDEMOS EN EL TIEMPO Y AVANDEMOS TAMBIÉN A LA SIGUIENTE ETAPA DEL EMBUDO DE CONTRATACIÓN: REANUDAR LA SELECCIÓN.



A FINES DE 2018, SE INFORMÓ QUE LA HERRAMIENTA DE RECLUTAMIENTO DE IA DE AMAZON, DESARROLLADA CON EL OBJETIVO DECLARADO DE AUMENTAR LA DIVERSIDAD DE LA FUERZA LABORAL, DE HECHO HIZO LO CONTRARIO:

EL SISTEMA SE ENSEÑÓ A SÍ MISMO QUE LOS CANDIDATOS MASCULINOS ERAN PREFERIBLES A LAS CANDIDATAS FEMENINAS.

PENALIZABA LOS CURRÍCULUMS QUE INCLUÍAN LA PALABRA "FEMENINO", COMO EN "CAPITANA DEL CLUB DE AJEDREZ FEMENINO".

Y REBAJÓ LA CALIFICACIÓN DE LAS GRADUADAS DE DOS UNIVERSIDADES EXCLUSIVAMENTE PARA MUJERES.

LOS RESULTADOS SE ALINEARON Y REFORZARON UN MARCADO DESEQUILIBRIO DE GÉNERO EN LA FUERZA LABORAL.

ESTE ES UN SESGO EMERGENTE EN ACCIÓN:

UN GERENTE DE CONTRATACIÓN A QUIEN UNA HERRAMIENTA DE IA SUGIERE REPETIDAMENTE EL MISMO TIPO DE SOLICITANTE DE EMPLEO COMO EL CANDIDATO PERFECTO

CON EL TIEMPO LLEGARÁ A CREER QUE ESA ES LA IMAGEN DE UN EMPLEADO PROMETEDOR.

TAMBIÉN ESTAMOS VIENDO UN SESGO PREEXISTENTE EN ESTE EJEMPLO: LA HERRAMIENTA DE IA SE ENTRENÓ CON DATOS HISTÓRICOS SOBRE EMPLEADOS ANTERIORES, QUE ERAN PREDOMINANTEMENTE HOMBRES.

AQUÍ HAY OTRO EJEMPLO, MÁS ADELANTE AÚN EN EL PROCESO DE CONTRATACIÓN, TAL VEZ DURANTE UNA VERIFICACIÓN DE ANTECEDENTES POSTERIOR A LA ENTREVISTA POR PARTE DE UN EMPLEADOR POTENCIAL:

LATANYA SWEENEY, PROFESORA DE INFORMÁTICA DE LA UNIVERSIDAD DE HARVARD,

DEMOSTRÓ QUE BUSCAR EN GOOGLE NOMBRES QUE SUENEN AFROAMERICANOS ES MÁS PROBABLE QUE GENERE ANUNCIOS QUE SUGIERAN ANTECEDENTES PENALES QUE BUSCAR EN GOOGLE NOMBRES QUE SUENEN BLANCOS,

E INCLUSO VERIFIQUE SI UNA PERSONA TIENE ANTECEDENTES PENALES.

ESTE ES UN SESGO PREEEXISTENTE EN JUEGO:

MANIFESTACIÓN DE PREJUICIOS RACIALES DE LARGA DATA DE LA SOCIEDAD.



[4] Racism is Poisoning Online Ad Delivery, says Harvard Professor. MIT Technology Review (2013)

LOS CASOS PRESENTADOS AQUÍ TIENEN UN DENOMINADOR COMÚN: MUESTRAN QUE LA IA PUEDE REFORZAR Y EXACERBAR LA DISCRIMINACIÓN ILEGAL CONTRA MINORÍAS Y GRUPOS HISTÓRICAMENTE DESFAVORECIDOS.

A MENUDO, ESTO SE DENOMINA "SESGO EN LA IA".

ENTONCES, ¿POR QUÉ LOS SISTEMAS SOFISTICADOS QUE TIENEN COMO OBJETIVO HACER QUE LA CONTRATACIÓN SEA MÁS EFICIENTE FALLAN EN ESO Y POSIBLEMENTE EMPEORAN LAS COSAS?

POR SUPUESTO, LOS PROBLEMAS DE SESGO EN EL EMPLEO NO SON NUEVOS. TAMBIÉN ESTABAN PRESENTES EN LA ERA ANALÓGICA.



POR EJEMPLO, EN SU CONOCIDO ESTUDIO DE 2004, MARIANNE BERTRAND Y SENDHIL MULLAINATHAN ENVIRIARON CURRÍCULUMS FICTICIOS A ANUNCIOS DE EMPLEO EN PERIÓDICOS DE BOSTON Y CHICAGO. [5]

PARA MANIPULAR LA RAZA PERCIBIDA, ASIGNARON AL AZAR NOMBRES QUE SONABAN AFROAMERICANOS O BLANCOS A LOS CURRÍCULUMS.



LOS NOMBRES BLANCOS RECIBIERON UN 50 POR CIENTO MÁS DE LLAMADAS PARA ENTREVISTAS.

ESTE CASO MUESTRA QUE EL SESGO PUEDE DEBERSE A DECISIONES HUMANAS.

REVISEMOS EL SESGO PREEEXISTENTE QUE A MENUDO SE MUESTRA EN LOS DATOS.

LOS DATOS SON UNA IMAGEN DEL MUNDO, SU REFLEJO EN EL ESPEJO.

CUANDO PENSAMOS EN EL SESGO EN LOS DATOS, CUESTIONAMOS ESTE REFLEJO.

UNA INTERPRETACIÓN DEL "SESGO EN LOS DATOS" ES QUE EL REFLEJO ESTÁ DISTORSIONADO:

PODEMOS SOBRERREPRESENTAR O SUBREPRESENTAR SISTEMÁATICAMENTE PARTES PARTICULARES DEL MUNDO EN LOS DATOS,

O DE LO CONTRARIO DISTORSIONAR LAS LECTURAS.

RECORDEMOS EL FRACASO DE LA IA DE RECLUTAMIENTO DE AMAZON PARA MEJORAR LA DIVERSIDAD DE LA FUERZA LABORAL.

ESTA HERRAMIENTA FUE ENTRENADA UTILIZANDO DATOS HISTÓRICOS: CURRÍCULUMS DE PERSONAS QUE FUERON CONTRATADAS EN EL PASADO.

ESE ENTRENAMIENTO ESTABA SUJETO A UN SESGO PREEEXISTENTE.

EN ESOS DATOS, HABÍA UNA SUBREPRESENTACIÓN DE MUJERES EN LA FUERZA LABORAL Y EN LOS ROLES TÉCNICOS.

UN PUNTO MÁS SUTIL TIENE QUE VER CON LAS DISTORSIONES.

CUANDO CONSIDERAMOS CARACTERÍSTICAS, COMO EL PUNTAJE DE UN INDIVIDUO EN UNA PRUEBA ESTANDARIZADA, ¿LAS TOMAMOS AL PIE DE LA LETRA?

¿LO TENEMOS EN CUENTA LAS DIFERENCIAS EN EL ACCESO A LAS OPORTUNIDADES EDUCATIVAS,

COMO HABER IDO A UNA MEJOR ESCUELA, O HABER TENIDO ACCESO A TUTORÍA PAGA?

OTRA INTERPRETACIÓN DEL "SESGO EN LOS DATOS" ES QUE INCLUSO SI PU-DIÉRAMOS REFLEJAR EL MUNDO PERFECTAMENTE EN LOS DATOS,

SEGUIRÍA SIENDO UN REFLEJO DEL MUNDO TAL COMO ES,



ES IMPORTANTE TENER EN CUENTA QUE UN REFLEJO NO PUEDE SABER SI ESTÁ DISTORSIONADO.



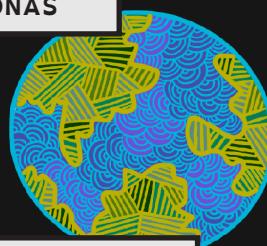
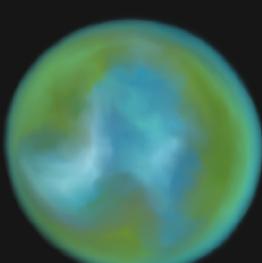
Y NO NECESARIAMENTE DE CÓMO PODRÍA O DEBERÍA SER.



LOS DATOS POR SÍ SOLOS NO PUEDEN DECIRNOS SI SON UN REFLEJO DISTORSIONADO DE UN MUNDO PERFECTO, UN REFLEJO PERFECTO DE UN MUNDO DISTORSIONADO,

O UNA COMBINACIÓN DE LAS DOS COSAS.

EL SEGUNDO PUNTO ES QUE NO DEPENDE DE LOS DATOS NI DE LOS ALGORITMOS, SINO DE LAS PERSONAS



—INDIVIDUOS, GRUPOS Y SOCIEDAD EN GENERAL—



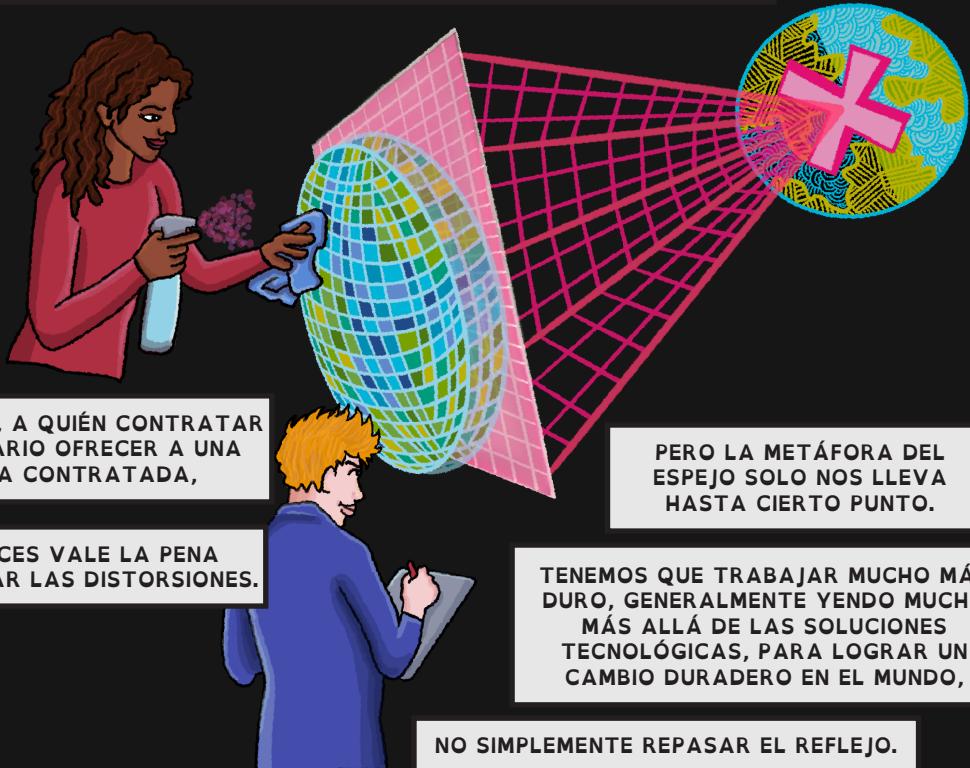
LLEGAR A UN CONSENSO SOBRE SI EL MUNDO ES COMO DEBE SER O SI NECESITA MEJORAR.

Y, SI ES ASÍ, CÓMO DEBERÍAMOS MEJORARLO.



EL PUNTO FINAL AQUÍ ES QUE CAMBIAR EL REFLEJO PUEDE NO CAMBIAR EL MUNDO.

SI EL REFLEJO MISMO SE USA PARA TOMAR DECISIONES IMPORTANTES:



POR EJEMPLO, A QUIÉN CONTRATAR
O QUÉ SALARIO OFRECER A UNA
PERSONA CONTRATADA,

PERO LA METÁFORA DEL
ESPEJO SOLO NOS LLEVA
HASTA CIERTO PUNTO.

ENTONCES VALE LA PENA
COMPENSAR LAS DISTORSIONES.

TENEMOS QUE TRABAJAR MUCHO MÁS
DURO, GENERALMENTE YENDO MUCHO
MÁS ALLÁ DE LAS SOLUCIONES
TECNOLÓGICAS, PARA LOGRAR UN
CAMBIO DURADERO EN EL MUNDO,

NO SIMPLEMENTE REPASAR EL REFLEJO.

Y VOLVIENDO AHORA AL DRAGÓN DE TRES CABEZAS DEL SESGO:

CUANDO HABLAMOS DE ABORDAR EL SESGO EN LA IA, TENDEMOS A ENMARCAR EL PROBLEMA COMO ENCONTRAR UNA MANERA DE MATAR AL DRAGÓN DEL SESGO.

PERO LO LARGO DE NUESTRO DEBATE SOBRE EL
VÍNCULO IRREVOCABLE ENTRE EL SESGO HUMANO Y
EL SESGO DE LA MÁQUINA,

NOS ENCONTRAMOS
CUESTIONANDO LA NATURALEZA
MISMA DE ESTA HISTORIA.

AL FINAL DEL DÍA,
TAL VEZ LA
PREGUNTA NO SEA:

¿CÓMO MATAR AL DRAGÓN Y
RESCATAR A LA PRINCESA?

LA PREGUNTA QUE
REALMENTE DEBERÍAMOS
HACERNOS ES:

¿QUÉ HACEMOS CON UNA SOCIEDAD QUE ENCIERRA
PRINCESAS EN CASTILLOS, EN PRIMER LUGAR?