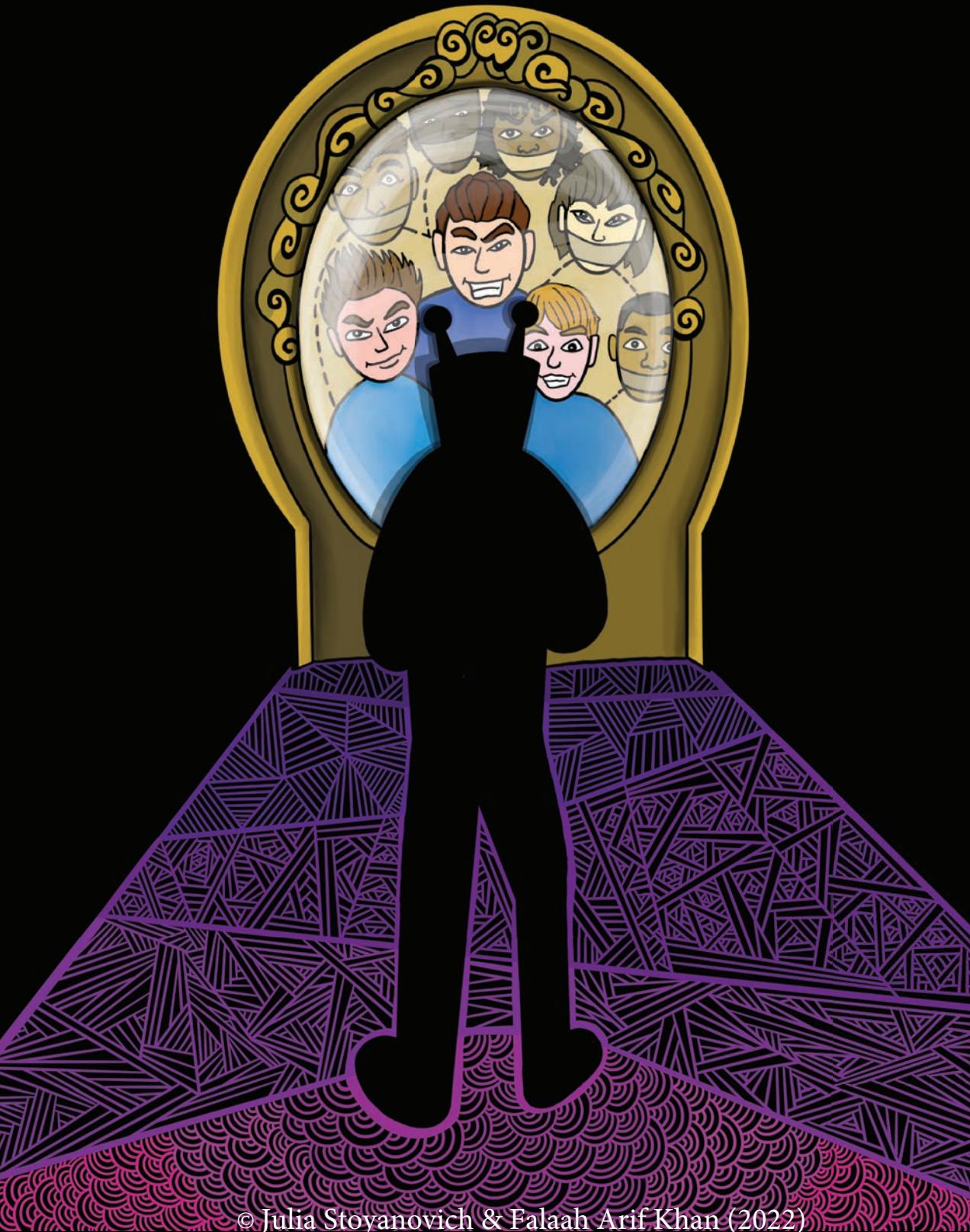


Somos IA n.º 4:

# TODO SOBRE LOS SESGOS



© Julia Stoyanovich & Falaah Arif Khan (2022)

Traducido por Daniel Domínguez Figaredo

# Términos de uso

Todos los contenidos gráficos/viñetas de este cómic están protegidos por una licencia CC BY-NC-ND 4.0. Consulte la página web de las licencias para obtener detalles sobre cómo puede usar este material gráfico.

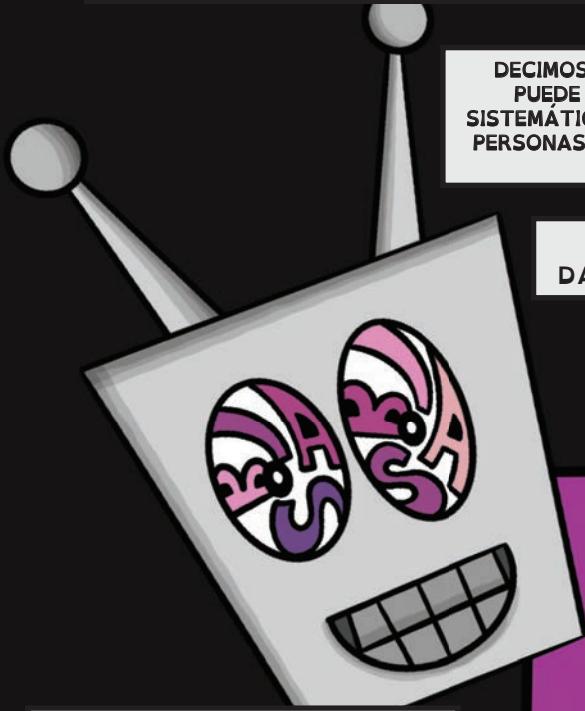
Se puede usar paneles/grupos de paneles en presentaciones/artículos, siempre y cuando:

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". We are AI Comics, Vol. 4 (2021) <http://r-ai.co/comics>

VAMOS A HABLAR DE LO QUÉ SON Y CÓMO SURGEN LOS LLAMADOS “SESGOS” DE LA IA.



DECIMOS QUE UNA IA ESTÁ SESGADA CUANDO SU USO PUEDE CONDUCIR A UNA DISCRIMINACIÓN INJUSTA Y SISTEMÁTICA CONTRA PERSONAS CONCRETAS O GRUPOS DE PERSONAS, DE MANERA QUE FAVOREZCA OTRAS PERSONAS O GRUPOS.

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

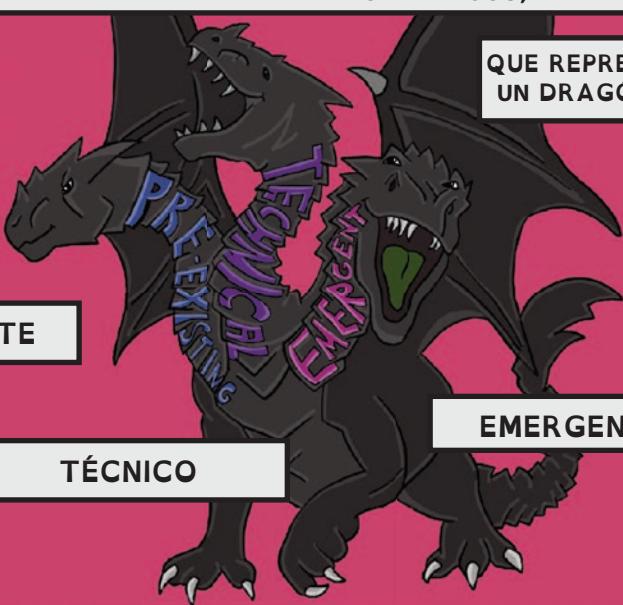
O DE CÓMO ESTÁ DISEÑADO EL ALGORITMO,

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 REPRESENTAMOS AQUÍ COMO UN DRAGÓN CON TRES CABEZAS:

PREEEXISTENTE

TÉCNICO

EMERGENTE

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

¡USEMOS AHORA LA MISMA METÁFORA PARA ENTENDER LOS SESGOS!



LOS SESGOS PREEXISTENTES EXISTEN INDEPENDIENTEMENTE DE LOS ALGORITMOS Y TIENEN SU ORIGEN EN LA SOCIEDAD.

SESGOS PREEXISTENTES  
(EN LOS DATOS)

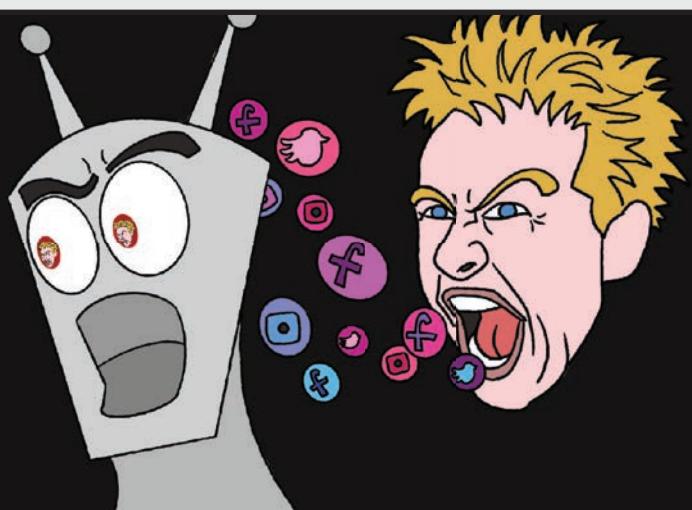
ESTOS SERÍAN LOS TIPOS DE SABORES QUE SE FILTRARÁN A NUESTRO PAN SI NO PRIORIZAMOS LA PUREZA Y LA FRESCURA DE LOS INGREDIENTES,

O SI DECIDIMOS USAR UNA MASA MADRE PREMEZCLADA Y LISTA PARA USAR.

ESOS SESGOS EXISTEN EN LA SOCIEDAD, VIENEN "PREOCINADOS" EN EL ALGORITMO,

Y PROCEDEN DEL SISTEMA DE DISCRIMINACIÓN SUBYACENTE EXISTENTE EN EL LUGAR DONDE SE RECOPIRARON LOS DATOS:

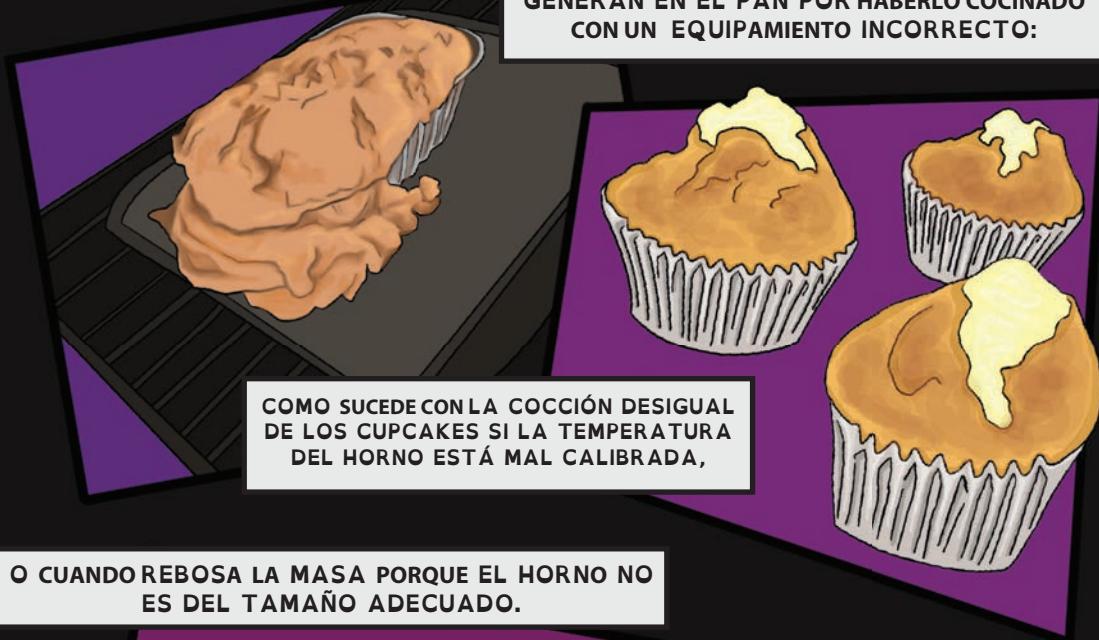
COMO POR EJEMPLO, LOS ESTEREOTIPOS RACIALES Y DE GÉNERO QUE LOS MODELOS LINGÜÍSTICOS RECOPEN CUANDO SE ENTRENAN CON DATOS PROCEDENTES DE LAS REDES SOCIALES.



## SESGOS TÉCNICOS (EN EL SISTEMA TÉCNICO)

LOS SESGOS TÉCNICOS LOS INTRODUCE EL PROPIO SISTEMA, DEBIDO A LA FORMA EN QUE ESTÁ DISEÑADO O A CÓMO FUNCIONA.

EQUIVALEN A LAS IMPERFECCIONES QUE SE GENERAN EN EL PAN POR HABERLO COCINADO CON UN EQUIPAMIENTO INCORRECTO:



Y VOLVIENDO A LOS SISTEMAS INFORMÁTICOS:

UN EJEMPLO DESTACADO SON LAS PLATAFORMAS DE REDES SOCIALES



QUE ACABAN PROMOCIONANDO LAS NOTICIAS FALSAS Y ARTÍCULOS EXTREMISTAS.

## SESgos Emergentes (Ocasionados por Decisiones)

LOS SESGOS EMERGENTES SURGEN CON EL PASO DEL TIEMPO, PORQUE LAS DECISIONES QUE SE TOMAN CON LA AYUDA DE LOS SISTEMAS INFORMÁTICOS LLEVAN A CAMBIAR COSAS DEL MUNDO REAL,

LO QUE A SU VEZ AFECTA EL FUNCIONAMIENTO DE ESOS MISMOS SISTEMAS EN EL FUTURO.

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

¿QUÉ PASARÍA SI NOS CONVERTIMOS EN MAESTROS REPOSTEROS E INCORPORAMOS, SIN DARNOS CUENTA, EL PAN COMO UN ALIMENTO QUE PERMANECE CONSTANTE EN NUESTRA DIETA?



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

O PENSEMOS SI NUESTRA OPINIÓN SOBRE "EL SABOR QUE DEBE TENER EL PAN" ESTÁ CONDICIONADA POR LA POPULARIDAD DE OTROS PRODUCTOS PARECIDOS Y CON SABORES CARACTERÍSTICOS COMO EL "WONDER BREAD".



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

ESTÁ CONFORMADA POR LOS ALGORITMOS QUE SELECCIONAN NOTICIAS DE LAS REDES SOCIALES A PARTIR DE LAS PUBLICACIONES QUE SON POPULARES Y "DE TENDENCIA".



PARA CONCRETAR UN POCO MÁS, VEAMOS EJEMPLOS DE LOS SESGOS ALGORÍTMICOS EN EL MUNDO REAL.

TOMEMOS EL SECTOR DE LA “CONTRATACIÓN” COMO UN CASO REPRESENTATIVO, DONDE 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, TRAS PUBLICARSE LOS RESULTADOS DEL ESTUDIO ADFISHER DE LA UNIVERSIDAD CARNEGIE MELLON [2].

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

—EN TÉRMINOS DE DATOS DEMOGRÁFICOS, INTERESES 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.

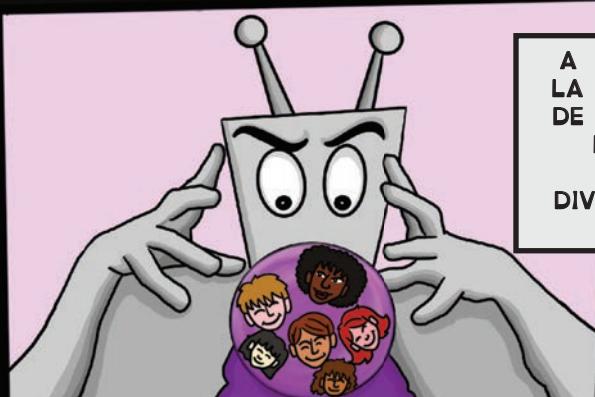
ESTO NOS RETROTRAE A LA ÉPOCA EN QUE ERA LÉGAL ANUNCIAR EN LOS PERIÓDICOS PUESTOS DE TRABAJO DIFERENCIADOS PARA CADA GÉNERO. ESA PRÁCTICA FUE PROHIBIDA EN LOS EE. UU. EN 1964, PERO PERSISTE EN EL ENTORNO DE LA PUBLICIDAD EN INTERNET.

MÁS TARDE SE DEMOSTRÓ QUE PARTE DE LA RAZÓN POR LA QUE ESO ESTABA SUCEDIENDO ES LA DINÁMICA DEL PROPIO SISTEMA DE ETIQUETADO DE LOS ANUNCIOS, UN MECANISMO ESENCIAL PARA LANZAR LAS OFERTAS DE EMPLEO.

ESTO ES UN SESGO TÉCNICO EN LA PRÁCTICA.

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

VAYAMOS ADELANTE EN EL TIEMPO Y AVANCemos TAMBIÉN A LA SIGUIENTE ETAPA DEL PROCESO DE CONTRATACION: LA REVISIÓN DE CURRÍCULUM.



A FINES DE 2018, SE CONOCÍÓ QUE LA HERRAMIENTA DE CONTRATACION DE AMAZON BASADA EN IA, QUE SE HABÍA DESARROLLADO CON EL OBJETIVO DE AUMENTAR LA DIVERSIDAD DE LOS TRABAJADORES, HIZO JUSTO LO CONTRARIO:

EL SISTEMA APRENDIÓ POR 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 SOLO DE MUJERES.

LOS RESULTADOS SE ALINEARON Y REFORZARON UN MARCADO DESEQUILIBRIO DE GÉNERO EN LOS TRABAJADORES DE LA EMPRESA.

ESTE ES UN SESGO EMERGENTE EN LA PRÁCTICA:

UN GERENTE DE CONTRATACION A QUIEN UNA HERRAMIENTA DE IA LE 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 ADECUADO.

TAMBÍEN SE PUEDE APRECiar UN SESGO PREEXISTENTE EN ESTE MISMO EJEMPLO: LA HERRAMIENTA DE IA SE ENTRENó CON DATOS HISTóRICOS SOBRE EMPLEADOS ANTERIORES, QUE ERAN PREDOMINANTEMENTE HOMBRES.

AQUÍ HAY OTRO EJEMPLO, QUÉ SUCEDA EN UNA FASE AÚN MÁS ADELANTADA DEL PROCESO DE CONTRATACION, DURANTE LA VERIFICACION DE ANTECEDENTES QUE SE HACE TRAS LA ENTREVISTA CON UN POTENCIAL EMPLEADOR:

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

DEMOSTRÓ QUE BUSCAR EN GOOGLE NOMBRES QUE SUENAN A PERSONAS AFROAMERICANAS ES MÁS PROBABLE QUE GENERE ANUNCIOS QUE SUGIEREN ANTECEDENTES PENALES, QUE BUSCAR EN GOOGLE NOMBRES QUE SUENAN A PERSONAS BLANCAS,

INCLUSO VERIFICANDO EN LA BÚSQUEDA SI UNA PERSONA TIENE O NO ANTECEDENTES PENALES.



ASÍ ES COMO FUNCIONA UN SESGO PREEXISTENTE EN LA PRÁCTICA:



MANIFESTANDO LOS PREJUICIOS RACIALES QUE ESTÁN AMPLIAMENTE ARRAIGADOS EN LA SOCIEDAD.



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

GENERALMENTE, ESTO SE DENOMINA "SESGOS DE LA IA".

¿POR QUÉ FAJAN ESTOS SISTEMAS SOFISTICADOS QUE PRETENDEN QUE LA CONTRATACIÓN SEA MÁS EFICIENTE, HACIENDO QUE LAS COSAS EMPEOREN TODAVÍA MÁS?

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 ENVIAJON CURRÍCULUMS FICTICIOS A OFERTAS DE EMPLEO EN PERIODICOS DE BOSTON Y CHICAGO. [5]

PARA MANIPULAR LA PERCEPCIÓN DE LA RAZA, 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 DE LOS DATOS, ESTAMOS CUESTIONANDO ESE REFLEJO.

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

PODEMOS SOBRERREPRESENTAR O SUBREPRESENTAR  
SISTEMÁTICAMENTE LAS PARTES PARTICULARES DEL  
MUNDO QUE MUESTRAN LOS DATOS,

O TAMBIÉN DISTORSIONAR LAS  
LECTURAS DE ESOS DATOS.

RECORDEMOS EL FRACASO DE LA IA DE CONTRACTACIÓN DE AMAZON  
PARA MEJORAR LA DIVERSIDAD DE LA PLANTILLA DE EMPLEADOS.

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 PLANTILLA EXISTENTE Y EN  
LOS ROLES TÉCNICOS.

UN PUNTO MÁS SUTIL TIENE QUE VER CON LAS DISTORSIONES.

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

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

COMO HABER IDO A UNA MEJOR ESCUELA, O  
HABER TENIDO ACCESO A TUTORÍAS PERSONALES?

OTRA INTERPRETACIÓN DEL "SESGO DE LOS DATOS" ES QUE INCLUSO SI PUDIERAMOS REFLEJAR EL MUNDO PERFECTAMENTE EN LOS DATOS,

SEGUIRÍA SIENDO UN REFLEJO DEL MUNDO TAL COMO ES,



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



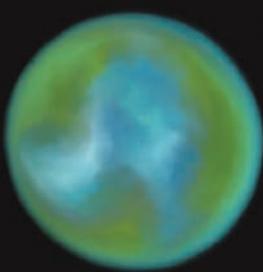
Y NO NECESARIAMENTE DE CÓMO PODRÍA O DEBERÍA SER.



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

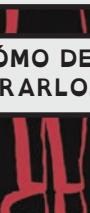


—INDIVIDUOS, GRUPOS Y SOCIEDAD EN GENERAL—



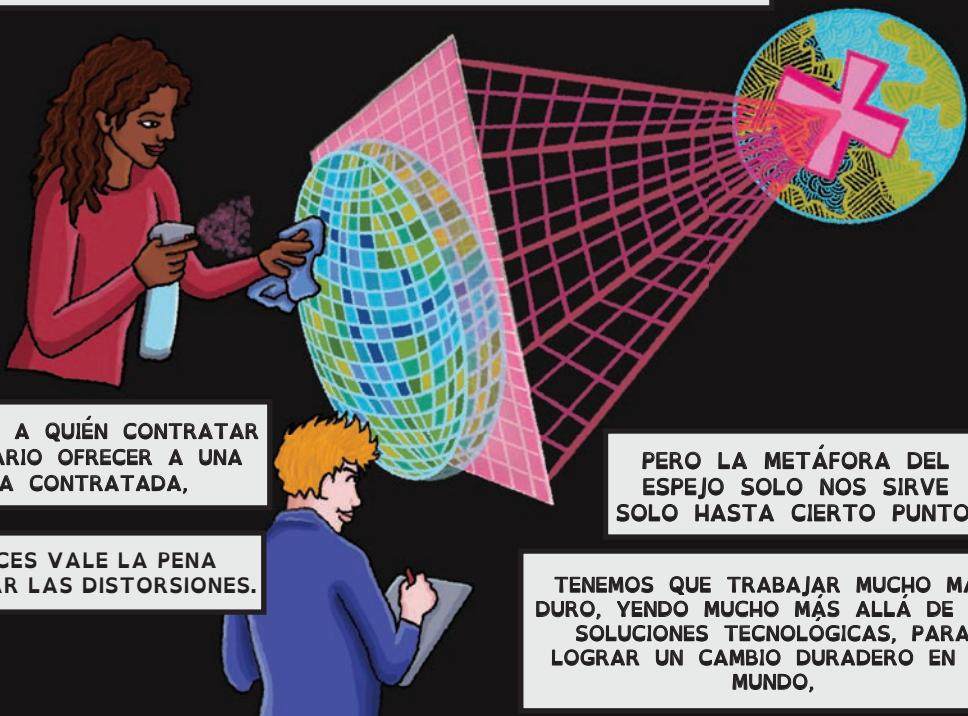
LLEGAR A UN CONSENSO SOBRE SI EL MUNDO ES COMO DEBE SER O SI NECESITA MEJORAR.

Y, SI ES ASÍ, CÓMO DEBERÍAMOS MEJORARLO.



LA CONCLUSIÓN AQUÍ ES QUE, POSIBLEMENTE, CAMBIAR EL REFLEJO DEL MUNDO NO EQUIVALE A CAMBIAR EL MUNDO.

SI EL REFLEJO 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 SIRVE  
SOLO HASTA CIERTO PUNTO.

ENTONCES VALE LA PENA  
COMPENSAR LAS DISTORSIONES.

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

NO ES SUFFICIENTE CON REVISAR EL REFLEJO, SIMPLEMENTE.

Y VOLVIENDO AHORA AL DRAGÓN DE TRES  
CABEZAS DE LOS SESGOS:

CUANDO HABLAMOS DE ABORDAR LOS SESGOS EN LA IA, TENDREMOS A  
ENMARCAR EL PROBLEMA EN COMO ENCONTRAMOS LA MANERA DE MATAR AL  
DRAGON.

PERO EN NUESTRO DEBATE SOBRE EL VÍNCULO  
ENTRE LOS SESGOS HUMANOS Y LOS SESGOS DE LAS  
MÁQUINAS,

NOS HEMOS LLEGADOS A  
CUESTIONAR LA NATURALEZA  
MISMA DE ESA RELACIÓN.

ASÍ QUE, EN ÚLTIMA  
INSTANCIA, TAL VEZ  
LA PREGUNTA NO SEA:

¿CÓMO MATAR AL DRAGÓN Y  
RESCATAR A LA PRINCESA?

LA PRIMERA PREGUNTA QUE  
REALMENTE DEBERÍAMOS  
HACERNOS ES:

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