SISTEMAS DE DECISIÓN AUTOMATIZADA JUSTOS

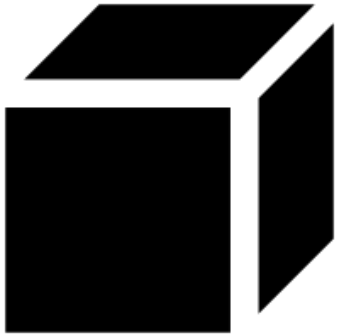
Patricia Díaz | @PatoDiazGNU

3 GRANDES DESAFÍOS

PRIVACIDAD

DIVERSIDAD

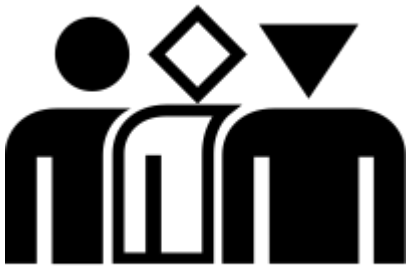
TRANSPARENCIA



3 GRANDES DESAFÍOS

PRIVACIDAD

DIVERSIDAD



TRANSPARENCIA

EL SESGO EN LOS ALGORITMOS AFECTA LA VIDA DE LAS PERSONAS

Resultados positivos	Resultados negativos
empleo ofrecido	empleo negado
préstamo aceptado	préstamo denegado
descuento ofrecido	descuento no ofrecido

Discriminación en los pactos Internacionales

DISCRIMINACIÓN DIRECTA (trato desigual) por motivos de Raza (*)
Convención Internacional sobre la Eliminación de todas las Formas de Discriminación Racial la define como:
"*toda distinción, exclusión, restricción o preferencia basada en motivos de raza, color, linaje u origen nacional o étnico que tenga por objeto o por resultado anular o menoscabar el reconocimiento, goce o ejercicio, en condiciones de igualdad, de los derechos humanos y libertades fundamentales.*"

(*) En la Convención sobre la eliminación de todas las formas de discriminación contra la mujer y la Convención sobre los derechos de las personas con discapacidad figuran definiciones similares.

DISCRIMINACIÓN INDIRECTA (impacto desigual):
Comité de Derechos Económicos, Sociales y Culturales define la discriminación indirecta como "*leyes, políticas o prácticas en apariencia neutras pero que influyen de manera desproporcionada en los derechos del Pacto afectados por los motivos prohibidos de discriminación*"

SESGO RACIAL

(Redlining)

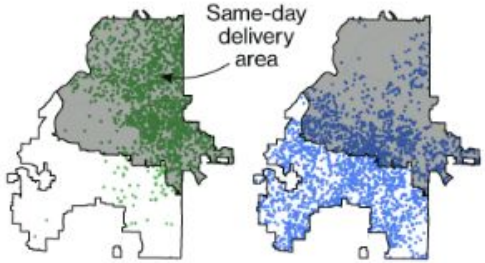
Bloomberg

Amazon Doesn't Consider the Race of Its Customers. Should It?

By David Ingold and Spencer Soper
April 21, 2016

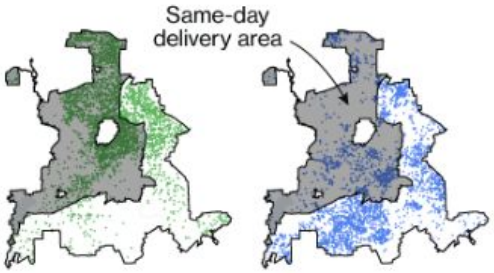
The northern half of Atlanta, home to 96% of the city's white residents, has same-day delivery. The southern half, where 90% of the residents are black, is excluded.

White residents | Black residents



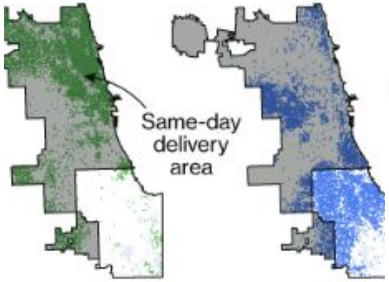
Dallas has the lowest overall coverage rate among cities with same-day delivery. White residents are more than twice as likely as black residents to have access to the service.

White residents | Black residents



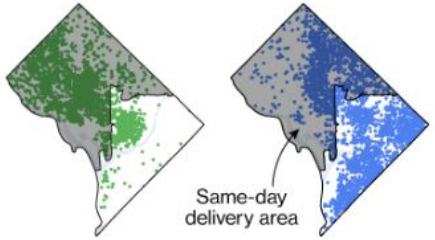
ɔut half of Chicago's black residents in the southern half of the city where ey do not have access to Amazon's same-day delivery service.

e residents | Black residents



Same-day service is unavailable in southeast Washington, D.C. including neighborhoods located blocks from the Capitol building and all areas across the Anacostia River.

White residents | Black residents



SESGO RACIAL y de GÉNERO

(en sistemas de Reconocimiento Facial Automatizado)

PROYECTO GENDER SHADES

Investigador del MIT que expone el sesgo en la tecnología de reconocimiento facial desencadena la ira de Amazon

Por Matt O'Brien | 8 de abril de 2019

SESGO RACIAL y de GÉNERO

(en sistemas de Reconocimiento Facial Automatizado)

PROYECTO GENDER SHADES

SESGO GÉNERO y EDAD
(en anuncios laborales)

Buscar empleos | Registrarse | Buscar | Edición internacional

The Guardian
For 200 years

England exams row timeline: was Ofqual warned of algorithm bias?

Scrapping followed series of warnings over algorithm's volatility and fairness

Ben Quinn *and* Richard Adams

Thu 20 Aug 2020 19.53 BST

▲ Students opposite Downing Street protesting against the downgrading of A-level results on 16 August. Photograph: Matthew Chattle/Rex/Shutterstock

CALIFICACIÓN A ESTUDIANTES DURANTE LA PANDEMIA EN UK

Efecto Matthew

"The rich get richer and the poor get poorer"

El sistema discrimina a los estudiantes que históricamente tuvieron malas calificaciones

(Informe Automating Society, 2020)

EL SESGO NO SIEMPRE ESTÁ EN LOS ALGORITMOS

Noticias hoy Dólar blue hoy Elecciones 2023 Ganancias Macri Encuestas Dónde voto Mendoza Champions League Sigue nuestro canal en Whatsapp Horóscopo hoy Virales Martes, 19 de Septiembre de 2023

Clarín

Suscribite por $30 Ingresar

En vivo | Sesión por Ganancias en Diputados | Elecciones 2023: 'Lo de Chaco fue apoteótico y el domingo ganamos en Mendoza', sentenció Bullrich

Sociedad

En Salta usan inteligencia artificial para prever embarazos adolescentes

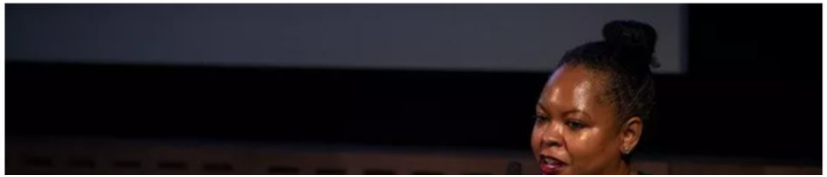
El gobernador Juan Manuel Urtubey dijo que hay niñas "predestinadas a un embarazo adolescente". Y que las pueden identificar con nombre, apellido y domicilio para trabajar con ellas y prevenirlo.

lm Internacional Cultura Política Clima Sociedad Kiosco Especiales Opinión Más

Sociedad

Simone Browne: «La vigilancia biométrica es similar a la del tráfico de esclavos»

- Entrevista a la socióloga Simone Browne, autora de 'Dark Matters: On the Surveillance of Blackness'



Los sistemas de **Procesamiento de Lenguaje Natural** también reproducen los sesgos sociales

EDIA:
Estereotipos y Discriminación en Inteligencia Artificial

Metodologías para explorar el sesgo en MODELOS DE LENGUAJE NATURAL EN ESPAÑOL

[Herramientas desarrolladas por Fundación Vía Libre y por la Universidad Nacional de Córdoba]

EDIA:
Estereotipos y Discriminación en Inteligencia Artificial
Metodologías para explorar el sesgo en MODELOS DE LENGUAJE NATURAL EN ESPAÑOL

EJEMPLOS de herramientas que utilizan PLN:

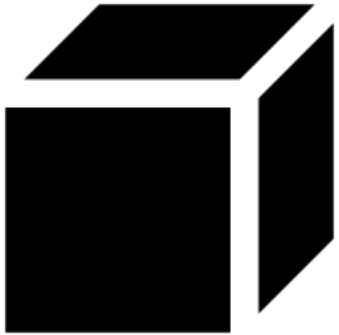
- Google Traductor
- ChatGPT
- Asistente de voz Siri, Alexa
- Chatbots
- Motores de búsqueda (búsqueda predictiva)

3 GRANDES DESAFÍOS

PRIVACIDAD

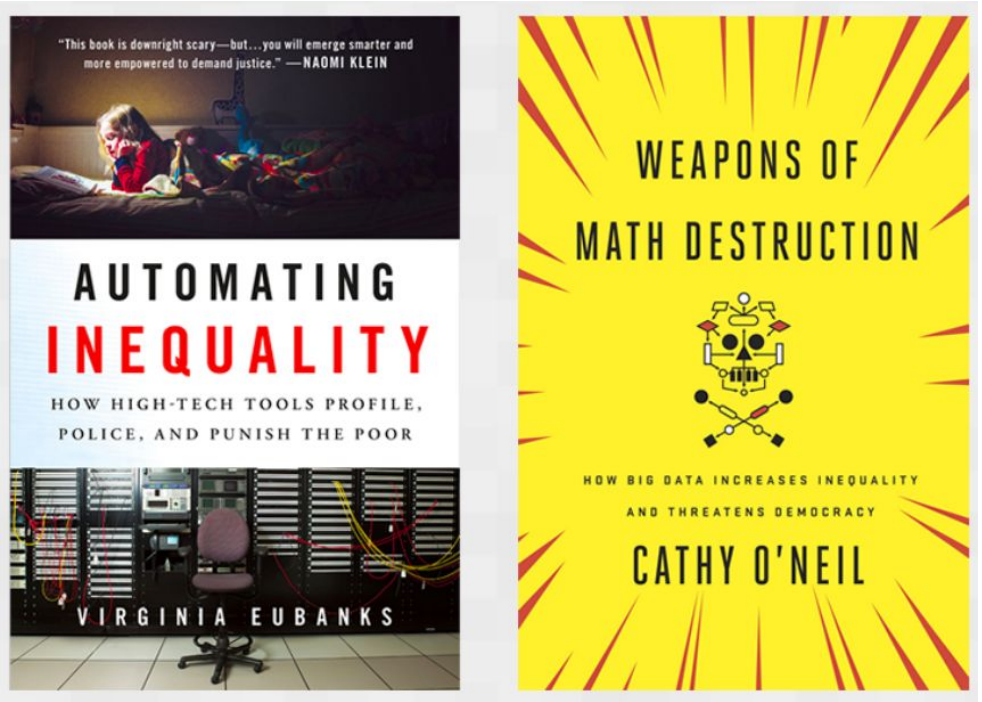
DIVERSIDAD

TRANSPARENCIA



“LOS ALGORITMOS SON OPINIONES EMBEBIDAS EN CÓDIGO”

(Cathy O’Neil, 2016)



RANKINGS DOCENTES EN LAS ESCUELAS DE USA

The Washington Post
Democracy Dies in Darkness

Education

‘Creative ... motivating’ and fired

Sarah Wysocki was out of work for only a few days after she was fired by DCPS last year. She is now teaching at Hybla Valley Elementary School in Fairfax County. (Jahi Chikwendiu/The Washington Post)

By Bill Turque
March 6, 2012

RANKINGS DOCENTES EN LAS ESCUELAS DE USA

¿Cómo detectar y corregir los sesgos si no comprendemos cómo se toman las decisiones automatizadas?

The Washington Post
Democracy Dies in Darkness

Education

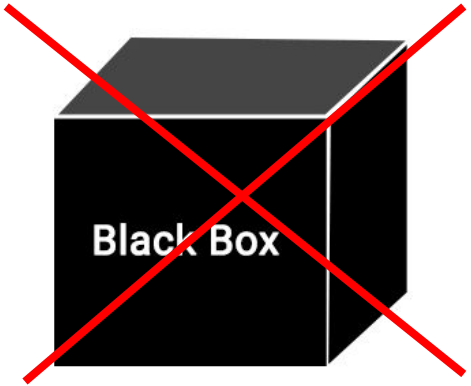
‘Creative ... motivating’ and fired

Sarah Wysocki was out of work for only a few days after she was fired by DCPS last year. She is now teaching at Hybla Valley Elementary School in Fairfax County. (Jahi Chikwendiu/The Washington Post)

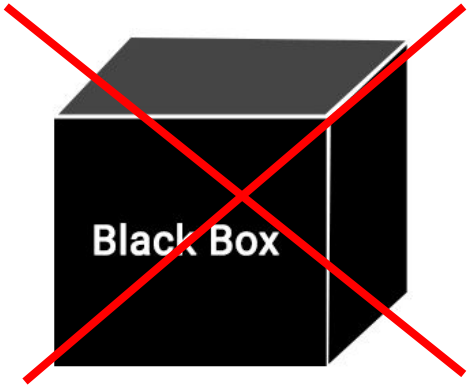
By Bill Turque
March 6, 2012

¿QUÉ ES LA TRANSPARENCIA ALGORÍTMICA?

- *transparencia algorítmica* ***no es sinónimo de liberar el código fuente***

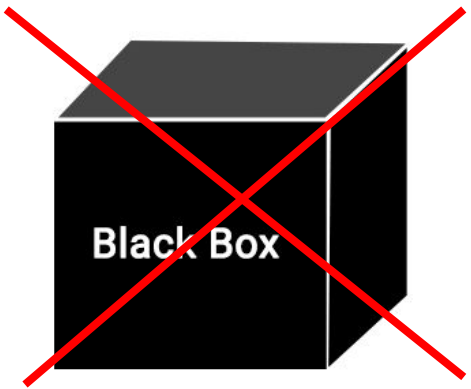


¿QUÉ ES LA TRANSPARENCIA ALGORÍTMICA?



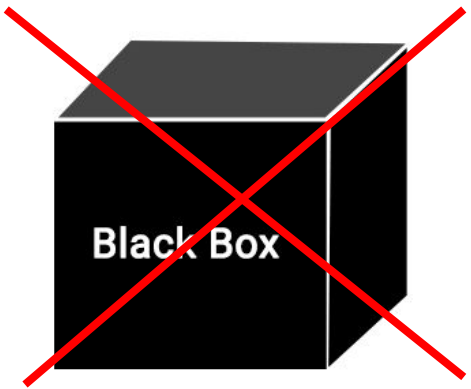
- *transparencia algorítmica* ***no es sinónimo de liberar el código fuente***
- *la transparencia algorítmica requiere transparencia de datos*

¿QUÉ ES LA TRANSPARENCIA ALGORÍTMICA?



- *transparencia algorítmica* ***no es sinónimo de liberar el código fuente***
- *la transparencia algorítmica requiere transparencia de datos*
- *la transparencia de datos* ***no es sinónimo de hacer públicos todos los datos***

¿QUÉ ES LA TRANSPARENCIA ALGORÍTMICA?



- *transparencia algorítmica* ***no es sinónimo de liberar el código fuente***
- *la transparencia algorítmica requiere transparencia de datos*
- *la transparencia de datos* ***no es sinónimo de hacer públicos todos los datos***
- *la transparencia procesable requiere* ***interpretabilidad / explicabilidad***

explicar los supuestos del sistema y sus efectos, involucrar a los actores que recibirán el impacto en la implementación

¿EXISTE EL DERECHO A LA EXPLICABILIDAD?
No exactamente...

URUGUAY Ley 18331
Artículo 16 (Derecho a la impugnación de valoraciones personales)

"Las personas tienen derecho a no verse sometidas a una **decisión con efectos jurídicos** que les afecte de manera significativa, **que se base en un tratamiento automatizado de datos** destinado a evaluar determinados aspectos de su personalidad, como su rendimiento laboral, crédito, fiabilidad, conducta, entre otros.

El afectado podrá **impugnar los actos** administrativos o decisiones privadas que impliquen una valoración de su comportamiento, **cuyo único fundamento sea un tratamiento de datos personales que ofrezca una definición de sus características o personalidad.**

En este caso, el afectado tendrá **derecho a obtener información del responsable de la base de datos tanto sobre los criterios de valoración como sobre el programa utilizado** en el tratamiento que sirvió para adoptar la decisión manifestada en el acto."

¿EXISTE EL DERECHO A LA EXPLICABILIDAD?
No exactamente...

URUGUAY Ley 18331
Artículo 16 (Derecho a la impugnación de valoraciones personales)

"Las personas tienen derecho a no verse sometidas a una **decisión con efectos jurídicos** que les afecte de manera significativa, **que se base en un tratamiento automatizado de datos** destinado a evaluar determinados aspectos de su personalidad, como su rendimiento laboral, crédito, fiabilidad, conducta, entre otros.

El afectado podrá **impugnar los actos** administrativos o decisiones privadas que impliquen una valoración de su comportamiento, **cuyo único fundamento sea un tratamiento de datos personales que ofrezca una definición de sus características o personalidad.**

En este caso, el afectado tendrá **derecho a obtener información del responsable de la base de datos tanto sobre los criterios de valoración como sobre el programa utilizado** en el tratamiento que sirvió para adoptar la decisión manifestada en el acto."

PERO...

- **no se imponen obligaciones de transparencia proactiva.**
- **el mecanismo solo aplica sobre sistemas que tratan datos personales y cuando la decisión automatizada tenga "efectos jurídicos" que afecten a un individuo de forma significativa.**

“Una evaluación y comunicación adecuadas de la funcionalidad debería ser un requisito mínimo para el despliegue masivo de sistemas algorítmicos”

ACM DIGITAL LIBRARY

☰ Navegación de artículos

La falacia de la funcionalidad de la IA

Inioluwa Deborah Raji , Universidad de California, Berkeley, EE. UU., deborahraji1@gmail.com
I. Elizabeth Kumar , Universidad de Brown, EE. UU., iekumar@brown.edu
Aaron Horowitz , Unión Americana de Libertades Civiles, EE. UU., ahorowitz@aclu.org
Andrew Selbst , Universidad de California, Los Ángeles, EE. UU., aselbst@law.ucla.edu

DOI: https://doi.org/10.1145/3531146.3533158
HECHO '22: Conferencia ACM 2022 sobre equidad, responsabilidad y transparencia , **Seúl, República de Corea, junio de 2022**

Los sistemas de IA desplegados a menudo no funcionan. Pueden construirse al azar, implementarse indiscriminadamente y promoverse de manera engañosa. Sin embargo, a pesar de esta realidad, los académicos, la prensa y los legisladores prestan muy poca atención a la funcionalidad. Esto conduce a soluciones técnicas y políticas centradas en implementaciones "éticas" o alineadas con el valor, a menudo omitiendo la pregunta anterior de si un sistema determinado funciona o proporciona algún beneficio. Para describir los daños de varios tipos de fallas de funcionalidad, analizamos un conjunto de estudios de casos para crear una taxonomía de problemas de funcionalidad de IA conocidos. Luego señalamos las políticas y las respuestas organizacionales que a menudo se pasan por alto y se vuelven más disponibles una