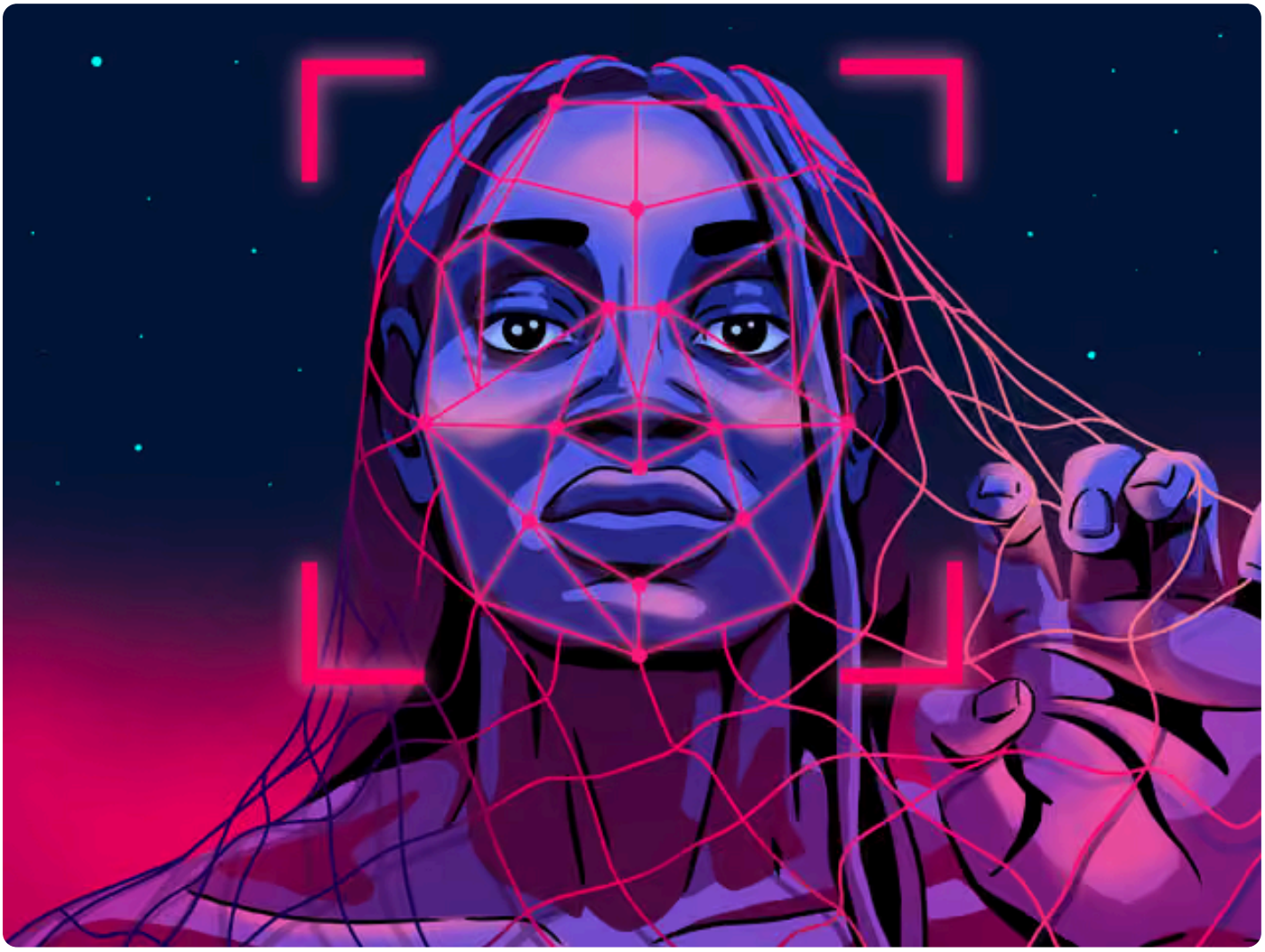


06/11/2020

Esto no es un pato

TXT CONSUELO LÓPEZ IMG SOFI SALAZAR

¿Estamos programando computadoras con prejuicios? ¿Cómo hace una inteligencia artificial para distinguir un pato de una minoría?

El futuro está acá: los algoritmos están adquiriendo habilidades que creíamos tan humanas como jugar al ajedrez o manejar un auto. ¿Es cuestión de tiempo hasta que se rebelen contra nosotros?

Probablemente no. Pero sí es cierto que **los avances tecnológicos y técnicos van mucho más rápido que los avances y las discusiones éticas sobre su uso**. Tenemos miedo cuando leemos que las máquinas aprenden más rápido que los humanos a pensar y resolver problemas, pero no les tememos a la discriminacion y a los sesgos que se perpetúan en esas inteligencias. Porque a la inteligencia artificial la crean humanos y **adopta tanto sus virtudes como sus defectos**. Así que estamos en un momento en el que hay que frenar un toque, mirar alrededor para entender dónde estamos parados y evaluar otros aspectos antes de seguir hacia el futuro. O, al menos, hacia 'ese' futuro.

El club de los niños blancos

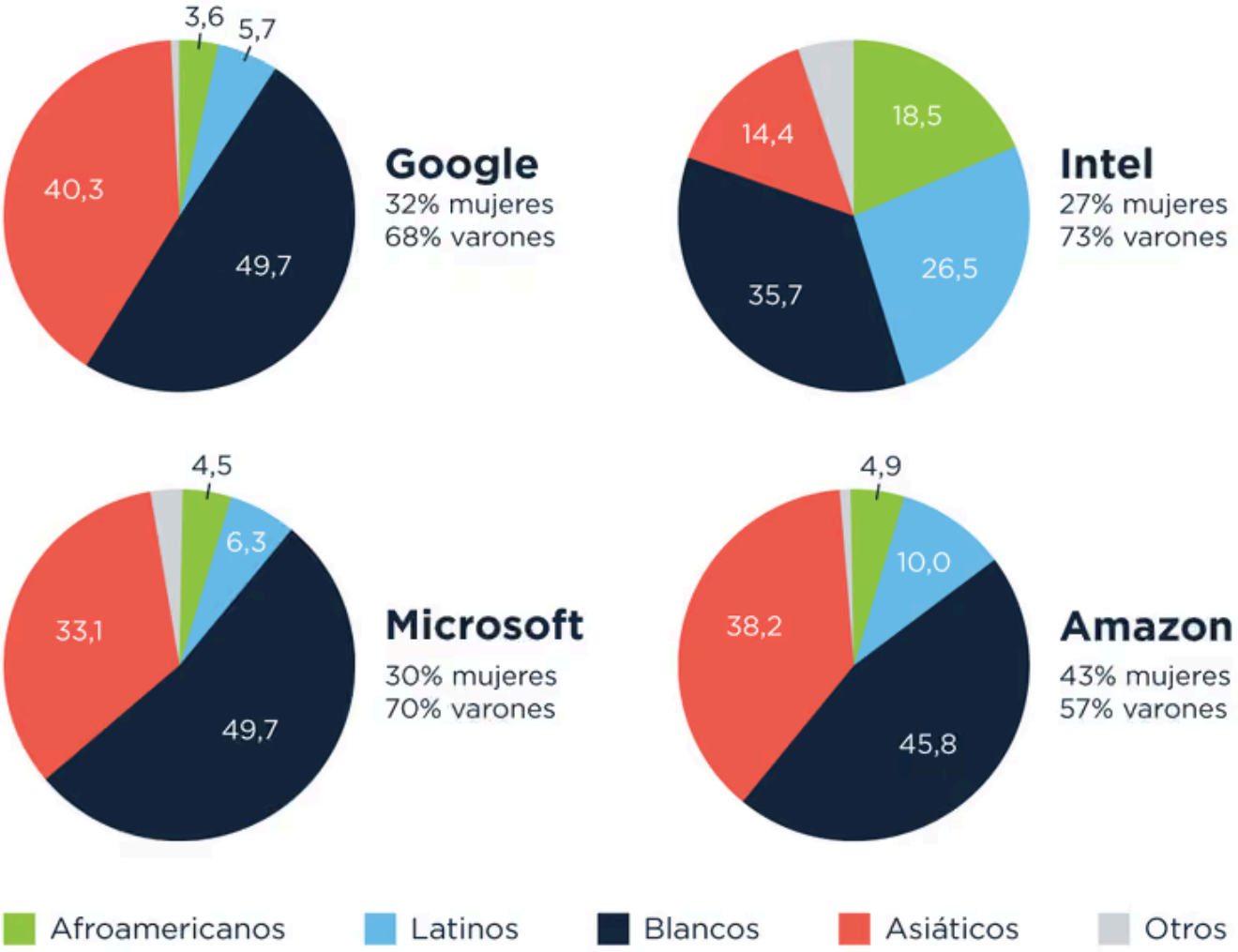
Hace unos meses nos enteramos del asesinato de George Floyd por parte de la policía estadounidense, cuando fue detenido por pagar con un billete falso de 20 dólares. El hecho fue grabado y registrado por una mujer de 17 años que pasaba por ahí. Floyd era una persona de la comunidad afroamericana y, con este suceso, Estados Unidos (y en consecuencia el mundo) tuvo un shock de realidad sobre el racismo.

El hecho fue la punta del iceberg de un problema mucho mayor, que es el racismo sistémico que se vive, principalmente, en la sociedad norteamericana. El acontecimiento repercutió en muchos ámbitos y vimos a miles de personas manifestarse en las calles, en medio de una pandemia, bajo la consigna *#BlackLivesMatter* (las vidas negras importan).

La cuestión es que Silicon Valley no fue la excepción a este cimbronazo; en las últimas semanas fuimos testigos de cambios históricos en la industria del *software* y de la tecnología. ¿Cuáles fueron estos cambios y por qué nos llaman tanto la atención?

Sillicon Valley es la mayor sede de compañías emergentes y globales de tecnología del mundo, es decir, **donde se cocina la mayoría de los cambios y avances tecnológicos del planeta**. El algoritmo que recomienda canciones que hacen acordar ex parejas, el que muestra la publicidad de eso que no se necesita pero que es tan lindo y el que hace el *match* con quien vamos a chatear solamente media hora: todos esos seguramente fueron creados por gente que vive y trabaja en ese área de la Bahía de San Francisco, en EE.UU.

El otro dato importante sobre este lugar es que Sillicon Valley ha sido, por mucho tiempo, el 'club de los niños blancos'. Si miramos los números de diversidad de los grandes monstruos tecnológicos, vemos que el porcentaje de empleados latinos, por ejemplo, tiene una sola cifra, y estos números empeoran cada vez más cuando se sube en la pirámide organizacional. Lo mismo pasa con otros grupos:



Fuente: Google, Microsoft, Amazon, Intel Nota: Entendemos que el género de una persona puede asumir valores no binarios. Pero no todas las empresas reportan géneros diferentes al femenino o masculino y es por eso la simplificación.

Estos datos (y estos otros) nos muestran que **las minorías subrepresentadas en las empresas de tecnología tenemos que enfrentar muchas más barreras para entrar, permanecer y crecer que cualquier hombre-cis-blanco con la misma formación y experiencia.** En consecuencia, no resulta descabellado pensar que los estereotipos racistas y sexistas subsisten en la tecnología que estas empresas desarrollan. Tecnología que consumimos todo el tiempo, y a través de la cual se toman decisiones, incluso del orden público.

Seteados en este contexto, volvamos al suceso de George Floyd. ¿Qué pasó al respecto en Sillicon Valley? La inmensa mayoría de las empresas más grandes de tecnología sacaron circulares manifestándose en repudio de los hechos de racismo, discriminación y violencia que se habían vivido en Estados Unidos, comunicando abiertamente su compromiso para incrementar la diversidad en sus equipos. Solamente para citar un ejemplo, Google donó 12 millones de dólares a organizaciones que trabajan contra las injusticias raciales.

Pero además de, finalmente, empezar a tratar como se debe el tema y a invertir recursos en mejorar la diversidad de los equipos que desarrollan tecnología, hubo otras noticias de 'tapa de revista de los que crean tecnología'. Un ejemplo de ello es la **revisión de la terminología que se usa en ingeniería de software y programación**. GitHub, la plataforma de desarrollo colaborativo más usada y más grande del mundo (que fue adquirida por Microsoft el año pasado), decidió que dejaría de usar la palabra *master* (que es equivalente a *amo* en inglés) para indicar que algo es principal, y comenzó a usar la palabra *main* que significa, justamente, principal. Otra notificación que nos llamó la atención fue un mail que mandó Airbnb (plataforma que permite buscar alojamiento temporario en casas, muchas veces compartiendo los espacios) a todos sus usuarios. La empresa nos contó que empezarían a usar nuestros datos para analizar si había comportamientos racistas entre usuarios y nos explicaban de qué forma lo harían, dándonos la opción de no querer participar.

"No resulta descabellado pensar que los estereotipos racistas y sexistas subsisten en la tecnología. Tecnología que consumimos todo el tiempo, y a través de la cual se toman decisiones, incluso del orden público..."

TWITTEAR

Pero, sin duda, las noticias que más llamaron la atención fueron las relacionadas a las herramientas de reconocimiento facial, uso muy popular de la Inteligencia Artificial. **Estos sistemas usan técnicas de inteligencia artificial que devuelven resultados racistas, sexistas y discriminatorios para otras minorías.** Y, por primera vez, las empresas reconocieron esto masivamente:

- El 8 de junio, el CEO de IBM mandó una carta al congreso diciendo, entre otras cosas, que iba a retirar del negocio sus productos de reconocimiento facial.
- A esto le siguió Amazon, que el 10 de junio prohibió, por el lapso de un año, el uso de *Rekognition* (su sistema de reconocimiento facial) para uso policial.
- No conformes con eso, y todavía no perdiendo el asombro por las noticias anteriores, al día siguiente se sumó Microsoft con una decisión en la misma línea

Pero no termina acá. De hecho, acá es cuando se pone bueno. Porque todo esto no quedó en la decisión de las empresas privadas, sino que trepó al siguiente nivel:

El 30 de junio la ACM (primera sociedad científica y educativa para educar acerca de la Computación), **llamó a una suspensión inmediata del uso privado y gubernamental actual y futuro de las tecnologías de reconocimiento facial, por razones técnicas y éticas**, a través de este comunicado.

La noticia generó shock. Hace tiempo que somos muchas (y cada vez más) organizaciones y personas que denunciamos la inequidad en el mundo de la programación y la tecnología, incluso investigadores e institutos de investigación que se dedican a estudiar el impacto social

de la IA y sus sesgos (AI Now Institute, Partnership on AI, Veronica Dahl) y hasta ahora no habíamos sido testigos de decisiones drásticas, tal como lo veníamos necesitando.

Esto no es un pato

Volvamos un pasito atrás para detenernos en los sistemas de reconocimiento facial. ¿Por qué devuelven resultados racistas? ¿Los programan así a propósito o sin querer queriendo?

El reconocimiento de imágenes es un uso muy popular de la IA. Voy a usar patos (sólo porque los carpinchos son muy Fase 1) para explicar brevemente cómo se crean los sistemas usando técnicas de aprendizaje de máquina (o *machine learning*), que son la forma a través de la cual una máquina *aprende* y se convierte en *inteligente*.

La inteligencia artificial permite a las computadoras adquirir ciertas habilidades propias de la inteligencia humana. ¿Cómo logran las máquinas este conocimiento? Veamos con un ejemplo:



Imagen de un pato. Este epígrafe lo puso la máquina, no yo. Ah re culpar al hermano.

La mayoría acordaríamos que eso es un pato y, si le preguntáramos a cualquier persona por la calle, habría altas probabilidades de que respondiera que es un pato. Pero ¿cómo hacemos que la computadora aprenda que eso es un pato? Tradicionalmente, le daríamos una secuencia de instrucciones: ¿tiene pico?, ¿tiene plumas? Y luego de aprenderse estas categorías, pediríamos que identifique, en una foto, si lo que ve es o no un pato. El problema es que, con esas dos instrucciones, esto también sería un pato para esta computadora:



Foto de una gallina. Tuve que buscar ‘alive chicken’ porque ‘chicken’ me devolvía pollos al horno.

Para crear sistemas de inteligencia artificial se usan técnicas con el objetivo de que la máquina aprenda por sí misma cuáles son las características que definen a un pato. Para eso, le vamos a mostrar miles o millones de fotos de patos y vamos a hacer que la máquina genere, a partir de eso, las reglas para poder identificarlo en el futuro. Que aprenda.



Collage de patos. Cuac.

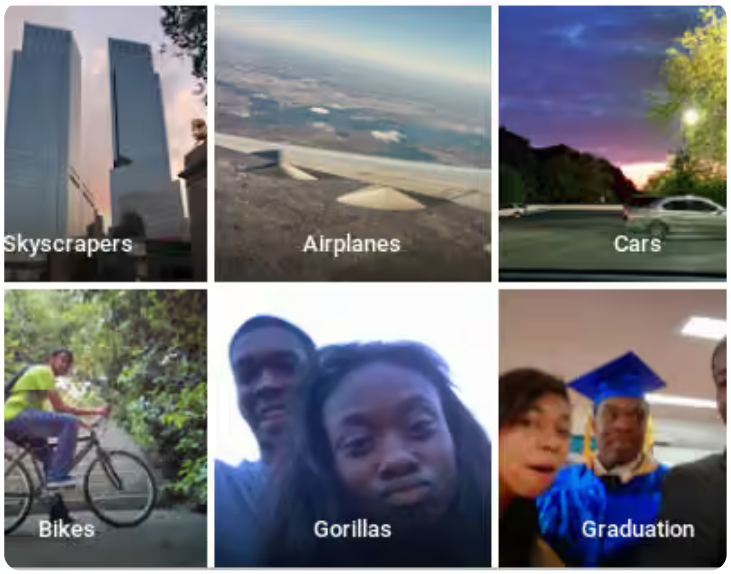
Esto fue un resumen muy simplificado de cómo aprende un sistema de Inteligencia Artificial. En base a eso que aprendió, ahora puede decidir autónomamente si lo que tiene enfrente es o no un pato. Los matices de este proceso están desarrollados acá.

Sólo para que nos hagamos una idea, la inteligencia artificial no es exclusiva de los androides y sistemas futuristas. También está presente en Siri, el traductor de Google, incluso Spotify y Netflix que son capaces de recomendar películas, series, o armar *playlists* después de analizar cientos de registros y recopilando material semejante al que hemos visto, escuchado o calificado positivamente. De hecho, Google ya anunció hace unos años que su estrategia es "lo primero es la IA" (*IA first,* en contraposición a la estrategia anterior "mobile first" que fue ubicar primero al desarrollo de la tecnología móvil), poniendo a la IA y **al aprendizaje de las máquinas a resolver la gran mayoría de sus problemas**.

Entrenar no te hace perfecto

Bien, ahora que más o menos quedó claro cómo se entrenan los sistemas de inteligencia artificial para que tomen decisiones, ¿por qué creemos que como resultado la IA es racista y sexista?

Resulta que esta forma de enseñarles a las máquinas cómo hacer para clasificar patos de no-patos al parecer tiene algunos problemas. Para las personas que trabajamos en sistemas, el ejemplo siguiente es un gran clásico: Google, cuando lanzó su sistema de reconocimiento de imágenes, fue denunciado por una usuaria de Twitter por taguear a sus amigos como gorilas.



Fuente: BBC

Un ejemplo más futurista está relacionado a los vehículos autodirigidos. Estos autos del futuro son mucho mejores detectando a personas blancas. Es decir que responden al siguiente patrón:



Esto ocurre porque, para que un sistema que usa reconocimiento facial pueda identificar, por ejemplo, a una persona potencialmente criminal en una estación de tren, se tuvo que entrenar ese sistema mostrándole descripciones de personas que alguna vez cometieron actos criminales. ¡Oh, vaya sorpresa! **Los sistemas se entrenan con datos cargados de sesgos, en este caso principalmente racistas.** Si los datos de los que aprende están sesgados (intencionadamente o no), el algoritmo, en consecuencia, va a decidir sesgado.

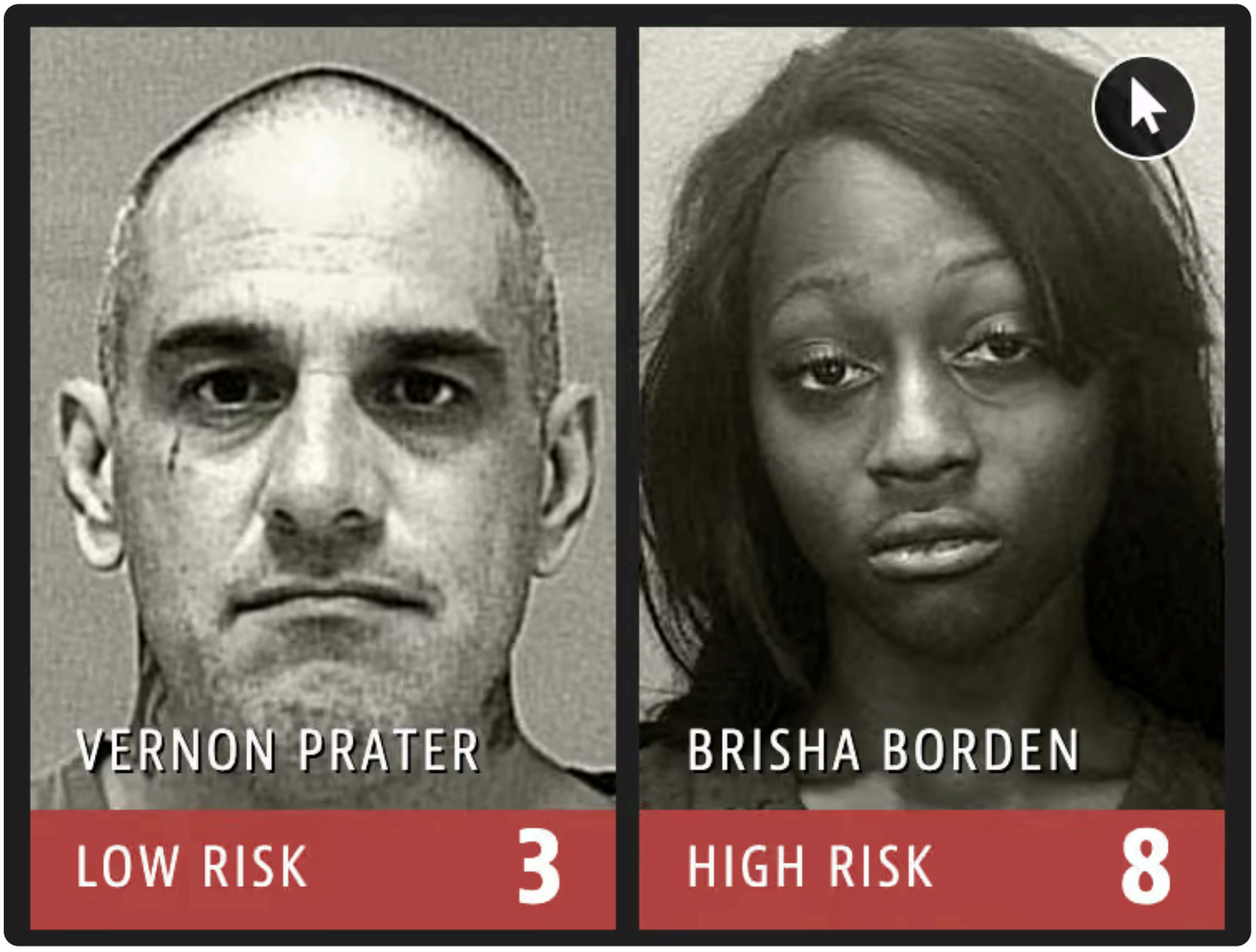
La discriminación que replican estos sistemas afecta a muchos grupos subrepresentados, principalmente en los sistemas que usan técnicas de inteligencia artificial para hacer predicciones. Un ejemplo muy concreto de un sistema con sesgo de género es un algoritmo que desarrolló Amazon para hacer más efectivo el proceso de reclutamiento. Para ello, creó un sistema que fue entrenado con los resultados de búsquedas de los últimos 10 años. Una aclaración necesaria aquí: hay muy pocas mujeres que estudian carreras relacionadas a la programación y a la tecnología (NSF), y las pocas que llegan tienen muchos más desafíos que sus pares varones para conseguir el puesto, es decir, los candidatos hombres casi siempre son elegidos por sobre las pocas candidatas mujeres. En consecuencia, la muestra no solo está estadísticamente desbalanceada (el sistema tiene muchísimas más entradas de aprendizaje de CVs de varones), sino que además aprendió de las prácticas sexistas de los equipos de reclutamiento. Como resultado, **el sistema procedió a descartar automáticamente todos los CVs de candidatas mujeres: un reflejo del dominio masculino en la industria tecnológica.** No solo eso, también rechazaba CVs de candidatos varones que hubiesen estado involucrados en alguna actividad con referencia a mujeres, por ejemplo, ser profesor de un bootcamp de programación para mujeres, como se cuenta acá.

Pero volvamos a los sistemas de reconocimiento facial que fueron el punto de partida de toda de esta cháchara. ¿Por qué se llegó a la decisión de su prohibición o suspensión?

Tu cara me suena

Uno de los usos más controversiales de este tipo de sistemas es el de determinar la potencialidad criminal de una persona y así poder tomar decisiones sobre una pena o

condena. En el 2016, la organización ProPublica denunció groseros sesgos raciales en los resultados del *software* COMPAS, que se utiliza para informar las decisiones sobre quién puede ser liberado en cada etapa del sistema de justicia penal, desde la asignación de pagos de fianza, hasta decisiones aún más fundamentales sobre la libertad de los acusados. Estos sistemas, que se alimentan de récords criminales históricos, toman datos personales (entre ellos, las fotos de las personas) para asignar un puntaje a las personas que acaban de cometer un crimen. Ese puntaje intenta predecir la probabilidad de que una persona reincida. En algunos estados de Estados Unidos, los resultados de tales evaluaciones se daban a los jueces durante la sentencia penal. En el informe muestran, entre otros, este ejemplo:



Estas dos personas cometieron delitos similares al momento en el que fueron evaluados por el algoritmo. La persona de la izquierda fue sentenciada anteriormente a 2 robos armados y un

intento de robo, mientras que la de la derecha fue condenada por delitos menores cuando era más joven. La persona de la izquierda, blanca, obtuvo menor puntaje según nivel de riesgo que la de la derecha. Analizaron el puntaje asignado a más de 7 mil personas arrestadas en Broward County, en Florida y contrastaron contra la cantidad de personas que efectivamente fueron acusadas por nuevos cargos en los años siguientes. **Solamente el 20% de las personas a las cuales el sistema les acusó de futuras criminales, efectivamente cometieron algún crimen.** Sumado a este elevado número de falsos positivos (en este caso, indicar erroneamente la reincidencia) siendo esta verdadera en la población, existía mayor probabilidad de que la fórmula señalara falsamente a los acusados negros como futuros delincuentes, etiquetándolos erróneamente de esta manera a casi el doble de tasa que los acusados blancos. No siendo lo suficientemente discriminatorio, los acusados blancos fueron mal etiquetados como de bajo riesgo con mayor frecuencia que los acusados negros

Y si bien esto suena terrible, es cierto que también parece algo muy propio de que ocurra en EE.UU. Parece, en otras palabras, un problema lejano. Hasta que nos enteramos de que en nuestro país vecino, Brasil, el 90% de las personas arrestadas porque fueron capturadas en cámaras y sistemas de reconocimiento facial eran negras. Sí, estos sistemas se usan en ciudades de carne y hueso para reducir la criminalidad. Estamos dando la responsabilidad a las máquinas de predecir cosas, pero en realidad lo que estamos haciendo es **usar datos del pasado para tomar decisiones del futuro.** Las máquinas, en este sentido, son como los niños que repiten todo. Estos niños son los adultos del mañana, los que van a hacer el mundo. Y en esta analogía, los sistemas de IA que se adoptan hoy, no sólo van a aplicar sesgos racistas aprendidos del pasado sino que también van a producir efectos reales en el futuro.



Fuente

Echarles la culpa a las máquinas que aprenden sería la salida fácil. El sesgo de los sistemas de inteligencia artificial se puede prevenir y también arreglar. Para ello es necesario que quienes crean y usan estos sistemas monitoreen continuamente la calidad de sus resultados y estén dispuestos a 'ajustar' las técnicas a través de las cuales la máquina aprende. Pero las máquinas están construidas por personas ¿Quiénes son estos humanos? Sumado a que los datos que se usan para alimentar estos sistemas están cargadísimos de patrones históricos de discrimación, **los equipos que construyen tecnología son muy poco diversos**, como vimos en el gráfico de demografía de Sillicon Valley.

En consecuencia, por lo general, no hay gente en esos equipos que pueda detectar esos errores porque probablemente ni se les ocurra probarlo. Faltan miradas, perspectivas y puntos de vista en cada una de las actividades de construcción de tecnología, que van desde la identificación de un problema o necesidad, hasta la distribución e implementación, pasando por varias etapas de desarrollo en la que intervienen, en su gran mayoría, equipos técnicos. Hace varios años que la industria del *software* y la tecnología está poniendo el foco en acciones que aumenten el número de 'mujeres en tecnología', pero es un enfoque un poco limitado, ya que —si nos basamos en lo visto hasta acá— probablemente privilegiará a las

mujeres blancas sobre otras. **Necesitamos reconocer cómo las intersecciones de etnia, género y otras identidades y atributos dan forma a las experiencias de las personas, en particular con la IA.** Este estudio, por ejemplo, demuestra que los algoritmos de reconocimiento facial realizados por Microsoft, IBM y Face++ tenían más probabilidades de identificar erróneamente el género de las mujeres negras que de los hombres blancos.

“...lo que necesitamos es que las miradas técnicas se complementen con otras miradas ricas y diversas para hacer un uso adecuado y responsable de la tecnología.”

TWITTEAR

Ojo, el mismo *software* que usa la policía para predecir la criminalidad, ayuda a rescatar a víctimas de trata de personas y reunir a niños desaparecidos con sus familias. Entonces, ¿qué hacemos? ¿Metemos a la Inteligencia Artificial en la columna de villanos? No, no queremos llegar a eso. Reconocer estos sesgos no nos hace pegarle un portazo al desarrollo de IA en la cara, sino incorporar otros (muy necesarios) criterios para su desarrollo. Las personas que estamos formadas en roles técnicos no tenemos la capacitación suficiente en ética: en la universidad no tenemos materias ni prácticas relacionadas a la ética. Entonces, en este escenario, **lo que necesitamos es que las miradas técnicas se complementen con otras**

miradas ricas y diversas para hacer un uso adecuado y responsable de la tecnología. No hay un manual ético para aplicar esto todavía. Lo importante es que es un debate que se está dando en las empresas tecnológicas, porque el daño que se genera es grande. La ACM misma dijo en su comunicado: "*El Comité concluye que, cuando se evalúa rigurosamente, con frecuencia la tecnología produce resultados que demuestran un sesgo claro basado en características étnicas, raciales, de género y otras características humanas reconocibles por los sistemas informáticos. Las consecuencias causan daños profundos, particularmente a las vidas y los derechos fundamentales de las personas en grupos demográficos específicos, incluidas algunas de las poblaciones más vulnerables de nuestra sociedad*".

La inteligencia artificial está en muchos lugares de nuestra vida, incluso en decisiones como el acceso a un crédito hipotecario, un tratamiento médico o una condena penal. Como estos sistemas funcionan a partir de algoritmos, **predomina la idea de que son objetivos** y neutrales y que, en consecuencia, van a tomar mejores decisiones y con más imparcialidad que si las tomaría una persona. En instituciones gubernamentales o de interés público, la adquisición e implementación de sistemas de IA suele realizarse sin escrutinio público ni la atención a preocupaciones de la sociedad civil [Los verdaderos riesgos de la IA]. Una pata importante de toda esta cuestión es incorporar criterios y miradas al desarrollo de estos sistemas que son muy prometedores en términos de modernización pero que rara vez son neutrales. La otra pata grande está en cuidar el proceso de implementación de estos sistemas en los diferentes organismos públicos.

En un contexto de crisis global, como la de la pandemia que estamos viviendo, se está hablando y tratando de entender si la IA podría ayudar a salvar la vida de las personas. Y sí, sería hermoso. Pero la evidencia nos muestra que no tenemos procedimientos éticos para que la Inteligencia Artificial se pueda usar de manera segura. Al menos, no todavía.

Referencias

Chelsea Barabas. Beyond Bias: Re-Imagining the Terms of ‘Ethical AI’in Criminal Law.

WEST, Sarah Myers; WHITTAKER, Meredith; CRAWFORD, Kate. Discriminating systems: Gender, race and power in AI. AI Now Institute, 2019.

Of course technology perpetuates racism. It was designed that way, MIT Technology Review.

“No tenemos procedimientos éticos sólidos para usar IA de forma segura”, MIT Technology Review.

“A challenging road”, The Margins.

“Data and Evidence on Gender Equality in Digital Access, Skills, and Leadership”, Equals.
