

## INTOLERANCIA ARTIFICIAL

Los expertos alertan de que las máquinas inteligentes pueden amplificar los estereotipos sexistas, racistas y clasistas

+4 más MAYTE RIUS



La inteligencia artificial ya se usa en prácticamente todos los grandes servicios online con los que interactuamos las personas y en muchos casos también está decidiendo quién tiene derecho a una hipoteca o a un seguro de vida, qué tratamiento médico debe recibir un enfermo, quién puede acceder a un puesto de trabajo o qué delincuente debe ser encarcelado por su alto riesgo de reincidencia. Y como funciona a partir de algoritmos, que son un conjunto de operaciones matemáticas, impera la idea de que estos son objetivos y neutrales y que, en consecuencia, las máquinas tomarán decisiones de una forma más eficaz e imparcial que las per-

sonas, superando los prejuicios o condicionantes sociales y culturales. Pero no es así.

Cada vez son más las voces que alertan de que la inteligencia artificial (IA) no es neutral, que a veces sus resultados están sesgados, impregnados de machismo o de racismo, y que se corre el riesgo de que las máquinas, a través del aprendizaje automático, refuercen los estereotipos sexistas, racistas y clasistas que subyacen en la sociedad y acaben siendo más "intolerantes" que los humanos.

"Ningún sistema de inteligencia artificial tiene intencionalidad, pero las decisiones que aprende están basadas en los datos con los cuales ha sido entrenado, y si esos datos están sesgados (intencionadamente o no), el algoritmo decidirá sesgado, y ese sesgo puede tener consecuencias muy drásticas que afecten a la vida de las personas", explica el director del Instituto de Investigación en Inteligencia Artificial del Consejo Superior de Investigaciones Científicas (CSIC), Ramón López de Mántaras.

"Hay dos formas en que un sistema de inteligencia artificial puede mostrar prejuicios: primero, porque se usen datos inadecuados y, segundo, porque el procesamiento de los datos sea inadecuado", afirma Carlos Castillo, que dirige el grupo de Ciencia Web y Computación Social en la UPF. Y detalla que los datos también pueden ser inadecuados de muchas maneras: porque contengan patrones históricos de discriminación –por ejemplo, que aprenda de las estadísticas que los cargos ejecutivos son mayoritariamente desempeñados por hombres blancos y a la hora de seleccionar candidatos para una vacante de este tipo descarte currículums de mujeres y de hombres de raza negra–, o porque se seleccionen mal. "Quizá monitorizar el tráfico de coches es más fácil que monitorizar el

de bicicletas o el de desplazamientos a pie, pero si seleccionamos solamente los datos de automóviles entonces propondremos políticas de movilidad más adecuadas para los viajes en coche que para otros viajes", ejemplifica Castillo.

Y agrega que tampoco hay que olvidar que las muestras de datos que se procesan no son necesariamente aleatorias ni representativas: "Por ejemplo, si usamos datos de redes sociales hay que tener en cuenta que cada una tiene su propia composición demográfica, sus propias comunidades y normas, de modo que las conclusiones que se obtengan usando esos datos no son aplicables al público en general" y si se hacen extensivas serán inadecuadas.

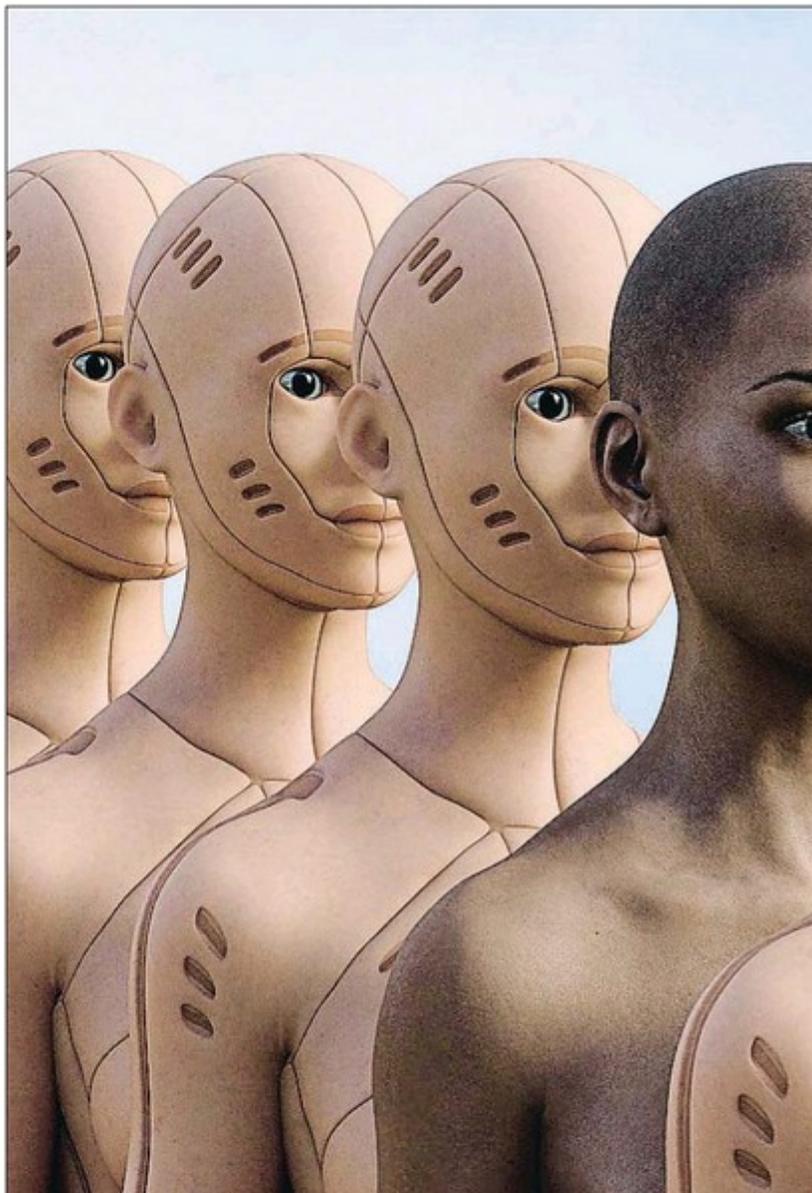
Son muchos los casos que ilustran que los sistemas de inteligencia artificial están trabajando con conjuntos de datos falsos, imprecisos o no representativos porque no tienen en cuenta, infrarrepresentan o sobrerrepresentan a determinados colectivos o circunstancias. Hace un año, un profesor de informática de la Universidad de Virginia notó un patrón en algunas de las suposiciones que hacía el sistema de reconocimiento de imágenes que estaba construyendo: las imágenes de cocinas las asociaba con mujeres y no con hombres, y se cuestionó si no estaría inculcando

Prejuicios en el programa. Por ello decidió revisar con otros colegas las dos grandes colecciones de fotos que se usan para "entrenar" a las máquinas inteligentes en el reconocimiento de imágenes. Y sus resultados, publicados este verano, son muy esclarecedores: en ambas colecciones las imágenes de compra y lavado están vinculadas a mujeres, mientras que las de prácticas deportivas y tiro están ligadas a los hombres. Y el programa que aprendía con estos datos no sólo reflejaba esos sesgos sino que los amplificaba

ba: si un conjunto de fotografías “generalmente” asociaba a las mujeres con la cocina, el software entrenado con ellas creaba una asociación aún más fuerte en sus etiquetas, porque detecta la tendencia subyacente y apuesta por ella para acertar. Y dado que muchas herramientas tecnológicas emplean fotos de redes sociales para identificar preferencias y patrones de los usuarios y pueden haber sido entrenadas con esas imágenes, se corre el riesgo de que refuercen los prejuicios sociales existentes.

“No construimos inteligencia artificial para que repliquen los errores que cometen las personas, como las actitudes sexistas, xenófobas o maleducadas; pero evitarlo es complejo porque a menudo las máquinas aprenden de la interacción con las personas y podrían reproducir algunos de sus comportamientos, sobre todo ahora que vivimos un boom de la inteligencia artificial pero la mayoría de disciplinas aún no están maduras y, como cualquier adolescente, son vulnerables”, justifica Miquel Montero, experto en inteligencia artificial y CEO y fundador de Atomian, firma que desarrolla y comercializa software de computación cognitiva.

Carme Torras, profesora de investigación en el Instituto de Robótica CSIC-UPC y autora, entre otros, de Enxarxats (Males Herbes), se muestra preocupada por esta “vulnerabilidad” de la inteligencia artificial y advierte que evitar los sesgos algorítmicos “es un poco responsabilidad de los informáticos que desarrollan los programas pero también de todos los que volcamos contenidos en internet”. “Con el machine learning o aprendizaje automático la máquina aprende cruzando datos y puede llegar a conclusiones nefastas o superbondadosas en función de los datos que haya de ti”, coincide la subdirectora del Observatorio de Bioética y Derecho de

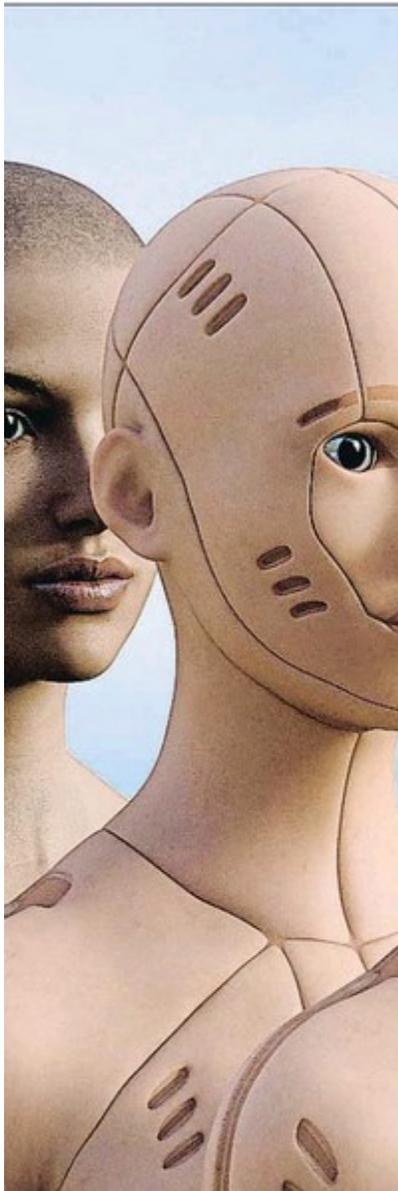


la UB, Itziar de Lecuona.

Alberto Robles, director general de Expert System Iberia, una de las firmas líderes en inteligencia cognitiva semántica, admite el riesgo de que algunos usos de la tecnología “roquen lo éticamente válido”, como podría ser que las aseguradoras depuren con inteligencia artificial la información que circula en las redes sociales para identificar a quién le gusta la escalada o los deportes de motor e incrementar así el precio de su seguro de vida por considerarlos más propensos a su-

frir accidentes. No obstante, asegura que la mayoría de empresas dedicadas a la inteligencia artificial ha desarrollado códigos éticos estrictos marcando los límites de lo permitido o no mediante el uso de la tecnología y se esfuerzan por no introducir sesgos y analizar las situaciones discriminatorias de mino-

El sesgo de los algoritmos constituye un problema social porque cada vez toman más



# Tende

## Los expertos alertan inteligentes pueden estereotipos sexistas

A menudo los sistemas de IA trabajan con conjuntos de datos falsos, imprecisos o no representativos

Los ciudadanos deben exigir transparencia a los algoritmos y los datos que hay detrás

Algunos expertos exigen que se revise que un algoritmo es ética y socialmente aceptable antes poder usarse

decisiones

Rías que pueda presentar la información que utilizan. En este sentido, indica que en Expert System aplican tres niveles de comprobación interna al entrenar al sistema de inteligencia artificial. "El sesgo lo puedes introducir cuando le enseñas un nuevo concepto –por ejemplo 'trabajar como un negro'– y en cómo se lo haces relacionar con otros –definiciones de la RAE, textos donde aparece...–, y lo que hacemos es que el trabajo de una persona es contrastado por otras tres para que lo que entienda el sistema sea correcto", detalla.

Montero, de Atomian, coincide en que los sistemas de aprendizaje supervisado y basado en conoci-

### Historias relacionadas

Inteligencia artificial, pero con el sexismo de siempre

La Vanguardia (1ª edición) 24 Oct 2017

Educar a las máquinas

La Vanguardia (1ª edición) 24 Oct 2017

# Tendencias

Los expertos alertan de que las máquinas inteligentes pueden amplificar los sesgos sexistas, racistas y clasistas

Los ciudadanos deben exigir transparencia a los algoritmos y los datos que hay detrás

Rías que pueda presentar la información que utilizan. En este sentido, indica que en Expert System aplican tres niveles de comprobación interna al entrenar al sistema de inteligencia artificial. “El sesgo lo puedes introducir cuando le enseñas un nuevo concepto –por ejemplo ‘trabajar como un negro’– y en cómo se lo haces relacionar con otros –definiciones de la RAE, textos donde aparece...–, y lo que hacemos es que el trabajo de una persona es contrastado por otras tres para que lo que entienda el sistema sea correcto”, detalla.

Montero, de Atomian, coincide en que los sistemas de aprendizaje supervisado y basado en conoci-



Tay, la bot de Microsoft, sólo vivió dieciséis horas después de lanzarla

miento simbólico minimizan los riesgos de que las máquinas discriminen o partan de conocimientos sesgados como ocurre en los sistemas de aprendizaje automático basados en las estadísticas y en que la

máquina aprenda a base de leer lo que hay por internet. Ambos especialistas en inteligencia artificial enfatizan, no obstante, que la única vía para evitar el sesgo algorítmico y que la tecnología actúe con preju-

icios es que los ingenieros y todos los involucrados en su desarrollo sean cada vez más humanistas, tengan más formación en valores, sean conscientes de su responsabilidad y del impacto decisivo de su trabajo en la sociedad, y se autorregulen.

Claro que, como también expresan López de Mántaras, Castillo y De Lecuona, la formación ética y la autorregulación son necesarias pero no suficientes, y para evitar que las máquinas inteligentes discriminen o perpetúen estereotipos hace falta también que los ciudadanos exijan legislaciones que obliguen a dar transparencia a los algoritmos y a los datos que están detrás de las decisiones de la inteligencia artificial, que se reconozca el “derecho a la explicación”. “Los algoritmos tendrían que pasar unos controles estrictos antes de ser utilizados por bancos o por compañías de seguros; igual que los medicamentos o los alimentos no se pueden vender sin que las autoridades sanitarias lo autoricen, debería haber una agencia certificadora que revise que un algoritmo es ética y socialmente aceptable antes de que pueda usarse”, enfatiza López de Mántaras.

Fuentes de Google –algunos de cuyos sistemas de inteligencia artificial han protagonizado denuncias de discriminación en los últimos años, tal y como recogen los ejemplos que acompañan esta información– aseguran que las máquinas también pueden usarse para reducir la discriminación, como es el caso de la herramienta que ellos han desarrollado para detectar prejuicios de género en los contenidos de cine, televisión o publicidad.