https://info.nodo50.org/Si-el-algoritmo-es-racista-es-porque-se-ha-entrenado-con-datos-racistas.html

**Entrevista a Ana Valdivia, investigadora en inteligencia artificial de la Universidad De Oxford**

# “Si el algoritmo es racista es porque se ha entrenado con datos racistas”

- Noticias - Noticias Destacadas -

Fecha de publicación en línea: Domingo 12 de mayo de 2024

**Copyright © Nodo50 - Todos derechos reservados**

**Ana Valdivia (Barcelona, 1990) es profesora e investigadora en Inteligencia Artificial, Gobierno y Políticas en el Oxford Internet Institute de la Universidad de Oxford. Matemática e informática, ha estudiado la influencia en las sociedades de la recopilación masiva de datos o el uso de algoritmos en las fronteras. Actualmente su trabajo se centra en los impactos medioambientales y sociales de la inteligencia artificial. Colabora con organizaciones como AlgoRace, que analiza los usos de la IA desde una perspectiva antirracista, y escribe en el blog** *La paradoja de Jevons*. Atiende a CTXT por videoconferencia.

Ya que es usted matemática e informática, me gustaría pedirle en primer lugar que nos explique qué es eso que llamamos inteligencia artificial y hasta dónde puede llegar, porque claro, tenemos a los señores de OpenAI hablando de “riesgos catastróficos para la humanidad”...

Pues a ver, ¿qué es la inteligencia artificial? A mí me gusta mucho la definición que está reflejada en la nueva Ley de Inteligencia Artificial de la Unión Europea. Explica que es un conjunto de *hardware y software* en el que un algoritmo se programa con un objetivo y llega a alcanzar ese objetivo de la manera más eficiente, algorítmicamente hablando, con datos.

O sea, básicamente es un algoritmo que se programa en un ordenador, o en un servidor, y que alcanza un objetivo aprendiendo de los datos que le han sido dados. No es algo nuevo. El concepto de inteligencia artificial se acuñó en 1956 en Estados Unidos, lo que pasa es que en aquella época la capacidad computacional de los ordenadores no era la que tenemos ahora, ni las sociedades estaban tan “datificadas”: hoy en día se recogen muchos más datos que en 1956. Entonces, se ha producido una explosión de esta tecnología porque hay ordenadores más potentes y hay datos con los que entrenar esos algoritmos en esos ordenadores potentes.

En esta definición de la ley europea también se explica que hay diferentes técnicas en las que se puede basar un algoritmo de inteligencia artificial, que son el aprendizaje automático, el aprendizaje profundo, algoritmos basados en reglas predefinidas o métodos más estadísticos como la heurística. Son conceptos muy técnicos, pero creo que esa definición está muy bien.

¿Hasta dónde puede llegar esta tecnología? Pues eso depende de la sociedad y de las manos en las que caiga. Yo publiqué un escrito en 2020 en el que hablaba sobre los mitos de la inteligencia artificial. Predecía que en los siguientes años los avances de la inteligencia artificial iban a recaer en manos de empresas privadas porque son las que tienen la capacidad de pagar la infraestructura para entrenar algoritmos como ChatGPT. Y es lo que está pasando. Desde las universidades ya nos hemos quedado muy cortas porque no tenemos esa capacidad computacional.

La IA llegará hasta donde quieran estas empresas privadas y hasta donde la regulación les permita. Ahora se le están poniendo trabas a OpenAI con los datos, por el tema del *copyright*. Uno de los talones de Aquiles de la inteligencia artificial son los datos. Sin ellos no puedes entrenar algoritmos. ChatGPT se ha entrenado extrayendo todos los datos de internet, pero muchos tenían un *copyright*…

**Ha mencionado la gran cantidad de recursos que hacen falta para sacar adelante estos sistemas y creo que eso es lo que está estudiando ahora mismo. No sé si podría contar un poco de eso, de la parte más “física” de la IA.**

Llevo muchos años investigando la inteligencia artificial y siempre he analizado la parte del código: cómo crear algoritmos que sean más transparentes o más justos, cómo mitigar los sesgos, etc.

En los últimos años me he ido dando cuenta de que la parte de la infraestructura, de la materialidad de la inteligencia artificial, estaba muy poco analizada. Y creo que es algo que se debe tener en cuenta en estos marcos de rendición de cuentas algorítmica. Cuando tú auditas un algoritmo, no solo tienes que auditar el código, también tienes que auditar qué empresa lo ha hecho, bajo qué *software*, cuánta agua se ha gastado, cuánto carbono se ha emitido, si ha tenido algún impacto en las comunidades locales... Es lo que estoy estudiando ahora, desde el origen: qué minerales se necesitan para crear las GPUs, que son los microchips con los que se entrenan algoritmos como ChatGPT porque tienen la capacidad de procesar algoritmos sofisticados de manera más rápida. Quién está fabricando GPUs a nivel mundial, que es Nvidia, con un 80% del mercado de GPUs. Casi toda la infraestructura de la inteligencia artificial recae en esta empresa. Luego, cuando esos microchips se envían a centros de datos, cuánta energía gastan, cuánta agua. Por último, el final del ciclo. Cada cinco años los centros de datos tienen que renovar su infraestructura, eso significa que cada cinco años las GPUs de un centro de datos se desechan; pues bueno, dónde se desechan, cómo se reciclan… Y qué impacto medioambiental y social está teniendo cada una de esas fases.

Estudio eso y también otra parte de la industria de la inteligencia artificial, que es quién está etiquetando los datos, quién está entrenando los algoritmos, etc. Siempre hace falta mucho trabajo humano.

**¿Hasta qué punto el consumo de recursos de la IA es superior al que ya tenía la industria de las tecnologías de la información y comunicación? Porque los megacentros de datos ya existían…**

Las GPUs, que son los chips que se utilizan para jugar a videojuegos y para la inteligencia artificial, consumen mucha más agua y mucha más electricidad porque son más sofisticados.

El primer móvil que tuvimos gastaba mucha menos electricidad que el móvil que tenemos ahora, porque ahora tenemos nuestra vida digital, claro. Se da la paradoja de Jevons. La tecnología cada vez es más eficiente pero cada vez hay más, cada vez necesitamos más centros de datos, tenemos cada vez más aparatos digitales, entonces sí, todo es más eficiente pero, al fin y al cabo, estamos consumiendo mucho más.

**El hecho de que el control de la tecnología esté quedando en manos privadas, ¿qué consecuencias puede tener en su desarrollo?**

En el campo de la inteligencia artificial siempre ha habido colaboraciones público-privadas de empresas tecnológicas con universidades. Por ejemplo, el primer *chatbot* que se codificó fue obra de IBM y la Universidad de Georgetown en Estados Unidos.

Pero últimamente está recayendo solo en manos privadas porque son los que tienen los datos y la capacidad computacional, lo vemos por ejemplo con Twitter. Twitter era una fuente muy rica de datos para las investigadoras académicas como yo, porque podías analizar ciertos comportamientos sociales en redes, pero esa información ya no está disponible.

Cuando Elon Musk decidió dejar de facilitarla, todos mis estudiantes entraron en pánico. Ahora tienes que pagar si quieres tener acceso a estos datos. Se han privatizado todas las fuentes de información, pertenecen a Microsoft, Amazon, o Google. Como ellos tienen la materia prima de los datos, ya nos queda muy poco que hacer a las universidades.

Copyright © Nodo50

Luego está el tema de la capacidad de cómputo. En mi departamento, en Oxford, ahora empezamos a tener GPUs, pero son muy costosas.

Las universidades jugamos en un segundo nivel en cuanto a desarrollo tecnológico. Ahora estamos auditando lo que están haciendo las empresas privadas. Poniendo el ojo crítico o desarrollando cosas a partir de lo que ellos han desarrollado. ¿Cuántos *papers* científicos están ahora analizando el ChatGPT, sus sesgos y sus aplicaciones? Esta va a ser la tendencia en los próximos años.

**Sobre el tema de los sesgos, he estado leyendo el informe Una introducción a la IA y la discriminación algorítmica para movimientos sociales**, de AlgoRace, del que es una de las investigadoras principales. Muchas veces en redes, cuando alguien habla de los sesgos de la inteligencia artificial, aparece otro que responde que las personas, los funcionarios, también tienen sesgos. La conclusión a la que se llega en el informe, si no he entendido mal, es que la inteligencia artificial va a ser racista mientras el sistema en su conjunto sea racista, pero no sé si se puede mitigar esto de alguna manera.

Veíamos que en España todo el mundo decía que la inteligencia artificial es racista, y para mí esa narrativa es una manera de escurrir el bulto de las grandes tecnológicas y de las personas que están diseñando esa inteligencia artificial racista, porque la inteligencia artificial en sí es una herramienta. Es como un martillo. Lo puedes utilizar para clavar un cuadro en la pared y poner tu casa más bonita o para hacer daño a una persona. La inteligencia artificial la puedes utilizar para seguir reproduciendo violencias estructurales, por ejemplo, con un algoritmo que haga más difícil pedir ayudas públicas a comunidades históricamente marginalizadas, o puedes crear un algoritmo que te analice movimientos financieros dentro de los partidos políticos e identifique quién está haciendo movimientos corruptos.

En España aún no tenemos un algoritmo que detecte ese tipo de corrupción, pero tenemos el algoritmo Bosco que dice si tienes derecho al bono social para la factura de la luz o no. La fundación Civio ha querido auditarlo, pero no nos han permitido acceder al código.

**Me ha sorprendido saber que se están utilizando ya muchos algoritmos de este tipo en España, no lo sabía.**

Sí, sí, se utilizan mucho. Está el Bosco; está Viogen, que es el de la violencia de género; hay otro de violencia de género en el País Vasco, yo misma lo estuve auditando con un juez y con una experta en temas legales y de tecnología. Una cosa que me resulta interesante es que la mayoría de veces la justificación para implementar un algoritmo en la vida pública es la falta de recursos. Por ejemplo, en el caso del algoritmo de violencia de género en el País Vasco, la justificación fue que la Ertzaintza no tenía suficientes expertos en violencia de género.

Entonces dijeron, bueno, como tenemos falta de recursos humanos, lo que vamos a hacer es poner un algoritmo que prediga el riesgo de violencia de género que tiene cualquier persona que venga a nuestra comisaría a reportar que está sufriendo esta violencia.

Una de las cosas que decimos Javi [Javier Sánchez Monedero, el otro investigador principal del informe] y yo es: no, primero mejora la infraestructura, pon los recursos necesarios en ese sitio. Una vez la Ertzaintza tenga los recursos para evaluar de una manera humana esos casos de violencia, entonces sí, pon el algoritmo como una herramienta extra, pero un algoritmo que esté bien diseñado, porque este algoritmo está muy mal diseñado, funciona muy, muy mal. Y está hoy en día asesorando casos de violencia de género en el País Vasco.

La inteligencia artificial no es una solución a ningún problema estructural. Se tendría que poner como una herramienta de ayuda ante un sistema público bien financiado y con trabajadores en buenas condiciones.

Copyright © Nodo50

**También puede ser una forma de escurrir el bulto, supongo, porque lo que haga la máquina no es culpa de nadie…**

A mis alumnos en la Universidad de Oxford siempre les desmitifico la idea del *black box*, de la caja negra, siempre les digo que los algoritmos no son una caja negra, que muchas veces las cajas negras son las instituciones, porque cuando yo, como informática y matemática, tengo acceso al código de un algoritmo, puedo preguntarle al algoritmo cómo está tomando las decisiones, lo puedo auditar. El problema es que muchas veces las administraciones o las instituciones no te dan permiso.

Pero es posible interrogar a los algoritmos igual que a las personas, incluso el algoritmo es un poquito más transparente porque no te puede mentir. Tú lo estás viendo, son fórmulas matemáticas, en cambio una persona sí que te puede mentir sobre cómo ha tomado una decisión.

Así que los procesos algorítmicos siempre son más transparentes que los humanos. Incluso esa transparencia permite detectar cuándo una institución tiene sesgos racistas.

Esto lo hemos visto en el Reino Unido, cuando el Ministerio de Interior decidió implantar un algoritmo para analizar las solicitudes de visados para entrar al país. Lo tuvieron que cancelar porque casi todas las personas de África recibían una puntuación muy alta, que indicaba que su caso debía examinarse en profundidad, lo que alargaba el proceso.

Entonces decidieron cancelarlo porque efectivamente la ley en el Reino Unido exige un trato igualitario sin tener en cuenta tu nacionalidad, tu género, tu orientación sexual, etc. Y se estaba vulnerando esa ley porque el algoritmo valoraba en función de la nacionalidad.

Vale, se canceló y todo bien, pero yo digo que tendríamos que ir más allá, porque el algoritmo estaba mostrando que históricamente las solicitudes que venían de África recibían un mayor escrutinio, que los humanos las estaban tramitando así. Los algoritmos pueden revelar patrones racistas o sexistas de nuestras instituciones. Si el algoritmo es racista es porque se ha entrenado con datos racistas, porque los humanos que han producido esos datos tenían comportamientos racistas.

**Me ha llamado la atención el ejemplo que ha puesto antes de un algoritmo para investigar la corrupción. ¿Cree que los sectores progresistas deberían utilizar más estas herramientas o explorar un poco ese tema?**

Pues sí, estaría muy bien que se crearan ese tipo de escrutinios algorítmicos. Sería importante ver quién los diseña, claro, pero un algoritmo te puede mostrar muchísimas cosas porque básicamente analiza patrones en los datos. Así que estaría muy bien que grupos progresistas de nuestro país abogaran por el uso de algoritmos. Vemos cómo estas tecnologías sirven a los poderosos, en vez de utilizarse como una herramienta del pueblo.

**Por último, ¿cómo valora la reciente Ley de Inteligencia Artificial de la Unión Europea?**

Tiene cosas buenas y malas. Es bueno que se regule esta tecnología. No sé hasta qué punto teníamos ya regulaciones que podían servir. Por ejemplo, la Ley de Protección de Datos. No sé si era necesaria una regulación específica o habría que fortalecer más las que ya teníamos, pero bueno, aun así está bien que se regule.

Por otro lado, vemos muchos vacíos. Por ejemplo, el artículo 83 de ese reglamento dice que todas las bases de datos destinadas a contexto migratorio de la Unión Europea están exentas de la regulación. O sea, que hecha la ley,

Copyright © Nodo50

hecha la trampa también. La Unión Europea se lava las manos y dice que se considerará de riesgo alto toda inteligencia artificial implementada en el contexto migratorio, pero que las suyas propias están exentas. Hay también otros temas como el del reconocimiento facial en vivo, que parecía que se iba a prohibir totalmente, pero al final la policía va a poder utilizarlo. En temas de seguridad también la legislación es muy laxa. Entonces, bueno, está bien, pero está mal. Y también, recordando lo que hemos hablado del impacto medioambiental, este reglamento dice que la inteligencia artificial se utilizará de una manera sostenible, pero no dice cómo de sostenible ni qué significa sostenible ni qué directrices existen.

Copyright © Nodo50