ARTICLES

Fotografia extreta de Freepik

ALGORITMO COMPAS: ANÁLISIS FILOSÓFICO Y ÉTICO SOBRE LA DISCRIMINACIÓN ALGORÍTIMICA A CIERTOS COLECTIVOS MARGINALES

Roger Fernández Fernández

El presente estudio tiene como intención un análisis sobre la discriminación algorítmica, en este caso, sobre un caso teórico concreto que es el COMPAS, una inteligencia artificial la cual se basa en un sistema de algoritmo de predicción de riesgos. La inteligencia artificial (IA) se refiere a la capacidad de las máquinas para realizar tareas que normalmente requieren inteligencia humana. Esto incluye el aprendizaje, el razonamiento, la resolución de problemas, el reconocimiento de patrones y la toma de decisiones. Es importante señalar que, junto con estos avances, también surgen desafíos y preocupaciones, como la pérdida de empleo debido a la automatización, cuestiones éticas relacionadas con la privacidad y la seguridad, y la necesidad de desarrollar marcos regulatorios adecuados. El impacto de la IA continúa evolucionando, y su adopción se espera que siga creciendo en diversos sectores.

La discriminación algorítmica que se dió en los Estados Unidos, relacionada con el sistema COMPAS, se hizo relevante en 2016 cuando una investigación detallada de ProPublica reveló la posibilidad de sesgos raciales en un algoritmo para la toma de decisiones judiciales. Este suceso destacó la intersección entre la inteligencia artificial y el sistema de justicia, planteando preguntas críticas sobre la equidad y la transparencia en la

implementación de estas tecnologías en contextos sensibles. El COMPAS, que significa "Correctional Offender Management Profiling for Alternative Sanctions", es un algoritmo de evaluación de riesgos utilizado en el sistema de justicia penal de los Estados Unidos. Fue desarrollado por Northpointe Inc. (ahora Equivant) y se utiliza para predecir la probabilidad de que un individuo reincida en actividades delictivas en el futuro.

El objetivo principal del COMPAS es ayudar a los jueces a tomar decisiones objetivas sobre sentencias y libertad condicional, proporcionando una evaluación del riesgo de reincidencia de un acusado. El algoritmo utiliza datos históricos para analizar factores como el historial criminal, la edad, el género y otros datos personales para generar una puntuación de riesgo. Sin embargo, el caso fue destacado en investigaciones y reportajes, como el realizado por ProPublica en 2016.

A partir de estos hechos, el uso del COMPAS ha generado varias controversias y críticas. Se puede notar que hoy en día, la inteligencia artificial tiene una influencia más que notable. En el entramado de las diferentes sociedades, surgen debates y un sin fin de opiniones de carácter filosófico y moral, sobre el uso y la incorporación de estas inteligencias en los ámbitos socioculturales.

Con las limitaciones actuales es difícil hallar un sistema inteligente ideal, pues por desgracia, las sociedades y expertos postmodernos, en esta disciplina, no tienen las capacidades suficientes como para asegurarse de que estas tecnologías cognitivas siempre hagan el bien y sean justas. Por lo que, indirectamente y sin voluntad de hacer el mal, existen consecuencias negativas y catastróficas para sus pobres víctimas. Es por ello por lo que existen una multiplicidad de cuestiones éticas que se dan en el caso COMPAS. Entre ellas están: el sesgo algorítmico, que es el carácter discriminatorio que tiene este algoritmo de predicción. Otro de los problemas, que contribuye a generar una gran desconfianza en esta inteligencia artificial, es la falta de transparencia, pues la empresa que diseñó el producto se niega a compartir el sistema de puntuación que utiliza el algoritmo para distinguir los potenciales reincidentes de los que no. En consecuencia, aquellos que son etiquetados como posibles reincidentes no pueden rebatir dicho resultado al no saber el criterio con el cual han sido juzgados. Hay que destacar que el uso de los algoritmos en la toma de decisiones judiciales abre la puerta al debate ético en relación a los derechos civiles y humanos, dado que la sola idea de que estos programas tengan la capacidad de influir significativamente en las vidas de las personas puede debilitar la equidad e imparcialidad del sistema judicial. La validación y la precisión en la predicción sobre la predisposición de reincidencia delictiva también es cuestionada.

Estas cuestiones no son las únicas, ya que también surgen preguntas de carácter moral en relación al uso de los datos con los que el algoritmo aprende, o bien otros temas como es el de la responsabilidad. Para empezar, está la base de datos, la cual puede traer serios riesgos, en caso de que estos datos estén sesgados. El caso COMPAS hace patente la necesidad de crear un sistema o unas políticas que puedan hacer frente a este problema. Es por ello que se debe hacer especial hincapié en la búsqueda de una forma que facilite la distinción entre los datos que pueden estar sesgados y aquellos que no. Hay múltiples factores por los que una IA puede tener sesgos, pero algunos surgen por una mala selección de datos los cuales pueden resultar de un conjunto de datos incompleto o no representativo.

Otra pregunta que surge en medio del debate es qué tipo de datos se deberían de utilizar, pues posiblemente la selección de ciertos datos puede facilitar que se dé el sesgo. Este es un tema que se ha abordado en los tiempos recientes desde los estudios de la filosofía de la inteligencia artificial. Uno de los referentes en este campo, como es Mark Coeckelberg, dedica parte de su libro Ética de la inteligencia artificial al algoritmo COMPAS y los problemas que conlleva. Su preocupación, al igual que la nuestra, es que las máquinas de inteligencia artificial perpetúen estos sesgos y, por tanto, no representen un dispositivo que aporte un progreso en el ámbito moral a la vez que en el científico.

Fotografia extreta de Freepik

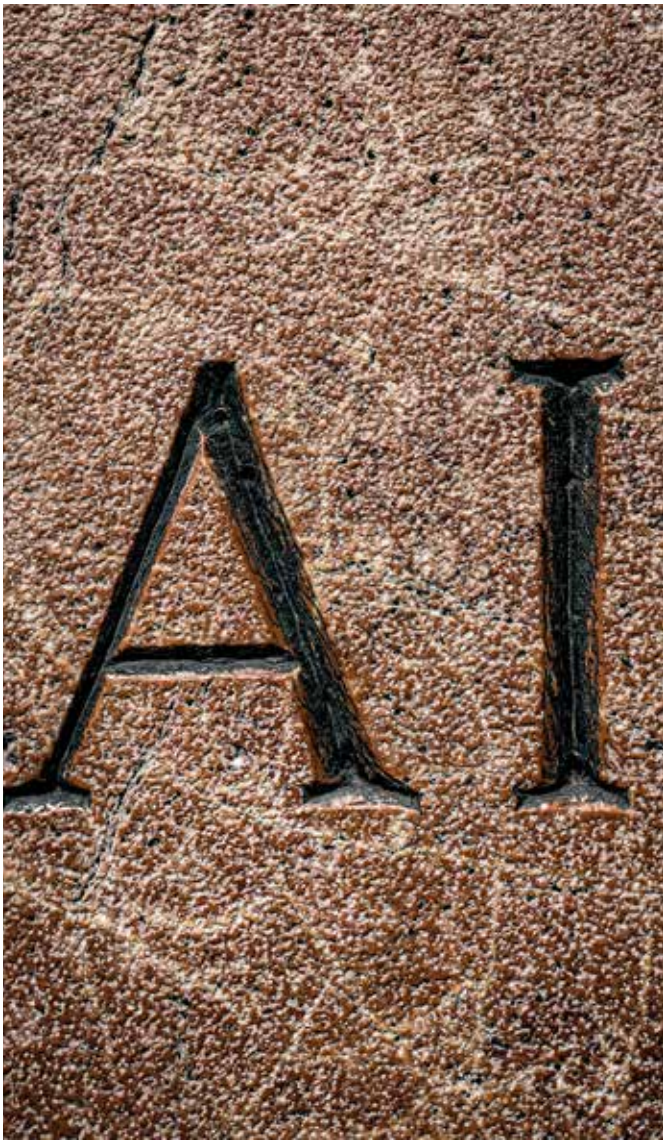
Coeckelberg hace gran hincapié en el problema de los datos a la hora de criticar los sesgos; los datos escogidos pueden ser incompletos y estar mal proporcionados o aplicados a casos donde el grueso de población es tan pequeño que la generalización no funciona. En definitiva, él muestra que se puede estar dando un problema con estos datos a la hora de diseñar y usar las IA, generando un problema clásico, como es el de la correlación sin causalidad. A lo que quiero llegar con ello, es que se debería de estudiar si la supresión de algunos datos podría ser favorable para evitar casos de discriminación. Por ejemplo, algunos datos como el lugar de residencia, si algún familiar ha sido arrestado, etc.

Estos datos pueden parecer inocentes en primera instancia, no obstante, debido a la intervención por parte de la policía en las comunidades de las minorías, puede tener como consecuencia que estos lugares, que luego se tienen en cuenta para decidir el porcentaje de reincidencia de una persona i en los que vive gente de minorías étnicas, puedan tener un gran peso en la decisión. Por lo que se puede dar más importancia al lugar y algunos factores sociales que a la persona en concreto, causando que se juzgue si las personas de una determinada área o familias tienen más probabilidad de reincidir. Esta cuestión puede suponer un gran problema puesto que estos individuos no serán juzgados por su disposición de reincidencia delictiva, sino por factores sociales.

Frente a ello, se abre una nueva cuestión y es la de la responsabilidad. ¿Quién es responsable de este sesgo? ¿Es culpa del algoritmo? ¿De los creadores de esta? ¿De los usuarios? ¿De factores sociales como la alta presencia policial en determinadas áreas? Estos elementos llevan a la necesidad de revisar nuevamente sobre la agencia de la responsabilidad para trazar un nuevo camino para una inteligencia artificial libre de sesgos discriminatorios.

CONCLUSIÓN

Desde un punto de vista general, el uso de los sistemas algorítmicos debería ser reducido a espacios controlados, hasta encontrar una tecnología óptima para el uso público, donde los beneficios sean mayores a las pérdidas. En el caso del COMPAS, sería el de reducir la cantidad en su máxima posibilidad, tanto de los falsos negativos, aquellos que supuestamente no iban a reincidir, pero sí lo hicieron, como de los falsos positivos, es decir, los que se predijo que iban a volver a delinquir pero, sin embargo, no lo hicieron.



Fotografia de Bernd Dittrich en Unsplash

BIBLIOGRAFIA

COECKELBERGH, M. (2021). "El sesgo y el significado de la vida". A: *Ética de la inteligencia artificial*. Madrid: Cátedra. pág. 111-115.

PASCUAL, G.M. (2021). "Los algoritmos que calculan quién va a reincidir discriminan a los negros (y no es fácil corregirlos)". *El País*. [https://elpais.com/tecnologia/2021-11-26/los-algoritmos-que-calculan-quien-va-a-reincidir-discriminan-a-los-negros-y-no-es-facil-corregirlos.html#]. [Fecha de consulta: 12/12/2023]

MÜLLER, V.C. (2020). "Ethics of Artificial Intelligence and Robotics". *The Stanford Encyclopedia of Philosophy.* [https://plato.stanford.edu/entries/ethics-ai/#AIRobo]. [Fecha de consulta: 13/1/24].