

Coded Bias y los futuros desafíos de las tecnologías

Sergio Peinado Cuevas

Resumen—In recent years, the so-called Coded Bias has gained relevance. The design and use of artificial intelligences (AIs) has led to questioning their application and the algorithms that make them up. Studies have shown that many of these technologies present racial and/or sexist biases, which affects the decisions made by these intelligences and the way facial recognition is handled. Furthermore, the information contained in the databases must be considered during the design of these technologies. The documentary *Coded Bias* presents specific cases that demonstrate these biases, such as teacher evaluation and the algorithm used by judges in the United States when determining sentences.

These coded biases have ethical implications and affect the thinking and development of future technologies. On the one hand, the reproduction of historical patterns limits innovation and perpetuates existing inequalities. On the other hand, people unfamiliar with the technology may feel bewildered and distrustful when discovering that the opinions of AIs vary according to ethnicities and genders. When it comes to developing new AI-based technologies, it is crucial that developers actively consider how their decisions will affect different groups of people.

I. INTRODUCCIÓN

En los últimos años el denominado “Coded Bias” o “Sesgo codificado” ha sido un tema que ha ido tomando mayor relevancia, pues bien, el diseño y uso de IAS ha conllevado a que se ponga en tela de juicio el uso que se le da a éstas y el tipo de algoritmos que poseen, ya que, según estudios muchas de éstas nuevas tecnologías poseen un sesgo racial y/o sexista, el cual tiene una incidencia al momento de la toma de decisiones por parte de éstas inteligencias y a la forma en cómo se maneja el reconocimiento facial, además cabe destacar que aquí se tiene en cuenta también las bases de datos y la información que se maneja al momento del diseño de éstas. A lo largo del documental “Coded Bias”, se nos presentan diferentes casos en donde se puede evidenciar lo descrito anteriormente, y esto nos lleva a preguntarnos ¿Cuáles son los desafíos por enfrentar en los próximos años en el diseño de nuevas tecnologías?, en esta oportunidad trataremos de dar respuesta a esta pregunta, y además con base en el documental “Coded Bias”, hablaremos de dos casos puntuales en donde se puede evidenciar el “Coded Bias”, como es el caso de la evaluación docente y el algoritmo que utilizan en EEUU los jueces al momento de determinar una condena.

II. DESARROLLO

II-A. Analisis

El “Coded Bias” ha sido un tema que ha ido tomando mayor relevancia en los últimos años, pues bien, los estudios realizados por Joy Buolamwini demostraron que existe un sesgo racial y/o sexista en diferentes tecnologías, ya sean reconocimiento facial, bases de datos y/o IAS, en donde este

“Coded Bias” provoca que se de una toma de decisión con base en sesgos. A lo largo del documental “Coded Bias”, se nos presenta esto de forma muy puntual, pues bien, para el caso del sistema judicial de Estados Unidos existe el uso de un algoritmo al momento de sentenciar a alguien por un crimen, el cual es conocido como COMPAS y es un sistema de evaluación de riesgo, en donde se tienen en cuenta diferentes variables como edad y/o antecedentes, en donde el algoritmo con base a unas preguntas otorga una calificación que va entre alto y bajo riesgo, y esto termina jugando un papel al momento de determinar si alguien debe ir a prisión de forma inmediata o no.

II-B. Casos reales

Ahora bien, este tipo de métodos ha sido cuestionado, puesto que según un estudio realizado por Karen Hao y publicado en MIT TECHNOLOGY REVIEW, este tipo de algoritmos ha demostrado tener un sesgo racial el cual se ve evidenciado en el estudio que hicieron de los acusados entre 2013 y 2014 para el artículo, mostraba que el 32 % de las personas encarceladas innecesariamente eran negras, mientras que el 11 % eran blancas, y siguiendo con este mismo balance, nos demuestran en este artículo que los liberados eran en un 65 % personas blancas y tan solo un 32 % personas negras, lo cual nos lleva a ver que se da un sesgo, aún cuando la raza no es una pregunta que entra directamente en este algoritmo. Frente al algoritmo Karen Hao (2021), escritora del artículo nos dice lo siguiente: “Se supone que COMPAS ayuda a los jueces a determinar si un acusado debe permanecer en la cárcel o quedar libre mientras espera el juicio. Se entrena con los datos históricos de los acusados para encontrar correlaciones entre factores como la edad y sus antecedentes en el sistema de justicia penal, y si esa persona ha sido arrestada antes. Luego usa esas correlaciones para predecir la probabilidad de que un acusado sea detenido por un otro delito durante el período de espera del juicio”

Entendiendo el funcionamiento del COMPAS, podemos evidenciar que esta programado para determinar un puntaje con base en una serie de variables, lo cual termina arrojando un resultado en donde califican a la persona como bajo o alto riesgo, y esto determina también las medidas para tener en cuenta frente a su proceso. Ahora bien, frente a lo dicho anteriormente y los datos que presentamos, Karen (2021) dice lo siguiente: “De acuerdo con la ley, COMPAS no incluye la raza en el cálculo de su puntuación de riesgo. Sin embargo, en 2016, una investigación de ProPública reveló que, aun así, la herramienta estaba sesgada contra los negros. ProPublica descubrió que, entre los acusados que nunca volvieron a ser arrestados, los negros tenían el doble de probabilidades que

los blancos de acabar en el grupo de alto riesgo de COMPAS” Como podemos evidenciar para este caso, se termina dando un sesgo por parte del algoritmo al momento de tomar decisiones, lo cual, podríamos relacionar directamente con la raza según los resultados que se obtuvieron en la investigación realizada por ProPública.

Otro de los casos que podemos tomar es el sistema de evaluación docente que se maneja en EE. UU., en donde en el documental “Coded Bias”, se nos presenta un algoritmo denominado “Valor agregado”, el cual es utilizado en los sistemas financieros e hipotecarios para identificar a personas con un alto riesgo de pérdida para así ofrecer este tipo de créditos, y ellos obtener un alto rendimiento al saber que estás personas no podrán con las cuotas o pagos determinados. Este tipo de algoritmos, causo bastante polémica, ya que no se concebía bien con base a que factores otorgaba la calificación a los docentes, lo cual provoco que, tras despidos masivos por obtener un bajo puntaje, se hiciera una denuncia ya que no se conocían estos factores que determinaban las calificaciones de ellos.

II-C. La etica de las IAs

Teniendo en cuenta los dos casos mencionados, podemos entrar a hablar de otro tema que surge con esto que se presenta en el “Coded Bias” y es la ética, y esto nos lleva a preguntarnos, ¿Existe un desarrollo ético al momento de crear las IAS?, pues bien, en el documental se nos presenta un caso en donde la IA de Microsoft comparte alrededor de 15 minutos con el “Mundo digital”, y este determina su comportamiento, volviéndola machista y misógina, lo cual nos lleva a preguntarnos ¿Somos conscientes de cómo el entorno puede afectar el desarrollo de una IA que se encuentra en constante aprendizaje?, desde mi perspectiva no, y aunque si bien las IAS se realizan con el objetivo de facilitar tareas, ayudar a las personas en diferentes circunstancias o brindar propiamente información, parece que aún no somos conscientes que la lógica que se maneja al momento de programar sus algoritmos y darles esa “Libertad de aprendizaje”, puede cambiar su comportamiento o fines, o bien, visto de otro modo, podríamos decir que el uso de una tecnología depende de quien la utilice, pero si esta experimenta el “mundo”, como fue el caso de la IA de Microsoft y con base a eso obtiene una visión u comportamiento, podríamos estar viendo serios riesgos.

Otro punto por tocar con el tema anterior estaría dado en como manejan las bases de datos los programadores y administradores al momento de crear este tipo de tecnologías, y examinar, si desde este punto de vista se esta generando o creando un sesgo en los algoritmos que poseen, lo cual es determinante al momento de ver la toma de decisiones que se viene dando por parte de este tipo de tecnologías.

II-D. Desafios en UI/UX

Ahora bien, con base en estos dos casos y en el análisis que hemos realizado frente a la ética de estas tecnologías, podemos entrar a preguntarnos, ¿Cuáles son los desafíos por enfrentar en los próximos años en el diseño de nuevas tecnologías?,

ya que, como hemos podido evidenciar se esta presentando un sesgo racial y/o sexista, entre otros casos, lo cual se debe contrarrestar. Si bien, tenemos en cuenta que hoy en día los IU y UX son diseñados con ayudas de IA se deben tener en cuenta diferentes factores al momento de utilizar esta tecnología para el diseño de los IU y UX, esto visto desde la forma en como se utilizan las bases de datos, y que estás no sean propiamente las generadoras de diferentes sesgos, los cuales resulten discriminatorios, sino que por el contrario permitan tener una amplia gama y diversidad de datos, lo cual resulte favorable para todas las personas que utilicen estas tecnologías. Otro de los puntos a resaltar, sería buscar nuevos patrones o algoritmos que permitan tener un margen de error más bajo, lo cual, lleve y permita una mejor toma de decisiones, en donde se favorezca más y sea claro el por qué de esa decisión.

III. ANALISIS ETICO

Ya que conocemos distintos casos en donde se produjeron sesgos codificados podemos formar un solido pensamiento de como estos errores tuvieron implicaciones eticas y como afectan el pensamiento y desarrollo de las tecnologias futuras.

Por el lado de los pensamientos y percepcion que pueden llegar a tener las personas no conocedoras del entorno tecnologico, al enterarse que esta tiene opiniones variadas hacia distintas etnias y generos. Generaria un gran desconcierto y recelo por las tecnologias que surgiran atravez de todos estos años que nos deparan como sociedad, ya que como usuarios indefinidos, corremos con los riegos reales presentados tanto en el documental, como en este documento y los miles de casos que han sucedido pero no se les da una voz masiva para representarlos.

En cuanto al desarrollo de las nuevas tecnologias en base a IAs, estos sesgos en los algoritmos pueden moldear un futuro de la tecnología de diversas maneras. Por un lado, la reproducción de patrones históricos puede limitar la innovación y perpetuar desigualdades existentes y por el otro los nuevos y viejos desarrolladores deben considerar activamente cómo sus decisiones afectarán a diferentes grupos de personas.

IV. CONCLUSION

El Coded Bias.º “sesgo codificado” ha surgido como un tema crucial en el diseño y uso de inteligencias artificiales (IA). Los estudios y casos presentados en el documental Coded Bias han evidenciado la existencia de sesgos raciales y sexistas en muchos algoritmos, lo que influye en la toma de decisiones y el reconocimiento facial. Este problema plantea serios desafíos éticos y tecnológicos para el futuro, ya que la perpetuación de patrones históricos a través de estos sesgos puede limitar la innovación y ampliar desigualdades.

El conocimiento de estos sesgos genera desconfianza y rechazo entre las personas que no están familiarizadas con la tecnología, incrementando la preocupación sobre el uso justo y ético de las futuras IA. Por lo tanto, es fundamental que los desarrolladores, tanto nuevos como experimentados, consideren cuidadosamente cómo sus decisiones afectarán a diversos

grupos de personas y trabajen activamente para mitigar estos sesgos, garantizando que las tecnologías emergentes sean más justas e inclusivas.

REFERENCIAS

- [1] MIT Technology Review en español, "Caso práctico: probamos por qué un algoritmo judicial justo es imposible", 2019. [Online]. Available: <https://www.technologyreview.es/s/13800/caso-practico-probamos-por-que-un-algoritmo-judicial-justo-es-imposible#:~:text=Se%20conoce%20como%20COMPAS%2C%20y,libre%20mientras%20espera%20el%20juicio>. [Accessed: May 23, 2024].
- [2] Politikon, "Justicia algorítmica: un experimento fallido", 2016. [Online]. Available: <https://politikon.es/2016/04/19/18152/>. [Accessed: May 23, 2024].
- [3] Shalini Kantayya, director, "Coded Bias", 2020, [Film]. Available: Netflix. [Accessed: May 23, 2024].