



Foto: Internet

RACISMO Y ROLES DE GÉNERO, CONDUCTAS PERPETUADAS EN ALGORITMOS DE INTELIGENCIA ARTIFICIAL

Por Daniela Mendoza

El desarrollo de nuevas tecnologías conlleva una gran responsabilidad; la inteligencia artificial se utiliza cada vez más para dar solución a un sinnúmero de problemas, es así, que ha permitido mejorar motores de búsqueda, realizar reconocimiento facial, desarrollar asistentes virtuales, vehículos autónomos, etc. Actualmente algoritmos de inteligencia artificial forman parte del día a día. Se utilizan en redes sociales, en procesos de contratación, detección de enfermedades y se puede decir, que ciertos aspectos, dependen de las predicciones hechas con IA. Ante esto, surge la necesidad de que estos algoritmos estén libres de sesgos, sean éticos e inclusivos. Es de esperar que los algoritmos que se desarrollen reflejen el pensar de sus autores, pues muchas veces están basados en su subjetividad y, por lo tanto, no abarcan ni representan a toda la población que deberían.

La poca representación de mujeres y minorías raciales fortalecen ideas culturales y de género preconcebidas en la sociedad, afectando a ciertos grupos y favoreciendo a otros, por lo que el sesgo podría representar un peligro en procesos de toma de decisiones, volviéndolos excluyentes y discriminatorios.

La implementación de algoritmos de inteligencia artificial no está exenta de errores humanos, a pesar de la creencia de que estos pueden pensar independientemente o aprender solos. Su fuente de conocimiento viene de datos recolectados por seres humanos y de su perspectiva, por lo que la tecnología implementada no es más objetiva que sus desarrolladores.

El sesgo en algoritmos de IA se puede dar por distintos motivos: empieza desde los datos que se recopilan para entrenar al sistema; estos datos expresan la subjetividad del equipo a cargo; se podría utilizar la analogía con un estudiante que aprende basado en un libro determinado; éste aprendería con la visión que tenga el autor del libro; es lo que sucede con los algoritmos de IA, estos no aprenden solos, sino dependen

de la información y bases de datos que se les proporcione, por lo que representarían de cierta manera el pensamiento y la realidad de sus creadores. (Kozyrkov, 2019)

Como se ha mencionado la IA no es imparcial, al igual que el ser humano, presenta un sesgo cognitivo, que condiciona su capacidad de decisión. Este tipo de sesgo incluye estereotipos que existen en la sociedad, predisposición y la percepción propia. El sesgo se puede dar de muchas maneras: por ejemplo, si se ingresan datos de un solo tipo, el algoritmo no reconocería otras entradas, así, si se ingresa únicamente imágenes de teléfonos inteligentes con pantalla táctil y sin botones, el algoritmo detectaría únicamente estos como teléfonos, descartando otros modelos.

Si todos los datos tienen una característica en común, como el tono de piel y se ingresa mayormente imágenes de personas de una determinada raza, se podría dar otro tipo de sesgo en el que el algoritmo tendría problemas para identificar a quienes no compartan estas características raciales. También se podría dar sesgo por selección, pues en el caso de reconocimiento facial es difícil incluir todas las estructuras óseas y características étnicas, por lo que podría limitarse a reconocer únicamente a personas de cierta

región (Sengupta, et.al, 2018).

Algunos ejemplos de sesgo en algoritmos se pueden evidenciar en los algoritmos de visión artificial de Google que identificó a algunas personas de ascendencia africana como gorilas; el sistema fue corregido después de este incidente. Asimismo, se desarrolló sistemas de comando por voz que funcionaban bien con voces masculinas, pero no identificaban voces femeninas. Igualmente, se podían ver roles de género en asistentes virtuales. Aquellos que realizaban funciones básicas como Siri o Alexa tenían voz femenina, pero sistemas más sofisticados para resolución de problemas como Watson de IBM tenían voz masculina (Martin, 2018).

La automatización de ciertos procesos podría representar un peligro cuando los datos que son utilizados para esto están sesgados; si un algoritmo se ha entrenado con un sesgo racial, va a seguir incorporando esto en distintos contextos. Algoritmos para selección de currículos podrían ser un ejemplo de esto. Si el algoritmo se entrena con currículos con nombres que sean usualmente de gente de raza blanca va a descartar automáticamente aplicantes de otras etnias (Fawcett, 2020). En Estados Unidos un algoritmo utilizado para brindar atención médica a pacientes fue acusado de presentar un sesgo racial. Dos senadores alegaron que da

mejor atención médica a pacientes de raza blanca; se pudo ver que pacientes de raza negra en estado más crítico recibían menor atención que pacientes blancos. El algoritmo no consideraba la raza en sí, pero consideraba el nivel socioeconómico y gasto en servicio médico según grupos raciales, definía que las personas con menos ingresos debían gastar menos en servicios médicos, dándoles menor atención (Domínguez, 2019).

Ciertos campos como ciencia, matemáticas, ingeniería e inteligencia artificial, se han visto representados mayoritariamente por hombres. Según estudios el 80% de los profesores de IA son hombres, y personas de color aún tienen una representación muy baja en compañías de tecnología (Fawcett, 2020). La industria de IA está llena de muchos ejemplos basados en estereotipos de género, lo cual se refleja en los resultados de estos algoritmos.

Por ejemplo, el sistema de traducción de Google al traducir de un lenguaje con género neutro como el inglés a un lenguaje con distinción de género como el español, atribuye un género al lenguaje y al hacerlo se refería como masculino; para describir profesiones como ingeniero, doctor o soldado y con género femenino para profesiones como profesora, enfermera o secretaria; de este modo, se puede demostrar que muchas veces la implementación de IA puede ayudar a fortalecer y reforzar ideas erróneas y estereotipos preconcebidos por la sociedad (Parsheera, 2018)

La inequidad en la contratación laboral al emplear IA no se da solamente por temas raciales; un algoritmo de selección de currículos de Amazon mostró un sesgo en contra de la población femenina, el algoritmo de Amazon se estuvo desarrollando desde 2014, para automatizar el proceso de contratación y encontrar los mejores talentos; se calificaba los currículos de 1 a 5.

En 2015, el equipo se dio cuenta que para puestos como desarrollador de software y otras plazas técnicas no se estaba siendo imparcial

respecto al género. El modelo fue entrenado para detectar patrones en currículos de quienes habían sido contratados en los pasados 10 años.

La mayoría de estos currículos eran de hombres, demostrando como el campo de la tecnología ha sido por años dominado por hombres. El sistema de Amazon basado en la información que tenía se entrenó para preferir perfiles masculinos, penalizando los currículos que tenían la palabra “mujeres” en ellos. Un currículo que contenía, por ejemplo: “capitana del club de ajedrez de mujeres”, o nombres de universidades con la palabra “mujeres” eran penalizados. El algoritmo fue arreglado, pero esto no es garantía que no se discrimine a las personas por otros términos en sus currículos (Dastin, 2018).

La extensión del uso de Inteligencia Artificial crea la necesidad de diversificar la fuerza laboral encargada de desarrollar esta tecnología, aunque en la actualidad se realizan esfuerzos por contratar más mujeres y minorías raciales en empresas de tecnología. La industria de IA aún está predominada por hombres de raza blanca, llevando esto a descartar currículos femeninos como ya se mencionó o confundir mujeres negras con hombres.

La elección de las bases de datos aún no representa correctamente grupos minoritarios, se necesita más diversidad en las bases de datos, y aún más importante en los equipos que desarrollan la tecnología, involucrando personas con distintas perspectivas e ideas (Shellenbarger, 2019).

Ante lo expuesto se puede llegar a varias conclusiones: empezando en que los errores y sesgos que se podrían presentar en los algoritmos tienen su base en las ideas humanas y en los modelos sociales y prejuicios creados. Los algoritmos reflejan el pensar de quienes los desarrollaron y de su entorno. Los sesgos raciales y de género presentes en IA representan cómo funciona la sociedad en general y como hay

áreas aún dominadas solo por ciertos grupos. Para minimizar sesgos en las aplicaciones de la inteligencia artificial, se debe empezar por minimizarlos en la sociedad y adoptar conductas más inclusivas para dejar de perpetuar ideas erróneas que se han mantenido.

La forma en la que se desarrolla los algoritmos determina su utilidad y aporte a la sociedad, así como pueden ser herramientas muy útiles que mejoren procesos e incluso los vuelvan más justos, también pueden sacar a la luz los problemas de la sociedad y agravarlos. Se debe considerar no solo las capacidades intelectuales del equipo a cargo de desarrollar los algoritmos, sino también su ética.

Al estar la inteligencia artificial involucrada en cada vez más ámbitos, es imperativo utilizar un enfoque ético que evite la exclusión o discriminación y favorezca a todos por igual. Es imposible que exista una persona sin una opinión subjetiva, pues las ideas se forman a partir de la realidad vivida.

Sin embargo, definitivamente se puede mejorar la manera en la que se ha venido desarrollando los algoritmos de IA, al diversificar los equipos de trabajo y los datos utilizados, para que de esta manera se logren reducir errores y sesgos.

Referencias bibliográficas

- Dastin, J. (2018). REUTERS. Obtenido de <https://www.reuters.com/article/usamazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>
- Domínguez, D. (2019). Economía Digital. Obtenido de https://www.economiadigital.es/tecnologia-y-tendencias/inteligencia-artificial-el-sesgo-racial-ya-es-un-problema-politico_20016359_102.html
- Fawcett, A. (2020). Educative. Obtenido de <https://www.educative.io/blog/racial-bias-machine-learning-algorithms#wrap-up>
- Kozyrkov, C. (2019). Medium. Obtenido de CIENCIA&DATOS: <https://medium.com/datos-y-ciencia/la-verdad-sobre-el-sesgo-en-inteligencia-artificial5e228be3aee7>
- Martin, M. (2018). VIVIDCOMM. Obtenido de Advanced Technology in Action: <https://vividcomm.com/2018/04/23/artificial-intelligence-partiality/#:~:text=Like%20the%20human%20brain%2C%20artificial,reasoning%2C%20resulting%20in%20reasoning%20errors.&text=AI's%20learning%20is%20shaped%20by,experience%20through%20interactions%20and%2>
- Parsheera, S. (2018). A GENDERED PERSPECTIVE ON ARTIFICIAL INTELLIGENCE.
- Sengupta, E., Garg, D., Choudhury, T., & Aggarwal, A. (2018). Techniques to Eliminate Human Bias in Machine Learning. Noida.
- Shellenbarger, S. (2019). THE WALL STREET JOURNAL. Obtenido de <https://www.wsj.com/articles/a-crucial-step-for-avoiding-ai-disasters-11550069865>