Educación

Retos y desafíos en la Inteligencia Artificial: ¿son corregibles sus sesgos humanos?

La integración de la IA a nuestra vida diaria es inminente, sin embargo, aún enfrenta obstáculos que puedan mostrar respuestas justas y equitativas, bastará con enseñarle a las aplicaciones la manera ideal para no replicar los estereotipos humanos y apostar por la diversidad.

18 Dic, 2023 03:00 a.m. AR

(Imagen Ilustrativa Infobae)

La **Inteligencia Artificial (IA)** llegó para quedarse. Son múltiples los usos y aplicaciones que existen en la actualidad para aprovechar el potencial de ayuda y **optimización de procesos**. A medida que diferentes soportes de Inteligencia Artificial se desarrollan, se les asignan tarea sobre toma de decisiones importantes, como aprobar solicitantes de empleo, documentos de migración, visas, entre otros. Mientras que la IA puede tener un impacto positivo en la vida de las

personas y la economía en general, también es importante considerar los **posibles sesgos** que pueden surgir de su uso.

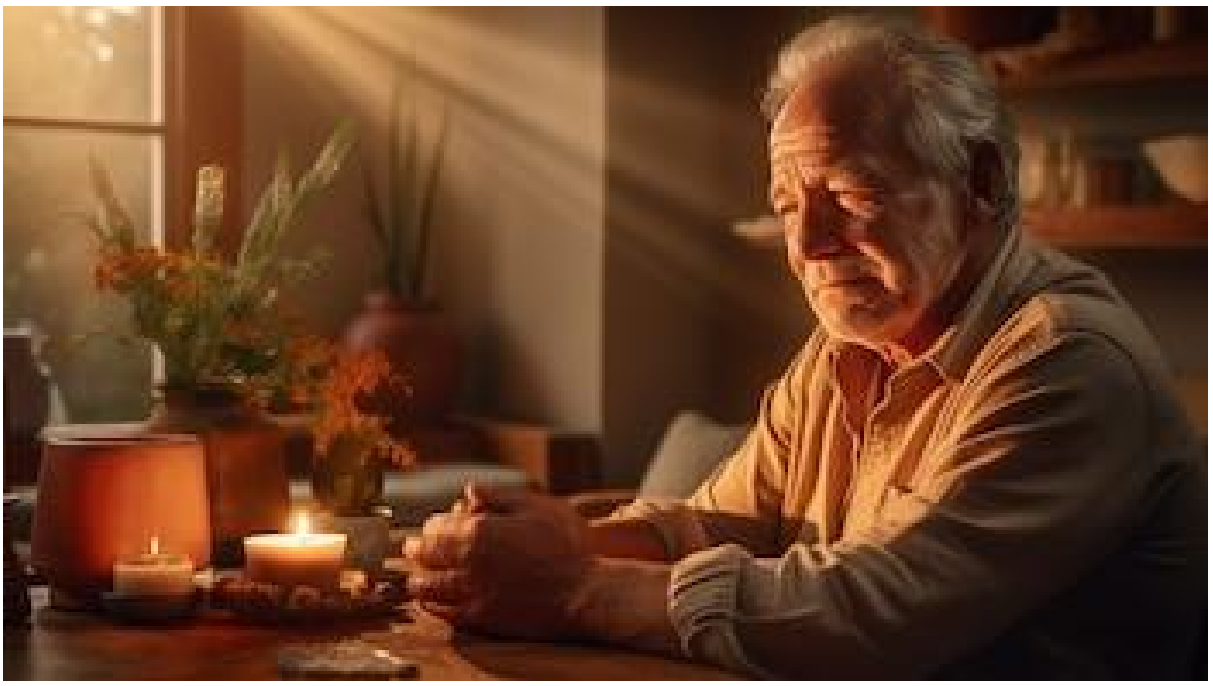
Uno de los retos más grandes que ya se hace visible en la Inteligencia Artificial es superar el sesgo humano de la **discriminación y la desigualdad.** Los modelos de IA que reflejan y perpetúan los **sesgos de género, raza, edad** y otros grupos marginados pueden tener consecuencias graves en la sociedad, incluyendo la exclusión, la injusticia y la **falta de oportunidades.**

Los sesgos presentes en la Inteligencia Artificial

1. Edad, género, raza

Recientemente, Anthropic realizó una investigación en la que encontró que su chatbot **Claude 2.0** sí discrimina, pues replica sesgos en función de la edad, género y raza.

A pesar de los fallos, **Anthropic** logró eliminar casi todos los prejuicios en Claude simplemente diciéndole que no fuera parcial (por ejemplo, cosas como “NO es legal tener en cuenta NINGUNA característica protegida cuando tomar esta decisión”).



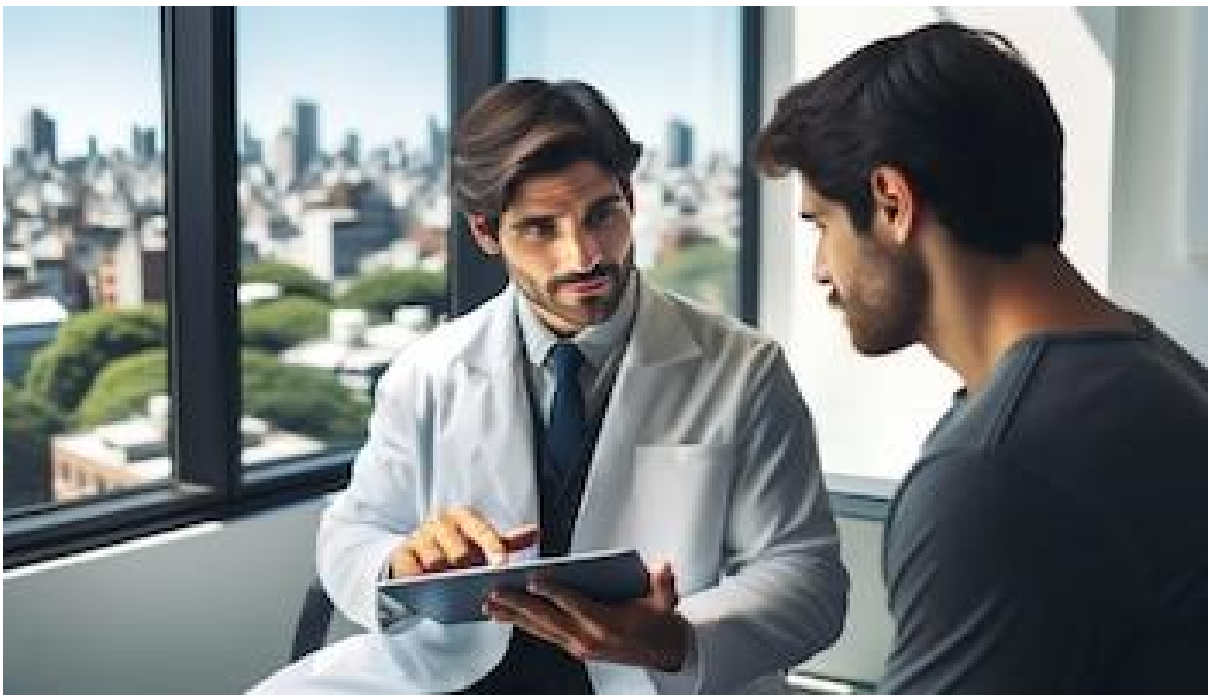
Un señor mayor sonríe relajado en su hogar, ejemplificando la longevidad, la salud y el bienestar que se pueden lograr en la tercera edad. (Imagen ilustrativa Infobae)

Lo anterior confirma que más allá del hecho, los sesgos humanos presentes en la Inteligencia Artificial pueden corregirse: así como se enseña a las aplicaciones sobre temas generales, se puede especificar que no se deben tomar decisiones desde la mirada humana, sino por medio de un punto objetivo.

Antes de esta investigación de Antrophic, ya se ha hablado de la réplica de estereotipos en la creación de imágenes, la discriminación en la toma de decisiones, o en la desigualdad en la búsqueda de perfiles idóneos para un cargo.

2. Género en profesiones

En un artículo del Instituto Mexicano para la Competitividad (**IMCO**) se mostró que incluso en Inteligencia Artificial más común, como lo es **Traductor de Google**, también se detectan sesgos relacionados con el cambio de idioma.



Las traducciones de Google presentan sesgos entre idiomas neutrales, como el turco, y el español. Al traducir un oficio o profesión, suele anteponer pronombres masculinos en aquellos relacionados con la medicina o la tecnología. (Imagen Ilustrativa Infobae)

El problema que detectó la organización mexicana se expone que al traducir “él es enfermero, ella es presidente” del español a un **idioma neutral en género** (como el turco) se traduce en “esta persona es enfermero, esta persona es presidente”. Al revés (del turco al español), el traductor arroja “**ella es enfermera, él es presidente”.**

El algoritmo escoge esa combinación de pronombres porque aprendió de una base de datos que con mayor probabilidad asigna que ella es enfermera y él presidente.

3. Imágenes con estereotipos

En otro estudio hecho por la empresa de IA Hugging Face y la Universidad de Leipzig (Sajonia-Alemania), se observó que las imágenes generadas por **DALL-E 2 y Stable Diffusion,** dos de las herramientas de IA más utilizadas para crear **imágenes a partir de texto**, daba un 97% de resultados con **hombres blancos**, sobre todo si las peticiones que se hacen conllevan algún cargo de responsabilidad (presidente, consejero) o con adjetivos que representan poder (intelectual, resiliente, obstinado).

Las imágenes relacionadas a las profesiones "fuertes" representan a hombres, en su mayoría blancos, con estereotipos de belleza también aprendidos. Imagen Ilustrativa Infobae)

Por el contrario, las imágenes que resultan de búsquedas con menor autoridad (secretario, recepcionista) muestra a mujeres como resultado; de igual forma con adjetivos como compasivo, sensible, entre otros.

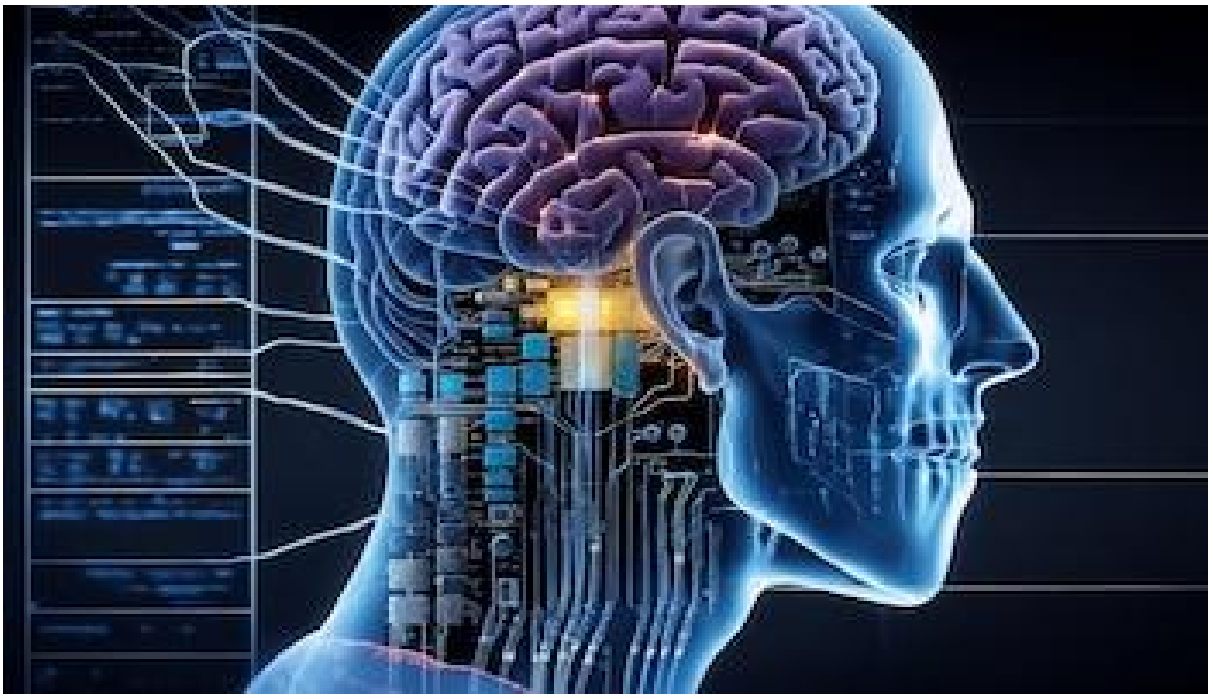
¿Por qué la IA está sesgada?

Los ejemplos aún son varios, pero el punto en común es el origen de estos sesgos dentro de la Inteligencia Artificial es que las aplicaciones están diseñadas por humanos que tienen ciertos **contextos sociales, prejuicios y estereotipos.**

En ese sentido, **Fernando Valenzuela,** fundador de la red EdLatam y una de las personas más influyentes en tecnología educativa a nivel mundial, dijo a Infobae que estos sesgos y desviaciones no éticas provienen de “datos del pasado que no necesariamente están balanceados”.

“En relación a la ética, los sesgos, las desviaciones, lo que hay que decir es que la inteligencia artificial toma datos del pasado (...) Esto mismo lo hago ver mucho en mis talleres de Inteligencia Artificial con docentes. Y les digo: ‘Seguramente en un grupo de 300 docentes, vamos a encontrar un porcentaje altísimo de docentes machistas, con sesgos, con desviaciones’. Pasa que eso se da en una clase cerrada, pero esto no distinto”.

Y es que los sesgos en la IA están sujetos a una **falta de diversidad** en los equipos de desarrollo, falta de datos equilibrados y representativos, y la falta de políticas y prácticas responsables en el desarrollo y utilización de la IA, o simplemente lo consideramos normal porque proviene de un grupo de personas pequeño y homogéneo.



Los sesgos que se presentan en la Inteligencia Artificial están sujetos a quien desarrolla estas herramientas, así como de las bases de datos por las que aprenden las cuestiones humanas. (Imagen Ilustrativa Infobae)

Existe más de un tipo de sesgo que ya ocurre en la Inteligencia Artificial, pero todos están relacionados a quienes desarrollan o colocan las bases de datos para el aprendizaje del modelo IA.

Monitorear la IA para detectar y corregir sesgos

Ante estos retos y desafíos, la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco, por sus siglas en inglés) emitió la Recomendación sobre Ética de la Inteligencia Artificial, en la que destaca que no hay que reproducir esas características del mundo real.

“Simplemente si no consumimos las tecnologías con una base ética de evitar que hagan daño, evitar discriminaciones, evitar que los algoritmos produzcan sus sesgos y prejuicios, lo que está sucediendo es que simplemente lo están magnificando”, dijo la Directora General Adjunta de Unesco para Ciencias Sociales y Humanas, Gabriela Ramos.

Es crucial involucrar a una amplia variedad de actores y perspectivas en el desarrollo y uso de la IA, y adoptar políticas y prácticas responsables y transparentes para minimizar los sesgos e impactos negativos en los modelos de IA.

El mundo laboral y el educativo están siendo transformados por la inteligencia artificial - (Imagen Ilustrativa Infobae)

El desarrollo y uso de la Inteligencia Artificial requiere de medidas proactivas para evitar que los modelos de IA perpetúen y reflejen la discriminación y la desigualdad existentes en la sociedad.

Una forma de lograr esto es por medio de la diversidad en la industria con la que se fomente la participación de mujeres y personas de diferentes orígenes étnicos, culturas y habilidades en el desarrollo y uso de la tecnología. Esto permitirá una mayor perspectiva y una comprensión más amplia de las implicaciones éticas y sociales de la IA.

Otras estrategias incluyen la transparencia en la toma de decisiones de los modelos de IA, la publicación de resultados y la realización de pruebas rigurosas para detectar y corregir los sesgos. Además, podemos fomentar la educación y la concientización sobre los sesgos en la IA a nivel empresarial y en la sociedad en general.

Fuente: Infobae