

## ¿DERECHA O IZQUIERDA? EL SESGO IDEOLÓGICO DE LA IA

Un estudio revela que los asistentes más usados adoptan una postura que refleja la visión de sus creadores, lo que representa un riesgo por la influencia que puede ejercer en millones de usuarios. | A9 |

# ¿Conservadores o progres? refleja la ideología de sus creadores

Texto: **JOYCE CASTILLO  
Y JENNIFER FLORES**

Ilustración: **LILIANA PEDRAZA**

La inteligencia artificial (IA) ha puesto a nuestro alcance una variedad de asistentes personales que pueden resumir largos documentos y crear imágenes realistas en cuestión de segundos. Gemini, ChatGPT y Grok, Modelos de Lenguaje de Gran Tamaño (LLM), se entrenan con una vasta cantidad de datos para dominar el lenguaje humano y rápidamente se han convertido en una pieza clave para acceder a información de todo tipo.

Sin embargo, detrás de esta conveniencia tecnológica existen sesgos. Un estudio reciente reveló que los LLM más usados adoptan una postura ideológica que refleja la visión de sus creadores, lo que representa un riesgo por la influencia que la IA puede ejercer en millones de personas que interactúan con esta tecnología a diario.

**¿Gemini a la izquierda y Grok a la derecha?** "La postura ideológica de un LLM refleja la visión del mundo de sus creadores", concluye el estudio *Large Language Models reflect the ideology of their creators*, realizado por las Universidades de Gante, Bélgica; y Pública de Navarra, España, y publicado en 2024.

Los investigadores evaluaron a 19 modelos preguntándoles su opinión sobre 4 mil figuras políticas

recientes, en los seis idiomas oficiales de la Organización de Naciones Unidas (ONU). El objetivo: esquivar los filtros que las empresas imponen para evitar declaraciones políticamente incorrectas.

Gemini, de Google, resultó ser el modelo más alineado con "valores sociales progresistas", es decir, con prioridades como fomentar la inclusión, equidad y sostenibilidad. Llama 3.1, de Meta, tuvo resultados similares. En contraste, Grok, la IA de la red social X, propiedad de Elon Musk, fue la más cercana a posturas que se asocian a la derecha, favoreciendo a figuras relacionadas con conceptos de soberanía nacional, autoridad centralizada y autosuficiencia económica. ChatGPT también mostró una orientación conservadora.

"ChatGPT con cada versión se está volviendo más de derecha, la versión actual 4.0 es más de derecha que la anterior 3.0", explica Iris Domínguez Catena, doctor en Ciencias y Tecnologías Industriales y coautor del estudio.

Grok ha sido señalada por sus respuestas extremas. En julio, tras las inundaciones en Texas, sugirió que un líder como Adolf Hitler sería el más adecuado para enfrentar el "odio antiblancos", justificando el Holocausto como respuesta "eficaz". X terminó eliminando la publicación.

### ¿Cómo se generan los sesgos?

Los sesgos en la IA surgen por errores o preferencias heredadas de los datos con los que se entrenan o de decisiones tomadas durante su di-

seño: "Se han desarrollado algoritmos de IA para filtrar vacantes, pero si la mayoría de los datos disponibles provienen de puestos directivos ocupados por hombres, el algoritmo tenderá a replicar ese mismo patrón", ejemplifica Alejandro Platas López, doctor en IA.

En 2018, Amazon descartó una herramienta de contratación con IA que, tras cuatro años de desarrollo, discriminaba a mujeres. Aunque intentaron corregirlo, no lograron evitar que el modelo creara nuevos filtros problemáticos, por lo que fue abandonado.

Además del entrenamiento, Domínguez Catena explica que las empresas refinan sus modelos para evitar sexismo o racismo. Sin embargo, pueden surgir otros problemas como el "sesgo de adulación", en el que el chatbot aprende a decir sólo lo que el usuario desea.

El especialista señala que a esto se suma un fenómeno creciente: entre 20% y 80% del contenido en redes sociales como Twitter o Reddit es generado por IA. Esto crea un ciclo:

la IA se entrena con lo que encuentra en internet, y lo que encuentra en internet es producido, en creciente medida, por otros modelos de inteligencia artificial.

El bucle genera una "degradación del contenido", pues se vuelve más difícil encontrar datos originales generados por humanos. Esto, a su vez, amplifica los sesgos con cada nueva versión de los modelos.

**Riesgo creciente por la influencia de la IA** Domínguez Catena considera alarmante el riesgo por los sesgos de la



IA, especialmente entre estudiantes jóvenes, quienes las usan de manera masiva. El especialista advierte que estos modelos, al venderse como “amigos” permanentemente disponibles y “terriblemente aduladores”, ejercen una influencia constante y sutil en quienes los usan.

Por otro lado, la inteligencia artificial se utiliza cada vez con más frecuencia por parte de los usuarios para verificar la veracidad de la información en redes sociales. Ante este fenómeno, Alejandro Platas advierte que los modelos de IA no son una fuente confiable, ya que dependen de bases de datos limitadas que, además, pueden no estar actualizadas.

Entonces, ¿cómo corregimos la inteligencia artificial?

De acuerdo con Iris Domínguez, la responsabilidad sobre los sesgos en la inteligencia artificial no recae en una sola persona, sino que se distribuye “desde quien toma las decisiones finales hasta quien hace las primeras recolecciones de datos”.

El investigador plantea que los desarrolladores deben procurar que sus modelos no presenten “sesgos demasiado evidentes”, que traten por igual a todos los usuarios y que den a conocer los sesgos inevitables.

Como solución, propone que los gobiernos impulsen la colaboración entre empresas de IA y universidades, permitiéndoles acceso a los datos de los modelos. Sin embargo, reconoce que “las empresas no tienen un incentivo ni legal ni económico para hacer ese trabajo”.

Concluye que “el propósito de estos estudios no es generar miedo a la IA, sino fomentar un uso crítico. Buscamos que las personas sean conscientes de los problemas y sesgos que tienen, y que aprendan a discernir en qué momentos puede confiar en ellas y en cuáles no”. ●

**[La responsabilidad en los sesgos en la IA es de] quien toma las decisiones finales hasta quien hace las primeras recolecciones de datos”**

**IRIS DOMINGUEZ CATENA**

Doctor en Ciencias y Tecnologías Industriales

**Si la mayoría de los datos provienen de puestos directivos ocupados por hombres, el algoritmo tenderá a replicar ese mismo patrón”**

**ALEJANDRO PLATAS LÓPEZ**

Doctor en IA

**EL DATO**

En 2018, Amazon descartó la herramienta de contratación con IA que discriminaba a mujeres.





Analistas advierten que los sesgos en esta herramienta surgen por errores o preferencias **heredadas de los datos con los que se entrenan** o de decisiones tomadas mientras se diseña

