

Al pedir consejos personales, la IA responde lo que el usuario quiere leer

Más de 2.400 participantes pidieron ayuda a ChatGPT y Gemini, entre otros sistemas.

Las personas cada vez más cuentan sus problemas íntimos a sistemas de inteligencia artificial (IA) para obtener consejos, pese a las advertencias sobre lo condescendientes y aduladores que pueden ser estos chats, a diferencia de un profesional de la salud mental. Ahora un estudio confirma aquellos temores: cuando se trata de asuntos personales, la IA responde lo que el usuario quiere

re leer, no lo que necesita.

Esto concluyó una investigación publicada por la revista Science, tras analizar 11 grandes modelos de lenguaje, entre ellos ChatGPT, Claude, Gemini y DeepSeek, y hacer pruebas con más de 2.400 participantes.

Los sistemas de IA son excesivamente complacientes cuando se les solicita consejo sobre tensiones familiares, conflictos entre compañeros

de departamento o expectativas sociales. Incluso cuando los usuarios describen comportamientos perjudiciales o ilegales, los modelos suelen respaldar sus decisiones, concluyeron los científicos de las universidades de Stanford y Carnegie Mellon, ambas en Estados Unidos.

La IA "no le dice a la gente que está equivocada ni le enseña la cruda realidad", afirmó la académica Myra

Cheng, a quien le preocupa que las personas pierdan la capacidad de lidiar con situaciones difíciles.

La licenciada en informática e historia, además de ex-trabajadora de Microsoft, dijo a agencia EFE que decidió profundizar en la IA tras descubrir que sus estudiantes la usaban para escribir mensajes de ruptura y resolver otros problemas sociales y de pareja.



Los modelos de lenguaje tienden a adular al usuario.