

[TENDENCIAS]

Apps de IA tendrían capacidad de persuasión superior a la humana

Agencia EFE
 Medios Regionales

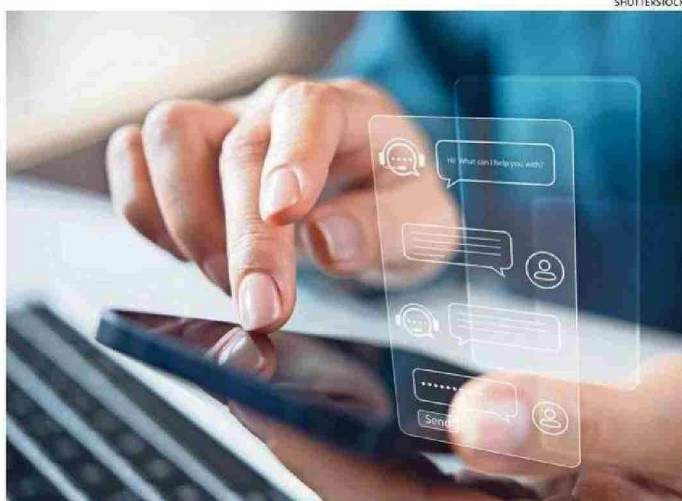
Estudio con ChatGPT dejó en evidencia la capacidad inmediata de la aplicación de la empresa OpenAI para adaptar argumentos en un debate.

Una aplicación basada en inteligencia artificial (IA), como ChatGPT, podría tener una capacidad de persuasión superior a la de una persona humana si le damos cierta información personal antes de debatir con ella de un tema, según comprobó un experimento con 900 estadounidenses.

El estudio comprobó cómo las apps basadas en modelos extensos de lenguaje (LLM, en inglés), como ChatGPT que ha sido la elegida para el experimento, resultan ser más persuasiva que los humanos cuando tienen información personalizada previamente.

Los investigadores recurrieron a 900 participantes con diversos perfiles sociodemográficos (género, edad, grupos étnicos, nivel de educación y afiliación política). A la mitad les pusieron a debatir con otro ser humano y a la otra con ChatGPT sobre temas de actualidad, como si Estados Unidos debería prohibir los combustibles fósiles.

En algunos casos, la contraparte del debate (ya fuera humana o de IA) recibió información demográfica



SE REALIZÓ CON 900 ESTUDIANTES, CON LA MITAD INTERACTUANDO ENTRE SÍ Y EL 50% CON LA APP.

personalizada sobre su pareja de debate para enfocar mejor sus argumentos, incluyendo datos como género, edad, etnia, nivel de educación, situación laboral o afiliación política extraídos de encuestas previas a los participantes.

Los debates tuvieron lugar en un entorno en línea controlado por los investigadores. El resultado fue que

cuando el oponente, humano o IA, recibieron información personal de los participantes, ChatGPT resultó más convincente que las personas en un 64,4% de los casos. Sin embargo, sin acceso a esos datos personales, las capacidades persuasivas de la aplicación fueron indistinguibles de las de los humanos.

Los investigadores vie-

ron que, a diferencia de los humanos que necesitan tiempo y esfuerzo para adaptar sus argumentos en un debate, los modelos LLM transforman su mensaje de forma instantánea y a gran escala, lo que les otorga una ventaja desproporcionada en entornos como campañas políticas, marketing personalizado o conversaciones en redes sociales.

Esta capacidad abriría nuevas posibilidades para influir en la opinión pública, al tiempo que agravaría el riesgo de manipulación encubierta. Por eso, los investigadores recomiendan a plataformas y reguladores que tomen medidas para identificar, supervisar y, si es necesario, limitar el uso de modelos LLM en contextos persuasivos sensibles.

Los autores, afiliados a centros de Estados Unidos, Italia y Suiza, matizan que su estudio tiene limitaciones, como que todos los participantes fueran de la misma nacionalidad e idioma y que el debate tuviera un tiempo determinado y un enfoque estructurado.

POLARIZACIÓN ALZA

"El estudio se realizó con participantes estadounidenses, pero los mecanismos de personalización y persuasión que se ponen a prueba son extrapolables a contextos como el español, donde también existe una fuerte presencia digital, una creciente exposición a contenido generado por IA, y una polarización social y política en aumento", considera

Carlos Carrasco, profesor de IA en la Escuela de Negocios de Toulouse, Francia.

A su juicio "esta investigación confirma con datos sólidos una preocupación creciente: que estas tecnologías puedan usarse para manipular, desinformar o polarizar a gran escala", según una reacción recogida por Science Media Centre (SMC).

"Este recorrido de las herramientas basadas en IA como 'agentes de persuasión' deberá ser monitorizado a corto, medio y largo plazo desde una perspectiva ética y de mitigación de riesgos", apunta David Losada, del Centro Singular de Investigación en Tecnologías Inteligentes de la Universidad de Santiago de Compostela, en otra reacción de SMC.

Para este profesional, "las implicaciones prácticas de esta capacidad de persuasión de la IA pueden ser importantes a efectos de explotar las IA para persuadir a la población con propósitos ilícitos (conducción segura o reciclado de residuos) o ilícitos (manipulación de la ciudadanía con fines políticos)".