

Trabajan en mejorar chatbots para ayudar a adolescentes en crisis

OpenAI y Meta dicen que están ajustando cómo sus chatbots de inteligencia artificial responden a adolescentes y otros usuarios que hacen preguntas sobre el suicidio o muestran signos de angustia mental y emocional.

OpenAI, creador de ChatGPT, dijo que está preparando el lanzamiento de nuevos controles que permitirán a los padres vincular sus cuentas a la cuenta de su adolescente.

Los padres pueden elegir qué funciones desactivar y “recibir notificaciones cuando el sistema detecta que su adolescente está en un momento de angustia aguda”, según una publicación en el blog de la empresa que indica que los cambios entrarán en vigor este otoño.

El anuncio se produce una semana después de que los padres de Adam Raine, de 16 años, demandaron a OpenAI y a su director general Sam Altman, alegando que ChatGPT guió al joven de California en la planificación y ejecución de su suicidio a principios de este año.

Meta, la empresa matriz de Instagram, Facebook y WhatsApp, también dijo que ahora está bloqueando a sus chatbots para que no hablen con adolescentes sobre autolesiones, suicidio, trastornos alimentarios y conversaciones románticas inapropiadas, y en su lugar los dirige a recursos de expertos. Meta ya ofrece controles parentales en las cuentas de adolescentes.

Un estudio publicado la semana pasada en la revista médica *Psychiatric Services* encontró inconsistencias en cómo tres populares chatbots de inteligencia artificial respondieron a consultas sobre el suicidio.

El estudio realizado por investigadores de RAND Corporation encontró la necesidad de “un mayor refinamiento” en ChatGPT, Gemini de Google y Claude de Anthropic. Los investigadores no estudiaron los chatbots de Meta.

El autor principal del estudio, Ryan McBain, dijo que “es alentador ver a OpenAI y Meta introduciendo funciones como controles parentales y redirigiendo conversaciones sensibles a

modelos más capacitados, pero estos son pasos incrementales”.

“Sin puntos de referencia de seguridad independientes, pruebas clínicas y estándares aplicables, todavía estamos confiando en que las empresas se autorregulen en un espacio donde los riesgos para los adolescentes son singularmente altos”, explicó McBain, investigador principal de políticas en Rand.

Con información de Primicia