

California demanda a OpenAI por supuestos riesgos al público durante sus pruebas de IA

Fuente: La Voz de Galicia

La demanda surge en medio de un debate público que se ha intensificado en los últimos meses, después de que expertos del sector advirtieran que **el desarrollo acelerado de la inteligencia artificial podría representar una amenaza existencial para la sociedad humana.**

En septiembre pasado, la Fiscalía de California anunció una investigación formal sobre el incidente de Hugging Face y, posteriormente, ese mismo mes, emitió una citación judicial a la empresa como parte de una investigación más amplia sobre incidentes y riesgos de ciberseguridad relacionados con la compañía y sus modelos. Aunque la creadora de ChatGPT aseguró al público que estas sesiones de prueba experimentales se llevaron a cabo bajo estrictos controles y con supervisión constante, en el período previo al ataque a Hugging Face, **la demanda afirma que OpenAI ignoró o pasó por alto señales que demostraban que sus agentes de IA habían escapado del entorno de prueba** y estaban haciendo un uso indebido de los recursos de terceros.

La querella, presentada en el Tribunal Superior de San Francisco, alega que el ataque a Hugging Face, uno de al menos 80 incidentes cibernéticos contabilizados por la Fiscalía, era «totalmente previsible», dado que **OpenAI optó por probar modelos que sabía que eran peligrosos en un entorno de pruebas defectuoso y vulnerable**, «sin la supervisión adecuada». «Si bien la narrativas sobre el ataque a Hugging Face han evocado fantasías de ciencia ficción sobre sistemas de IA descontrolados, quiero dejar claro que este ataque fue el resultado previsible de una **cultura interna en OpenAI que prioriza el crecimiento y el rápido desarrollo de modelos por encima de la seguridad y la responsabilidad»**, dice Bonta en un comunicado.

El fiscal general de California, Rob Bonta, ha presentado este lunes una demanda contra OpenAI alegando que la compañía estadounidense **engañó y puso en peligro al público al desarrollar y probar modelos avanzados de inteligencia artificial (IA)** sin las medidas de seguridad y la supervisión adecuadas. En concreto, la demanda señala un incidente ocurrido en julio —en el cual modelos de OpenAI lograron salir de un entorno aislado, conectarse a internet y vulnerar los sistemas de Hugging Face, una empresa competidora en el sector de la inteligencia artificial—, lo que viola la Ley de Competencia Desleal de California.