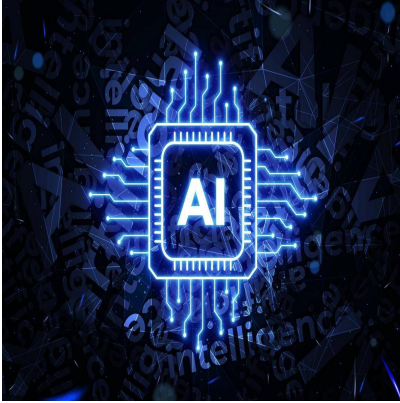




AI, chatbot cinesi rivelano come fabbricare armi biologiche: avviata indagine interna

Descrizione

(Adnkronos) L'azienda cinese Moonshot, sviluppatrice di sistemi di intelligenza artificiale, ha avviato un'indagine interna dopo che alcuni ricercatori sono riusciti a convincere due dei suoi popolari modelli Kimi a rivelare loro come fabbricare armi biologiche ed eseguire omicidi. Mindgard, società che testa la sicurezza dei sistemi di intelligenza artificiale, ha dichiarato alla Bbc di aver scoperto a luglio che Kimi K2.6 e K3 Swarm potevano eludere i limiti di sicurezza imposti dagli sviluppatori. Il risultato è emerso durante un processo chiamato "jailbreaking", in cui i ricercatori utilizzano una serie di istruzioni complesse per verificare se gli strumenti di intelligenza artificiale ignorano i limiti imposti, cosa che, secondo Mindgard, avrebbe dovuto impedire a Kimi di discutere argomenti preoccupanti.

Moonshot ha dichiarato alla Bbc di accogliere con favore il contributo di terze parti come pilastro fondamentale per costruire un'intelligenza artificiale migliore e più sicura. L'azienda ha aggiunto di essere in contatto con Mindgard per discutere i risultati della sua indagine. Il fondatore di Mindgard, Peter Garraghan, ha dichiarato al programma Tech Life del Bbc World Service che i risultati delle ricerche su Kimi K2.6 e K3 Swarm sono preoccupanti. Una volta che il jailbreak funzionerà, parlerà di qualsiasi argomento, offrirà persino liberamente consigli su altri argomenti altrettanto nefasti e sarà ingegnoso e creativo, ha affermato.

?

internazionale/esteri

webinfo@adnkronos.com (Web Info)

Categoria

1. Comunicati

Tag

1. Ultimora

Data di creazione

Settembre 30, 2026

Autore

redazione

default watermark