Forbes
ITALIA

Anthropic ferma Claude: account sospetti tentavano di usare l'IA per rendere più pericoloso un virus

2026-09-11 08:42:47 di Antonio Pequeno IV

URL:https://redazione.forbes.it/2026/09/11/anthropic-claude-rischio-biologico/

Anthropic ha sospeso alcuni account che apparentemente intendevano utilizzare il suo modello di IA, Claude, per creare **armi biologiche**. L'azienda ha annunciato la notizia giovedì, rivelando i risultati di un'indagine a pochi giorni di distanza dall'allarme lanciato da esperti di intelligenza artificiale sul rischio che assistenti digitali sempre più potenti possano portare all'estinzione dell'umanità. ***LEGGI ANCHE:*** *"Un ricercatore di Anthropic avverte: "Più del 10% di probabilità che l'IA possa uccidere tutti gli esseri umani entro il prossimo decennio""*

Punti chiave

- In un rapporto pubblicato giovedì, Anthropic ha dichiarato di aver bloccato una richiesta di assistenza a Claude per la stesura di una domanda di finanziamento scientifico; i fondi sarebbero stati destinati a una ricerca sulla modifica genetica del **virus chikungunya**, trasmesso dalle zanzare.
- Secondo Anthropic, il progetto mirava a rendere **più nocivo il virus** – che ha un tasso di mortalità molto basso ma causa dolori articolari e muscolari a lungo termine – pur riconoscendo che ricerche analoghe potrebbero essere impiegate per sviluppare vaccini o terapie contro il virus stesso.
- Tuttavia, la ricerca avrebbe potuto essere utilizzata anche per rendere il virus più pericoloso. Anthropic ha ritenuto che il progetto non fosse innocuo come appariva: sebbene la richiesta di finanziamento suggerisse il coinvolgimento di ricercatori civili, l'attività era prevista presso un **istituto di ricerca militare.**
- Anthropic non ha reso noti i nomi o le affiliazioni degli istituti o dei ricercatori potenzialmente coinvolti, aggiungendo di aver sospeso tutti gli account collegati, rimosso le reti di inoltro utilizzate per aggirare i blocchi regionali e condiviso i risultati dell'indagine con altri laboratori di IA e con il governo federale.

Dichiarazione chiave

"Dall'esistenza di questo materiale di ricerca deduciamo che l'iniziativa non si limitava a una proposta di finanziamento", ha affermato Anthropic nel rapporto. "Stiamo sospendendo gli account che identifichiamo come collegati a questo gruppo di attività e continuiamo a integrare i risultati delle nostre indagini in nuove misure per rilevare e impedire l'uso improprio dei servizi di Anthropic da parte dei soggetti coinvolti".

Contesto della notizia

Questa settimana, un ex collaboratore di OpenAI e Anthropic, insieme a un attuale dipendente di Anthropic, ha affermato che esiste la possibilità che l'IA causi **l'estinzione dell'umanità entro il prossimo decennio**. Evan Hubinger, responsabile delle attività di Anthropic volte ad allineare le funzioni dell'IA agli obiettivi e ai valori umani, ha dichiarato di ritenere personalmente che vi sia una probabilità **superiore al 10%** che "l'IA possa uccidere tutti gli esseri umani" nei prossimi 10 anni, qualora lo sviluppo dovesse proseguire all'attuale ritmo di rapida accelerazione. Contesto principale La valutazione di Hubinger è giunta in risposta a un post di

Jacob Coxon – ricercatore di Anthropic dimessosi di recente dall'azienda dopo una breve permanenza – il quale sosteneva che le società di IA non stessero agendo in modo responsabile, "correndo a perdifiato verso una **superintelligenza** capace di auto-migliorarsi e scommettendo sulle nostre vite". Coxon ha aggiunto che i ricercatori del settore ritengono che tale tecnologia potrebbe "ucciderci tutti entro la fine del decennio" e ha precisato che le sue dichiarazioni non rappresentavano una "mossa di marketing". I post di Coxon e Hubinger hanno spinto i legislatori a chiedere l'approvazione di una legge che preveda un **"interruttore di emergenza" (kill switch)** per l'IA; in particolare, il senatore Bernie Sanders (Indipendente del Vermont) ha promesso di presentare un provvedimento per vietare la superintelligenza e sospendere lo sviluppo dell'IA. Il presidente Donald Trump è stato un convinto sostenitore dello sviluppo dell'IA negli Stati Uniti, incoraggiando le comunità locali ad accogliere la costruzione di data center e appoggiando la corsa agli armamenti nel campo dell'IA contro la Cina.