



Allarme ricercatori Anthropic su rischio estinzione umana con IA, Musk: «Una montatura»

Descrizione

(Adnkronos) «Dopo le dichiarazioni «apocalittiche» dell'ex ricercatore di Anthropic Jacob Coxon

sull'intelligenza artificiale, altri ricercatori e membri dello staff della startup specializzata in IA si sono detti pubblicamente d'accordo con lui, pubblicando a loro volta allarmi altrettanto gravi. In risposta, Elon Musk e altre figure conservatrici hanno definito su X il coro di preoccupazioni una «montatura» e un'operazione psicologica. Gli addetti ai lavori di Anthropic temono che la tecnologia che stanno sviluppando possa diventare cosí avanzata e pericolosa da causare lâ estinzione umana entro un decennio e affermano che molti dei loro colleghi la pensano allo stesso modo, pur non avendolo dichiarato pubblicamente.

Questi ricercatori e membri dello staff scrive il Guardian stanno rispondendo a un thread diventato virale del ricercatore di Anthropic Jacob Coxon, che mercoledì ha annunciato le sue dimissioni dall'azienda perché nÃ Anthropic nÃ la sua concorrente OpenAI, anch'essa suo ex datore di lavoro, stanno sviluppando modelli di intelligenza artificiale in modo responsabile e stanno mettendo a rischio le nostre vite. Anna Wang, che si occupa di sicurezza dell'intelligenza artificiale generale presso Anthropic e in precedenza ha lavorato presso DeepMind di Google, ha dichiarato ieri che molti in Anthropic desiderano rallentare lo sviluppo per elaborare un piano che mitighi i rischi potenzialmente associati a questi modelli di intelligenza artificiale.

«Non esiste ancora un piano scientifico valido per risolvere i rischi derivanti dall'intelligenza artificiale che si auto-migliora in modo ricorsivo», ha scritto Wang su X. Drake Thomas, un altro dipendente di Anthropic, ha affermato di rispettare la scelta di Coxon di astenersi dalla costruzione di questi modelli se teme che possano comportare rischi su scala planetaria. «Le cose si stanno muovendo troppo velocemente, non abbiamo neanche lontanamente il grado di certezza che desidereremmo per lâ Asi (superintelligenza artificiale)», ha affermato Thomas. Musk ha assunto una posizione piÃ conflittuale. «Sembra una trappola», ha scritto nel suo post. Coxon ha risposto

all'ultimo uomo più ricco del mondo, scrivendo: «Esisto davvero e queste sono le mie vere convinzioni. Potresti chiedere ai tuoi ricercatori di xAI cosa ne pensano di me, se non li avessi licenziati.»

In una dichiarazione al Guardian, un portavoce di Anthropic ha difeso la strategia dell'azienda. «Siamo sempre stati trasparenti sul fatto che l'intelligenza artificiale porterà enormi benefici ma anche rischi senza precedenti. Per affrontare questi rischi, continuiamo a sviluppare modelli con alcune delle più solide misure di sicurezza del settore», si legge nella dichiarazione. Musk e altri con una visione più positiva dell'IA hanno iniziato a diffondere teorie secondo cui il post di Coxon e le conseguenti polemiche farebbero parte di un'operazione psicologica («psychop» per influenzare negativamente l'opinione pubblica sull'IA. «Credo che le basi per questa operazione psicologica (per mancanza di un termine migliore) siano state gettate da tempo», ha scritto Musk su X. «Questo è stato solo il fiammifero che ha acceso il fuoco».

»

internazionale/esteri

webinfo@adnkronos.com (Web Info)

Categoria

1. Comunicati

Tag

1. Ultimora

Data di creazione

Settembre 11, 2026

Autore

redazione