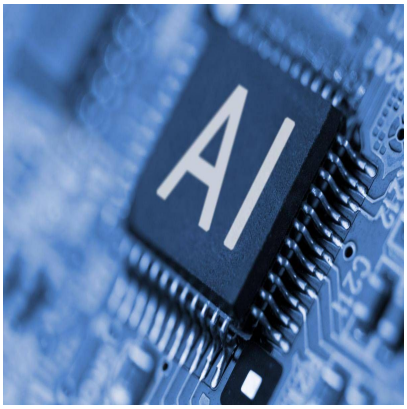

â??Lâ??Intelligenza Artificiale puÃ² ucciderci tutti in 10 anniâ?•, lâ??allarme del ricercatore

Descrizione

default watermark

(Adnkronos) â??
â??Lâ??Intelligenza Artificiale potrebbe uccidere tutti gli esseri umaniâ?• nel prossimo decennio. Eâ?? lâ??allarme che Evan Hubinger, un ricercatore di punta di Anthropic, uno dei colossi dellâ??AI, fa scattare con un post su X. â??Crediamo davvero che lâ??AI possa uccidere tutti gli esseri umani. Personalmente ritengo che la probabilitÃ sia superiore al 10% nel prossimo decennioâ?•, scrive in un post che viene rilanciato dallâ??emittente Cbs. Hubinger si presenta come ricercatore dellâ??azienda con sede a San Francisco. â??Credo che Anthropic stia facendo del suo meglio, ma non abbiamo ancora un piano per risolvere il problema dellâ??allineamento per la superintelligenza e non siamo chiaramente sulla strada giusta per riuscirci. Penso che il rischio dai modelli attuali sia basso. Quello che mi preoccupa Ã¨ la superintelligenza derivante dal miglioramento autonomo: come abbiamo detto, sta accadendo piÃ¹ velocemente di quanto pensassimoâ?•, aggiunge commentando il post di un collega che ha appena presentato le proprie dimissioni.

â??Oggi mi sono dimesso da Anthropic. Ho trascorso gli ultimi tre anni a occuparmi di ricerca sul pre-addestramento sia presso OpenAI che presso Anthropic. Nessuna delle due aziende sta agendo in modo responsabileâ?•, le parole di Jacob Coxon in un post su X. â??Stanno correndo a tutta velocitÃ verso una superintelligenza capace di auto-migliorarsi, mettendo a repentaglio le nostre vite. In OpenAI, molti non hanno interiorizzato totalmente la portata della posta in gioco per la civiltÃ . In Anthropic, la posta in gioco Ã¨ ben compresa, ma lâ??azienda Ã¨ incastrata in una corsa per arrivare prima degli altri: ritengono che nessun altro agirÃ responsabilmente e che quindiâ?• non â??debbano farlo nemmeno loro, nonostante i rischiâ?•.

La Cbs ha chiamato in causa lâ??AI Security Institute, lâ??organizzazione internazionale creata dai governi e incaricata di analizzare, testare e ridurre i rischi legati ai sistemi di intelligenza artificiale piÃ¹ avanzati. â??Lâ??AI Security Institute continua a collaborare strettamente con i partner del settore, tra cui Anthropic, per rendere i modelli piÃ¹ sicuriâ?•, le parole di un portavoce del Cabinet Office britannico alla Cbs, sottolineando che â??proprio la scorsa settimanaâ?• Ã¨ stato testato â??GPT-6 Astraâ?•, il

modello â??piÃ¹ potenteâ?• di OpenAI, prima del suo rilascio al pubblico. â??I rischi non si fermano ai confini nazionali e nessun Paese puÃ² affrontarli da solo. Il Regno Unito continuerÃ a testare i modelli piÃ¹ avanzati, a sviluppare una rigorosa comprensione scientifica delle loro capacitÃ e dei relativi rischi, e a garantire che le decisioni politiche si basino su evidenze concreteâ?•, ha affermato il portavoce.

â??

internazionale/esteri

webinfo@adnkronos.com (Web Info)

Categoria

1. Comunicati

Tag

1. Ultimora

Data di creazione
Settembre 9, 2026
Autore
redazione

default watermark

EI SRL - P.IVA IT16017671005