



Intelligenza artificiale, allarme Onu: "Rischio paralisi dei servizi essenziali e minaccia per le democrazie"

Descrizione

(Adnkronos) "L'intelligenza artificiale potrebbe costituire un rischio per l'umanità. A lanciare l'allarme è l'Alto commissario delle Nazioni Unite per i diritti umani, Volker Turk, in una lettera inviata agli Stati e alle aziende leader a livello mondiale nel campo dell'intelligenza artificiale. "Siamo alla vigilia di un cambiamento irreversibile, che non riguarderà solo noi, ma anche le generazioni future dell'umanità", avverte. "L'attuale corsa verso un'intelligenza artificiale sempre più potente rappresenta un salto verso maggiori rischi esistenziali per ogni aspetto delle nostre vite".

I fallimenti che coinvolgono sistemi di intelligenza artificiale, per non parlare di sistemi deliberatamente progettati in tal senso, scrive l'Alto commissario delle Nazioni Unite, "potrebbero paralizzare i servizi essenziali, frammentare le comunicazioni, destabilizzare le infrastrutture critiche e colpire al cuore stesso delle istituzioni democratiche".

"Oggi mi sono dimesso da Anthropic" ha scritto sui social nei giorni scorsi il ricercatore britannico 27enne Jacob Coxon, specializzato nei modelli di pre-addestramento dell'intelligenza artificiale. "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 responsabile" denuncia in un post su X. "Stanno correndo a tutta velocità verso una superintelligenza, capace di auto-migliorarsi, mettendo a repentaglio le nostre vite".

L'azienda di intelligenza artificiale Anthropic ha dichiarato, inoltre, di aver bloccato, negli ultimi mesi, diversi tentativi da parte di ricercatori di utilizzare versioni del suo software Claude in relazione al possibile sviluppo di armi biologiche. La società, concorrente di OpenAI (sviluppatrice di ChatGPT), ha tuttavia avvertito che è difficile stabilire le intenzioni alla base di tali richieste, poiché la stessa ricerca può servire sia a scopi medici che militari.

Dario Amodei, ceo di Anthropic, avverte che il ritmo di sviluppo dell'AI sta accelerando, oltre la capacità dei sistemi di sicurezza. Per questo propone di "regolare il ritmo di avanzamento" con un piano in tre fasi: valutatori integrati, coordinamento democratico e coordinamento globale. La prima fase "è un impegno unilaterale di Anthropic, la seconda richiede cooperazione tra aziende, la terza un accordo internazionale.

Amodei sostiene che la sicurezza deve diventare un terreno di competizione industriale, "una corsa verso la vetta". Rivendica, inoltre, la prudenza adottata finora, ma ritiene che non basti più: serve rallentare lo sviluppo delle capacità dei modelli per permettere alla prevenzione di tenere il passo. "Dobbiamo dosare il ritmo e usare con attenzione il tempo che guadagniamo" sottolinea.

Donald Trump interviene sul dibattito aperto dal ceo di Anthropic, Dario Amodei, che ha chiesto di rallentare lo sviluppo dell'AI per motivi di sicurezza. Parlando all'Irish Open, il presidente Usa ha escluso qualsiasi frenata: gli Stati Uniti, ha detto, sono "all'avanguardia rispetto alla Cina" e devono restare il Paese più avanzato al mondo. "Chi vince sull'Intelligenza artificiale, vince su tutto", ha dichiarato, ribadendo che la competizione con Pechino "è centrale. La sua linea è chiara: nessun rallentamento allo sviluppo dell'AI per mantenere il primato tecnologico americano.

Elon Musk approva l'appello lanciato dal ceo di Anthropic, Dario Amodei, sulla necessità di rallentare il ritmo di sviluppo dei modelli di intelligenza artificiale "Dario ha ragione" scrive l'imprenditore su X rispondendo al post di Amodei, pubblicato oggi sempre su X. Il ceo di Anthropic infatti ha argomentato con un lungo post i motivi di questa posizione determinata dalla consapevolezza che "il rischio che la velocità con cui un determinato processo della Ai, in grado di auto migliorarsi, possa superare la nostra capacità di comprendere e governare questi sistemi" è.

??

economia

webinfo@adnkronos.com (Web Info)

Categoria

1. Comunicati

Tag

1. Ultimora

Data di creazione

Settembre 14, 2026

Autore

redazione