



Intelligenza artificiale, l'allarme di Anthropic: Tra un anno l'AI potrebbe prendere il controllo di internet?

Descrizione

(Adnkronos)

Entro un anno l'AI potrebbe arrivare a prendere il controllo dell'intera internet, con danni potenziali per centinaia di miliardi di dollari. È l'allarme lanciato dal ceo di Anthropic, Dario Amodei. Dato il ritmo accelerato di sviluppo delle capacità dell'intelligenza artificiale, temo che tra 6-12 mesi uno sciame di agenti, come quello che provocato l'incidente OpenAI-Hugging Face, potrebbe essere in grado di prendere il controllo dell'intera internet causando potenzialmente centinaia di miliardi di dollari di danni che continuerebbero ad aumentare se l'intelligenza artificiale diventasse più potente senza i necessari meccanismi di protezione. È quindi necessario regolare il ritmo di avanzamento, scrive in un saggio, proponendo un piano in tre fasi per attuare tale rallentamento: valutatori integrati, coordinamento democratico e coordinamento globale.

La prima fase è un impegno unilaterale di Anthropic, e invita i governi a richiedere alle altre aziende di frontiera di fare altrettanto, spiega Amodei. La seconda fase richiede un coordinamento a livello di settore e il terzo richiede un coordinamento globale. Vogliamo dimostrare che è possibile sviluppare con cautela e avere successo commerciale, e fare della sicurezza un elemento su cui le aziende di intelligenza artificiale competono. In altre parole, creare una corsa verso la vetta, spiega Amodei.

Abbiamo sempre dedicato una parte sostanziale dei nostri sforzi allo studio, all'analisi e all'informazione del pubblico sui rischi dell'intelligenza artificiale, nonché alla promozione di una regolamentazione ponderata, anche quando questo ci ha attirato accuse di sensazionalismo, pessimismo o di strumentalizzazione normativa. Abbiamo cercato di dare priorità alla cautela rispetto alla velocità e alla prudenza rispetto al profitto. Negli ultimi mesi, però, mi sono convinto che affrontare appieno i rischi richieda ancora più prudenza: non solo investire nella prevenzione dei rischi, ma anche dosare il ritmo di sviluppo delle capacità in modo che la prevenzione abbia il tempo di adeguarsi, spiega il ceo. Dobbiamo rallentare il passo con cui miglioriamo le capacità dei modelli di intelligenza artificiale e fare un uso oculato del tempo che guadagniamo.

â??

economia

webinfo@adnkronos.com (Web Info)

Categoria

1. Comunicati

Tag

1. Ultimora

Data di creazione

Settembre 13, 2026

Autore

redazione

default watermark