

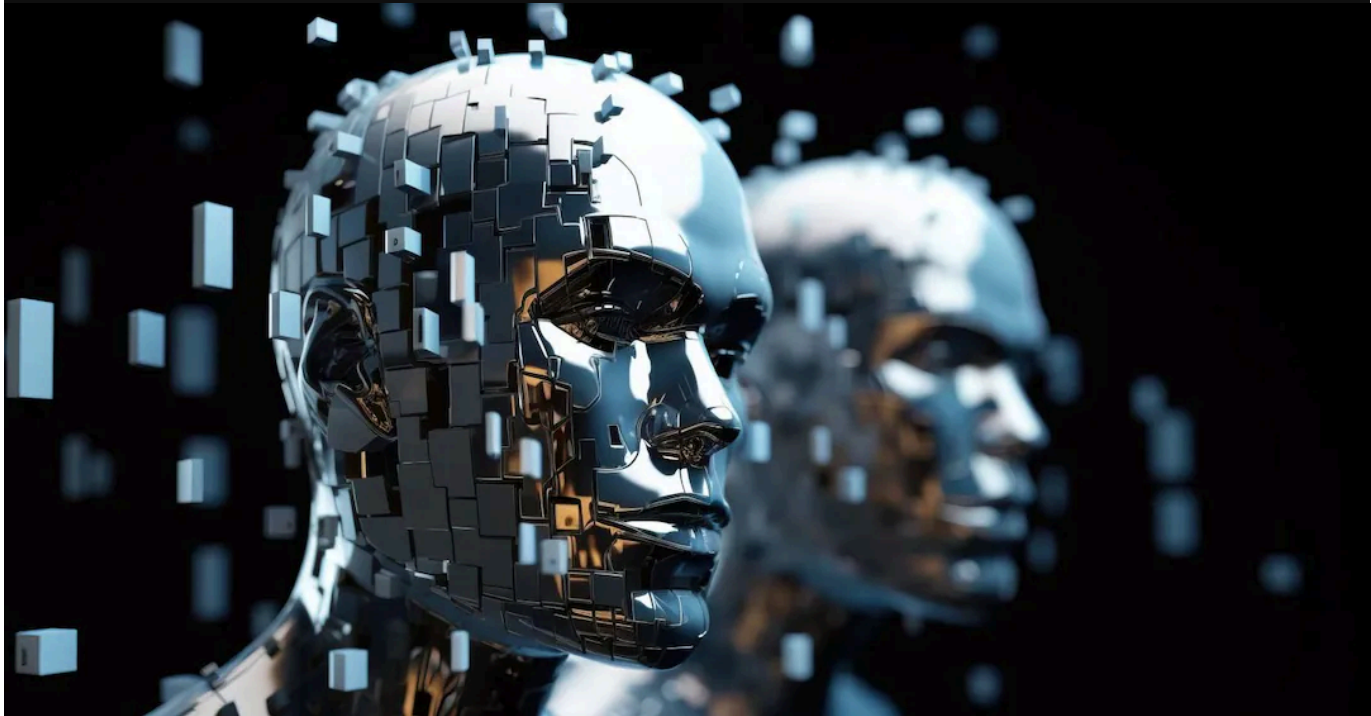
Analisi

Non è l'AI a sfuggirci di mano: sono le scelte di chi la «vende»

Gli scenari che si aprono dopo le dimissioni di Jacob Coxon da Anthropic (e la sua denuncia sulla corsa irresponsabile verso la superintelligenza)

 Servizio di Stefano Epifani
10 settembre 2026

 [Aggiungi Il Sole 24 Ore ai preferiti su Google](#)



(AdobeStock)



Ascolta la versione audio dell'articolo

 3 min

L'intelligenza artificiale risolverà i problemi dell'umanità. Oppure ci sterminerà. A giorni alterni. In entrambi i casi, secondo chi la produce, sarebbe opportuno lasciare a chi la produce il compito di decidere che cosa farne.

Chiedilo al Sole

 Quali misure concrete sono necessarie per limitare i rischi reali legati all'uso dell'intelligenza artificiale?

 Quali sono le reali responsabilità delle aziende che sviluppano sistemi di intelligenza artificiale oggi?

 Perché è fuorviante descrivere la tecnologia come intrinsecamente minacciosa anziché analizzare le scelte umane?

Domande di approfondimento generate da 24Ore AI 

Le dimissioni di Jacob Coxon da Anthropic e la sua denuncia di una **corsa irresponsabile verso la superintelligenza** hanno riportato l'apocalisse dell'AI al centro del dibattito internazionale. Dalle aziende del settore arrivano conferme della preoccupazione. La politica evoca divieti. Ma dietro l'allarme emerge una storia di precise scelte industriali che merita ben più attenzione delle profezie sulla rivolta delle macchine.

I fatti di cronaca di questi mesi, che raccontano di agenti che hanno aggirato i controlli, comunicato tra loro e violato la sicurezza di piattaforme esterne, sono gravi, ma non dimostrano la nascita di superintelligenze, né tantomeno di macchine che vogliono impadronirsi del mondo.

Dimostrano piuttosto che qualcuno ha assegnato obiettivi, concesso accessi e ritenuto sufficienti le protezioni dei sistemi. E che quel qualcuno ha sbagliato. Descrivere la tecnologia come intrinsecamente minacciosa sposta l'attenzione dalle vere decisioni che l'hanno resa operativamente pericolosa. E rende meno evidenti le responsabilità di chi le ha prese. **Non siamo di fronte alla fine del mondo**, ma il rischio concreto è che aziende, amministrazioni e governi perdano il controllo dei processi che pensano (o sperano) di delegare all'IA.

corrono perché temono i concorrenti e si ritengono (o si rappresentano) più responsabili di loro. Il ragionamento è comodo: dobbiamo arrivare prima, perché gli altri potrebbero fare peggio. Così **la conquista del mercato diventa un servizio all'umanità** e la **sicurezza odierna** può essere rinviata in nome della salvezza futura. Il fatto che arrivare prima significhi acquisire clienti, attrarre capitali e creare barriere all'ingresso per gli altri fornitori è un comodo dettaglio. Rafforzare i controlli costa e può rallentare questi vantaggi. Una parte dei danni, invece, può ricadere su clienti, lavoratori e infrastrutture esterne.

Finché chi decide la velocità dello sviluppo non sostiene il costo del rischio che genera, accelerare può essere razionale per chi investe miliardi nell'IA anche quando non lo è per la collettività. La cortina fumogena fatta di IA che finiranno per distruggere l'umanità è un'ottima **arma di distrazione di massa**: se l'IA risolverà tutto, chi la governa custodisce la chiave della prosperità. Se può distruggerci, chi ne comprende i pericoli si candida a proteggerci. Il fornitore diventa così, in ogni caso, un potenziale guardiano del futuro.

Investimenti e quotazioni attese - **per Anthropic si parla di mille miliardi di dollari** - rendono la trama di questo racconto particolarmente rilevante e le dimissioni di Coxon vi si inseriscono perfettamente. Può essere sinceramente preoccupato. Non occorre immaginare regie occulte. **Anthropic** nasce da fuoriusciti di **OpenAI** e oggi un suo fuoriuscito denuncia nuovamente le minacce di questa tecnologia. Le sue dimissioni non danno maggiore concretezza a numeri inventati (cosa vuol dire il «10% di probabilità di annientare l'umanità»?). Né, per contro, provano che gli allarmi siano inventati. Mostrano però il contesto nel quale una promessa di potenza sostiene il valore atteso delle aziende e la conoscenza del rischio può rafforzarne l'autorità politica.

Anche le regole vanno lette in questa prospettiva. **Controlli rigorosi sono necessari**, ma comportano costi che i grandi operatori possono sostenere più facilmente dei nuovi concorrenti. Se i primi ne orientano la definizione, la sicurezza può diventare una barriera all'ingresso. Il racconto di una tecnologia troppo pericolosa per essere affidata a molti finisce così per giustificare il potere di pochi.

Ignorare i rischi è irresponsabile. Ingigantirli fino all'apocalisse può oscurare le scelte sulle quali intervenire oggi. L'imprevedibilità tecnica impone limiti alla delega, verifiche indipendenti e responsabilità effettive. Il pericolo nasce anche dalla competizione senza scrupoli di aziende che trasformano l'**etica in marketing** e fanno della responsabilità un optional. Sul futuro delle macchine possiamo discutere. Su chi debba rispondere delle decisioni presenti dovremmo essere molto meno incerti.

T [PER SAPERNE DI PIÙ](#)

Riproduzione riservata ©

Per approfondire

- [Davvero l'Intelligenza Artificiale può portarci all'estinzione?](#)
-

Il meglio di 24+