Dario Amodei alla Cbs: «Per troppo tempo l'industria ha mentito sui rischi dell'intelligenza artificiale» di Michela Rovelli

Il Ceo di Anthropic insiste sulla necessità di un freno allo sviluppo dell'AI e spinge per una regolamentazione condivisa (Fonte: https://www.corriere.it/ 13 settembre 2026)

L'uno è diventato ben presto un due, che poi si è trasformato in un 4, poi arrivato subito a 8. Poi 16, 32. **Un ritmo esponenziale.** Usa i numeri **Dario Amodei** per spiegare **quanto veloce sta viaggiando lo sviluppo dell'intelligenza artificiale**. E quanto, dunque, **sia urgente rallentare**. Poche ore dopo la **pubblicazione del lungo post «We must pace the frontier»** il Ceo di Anthropic è entrato negli studi dell'emittente statunitense *Cbs* e ha risposto alle domande del corrispondente esperto di tecnologia **Jo Ling Kent.** Confermando quanto dichiarato per iscritto: la corsa all'AI va frenata, e regolamentata. Ma portando avanti anche accuse, che riguardano anche la sua stessa azienda: «**Non vi mentirò i pericoli sono reali. E credo che per troppo tempo il settore abbia mentito** alle persone sul fatto che questa tecnologia comportasse dei rischi», ammette. «**Questo non significa che oggi dobbiamo farci prendere dal panico** - aggiunge - Non significa che dobbiamo bloccare tutto». Ma lo definisce un «**segnale d'allarme**». E aggiunge: «**Senza misure di sicurezza l'AI potrebbe controllare Internet in sei mesi**».

Amodei **dopo le parole di un suo ex ricercatore** che hanno fatto il giro del mondo, ha aspettato qualche giorno per intervenire nel dibattito sul potenziale pericolo che l'AI rappresenta per l'umanità. **Jacob Coxon**, dopo essersi dimesso, su X ha raccontato come all'interno delle società che sviluppano questa tecnologia quotidianamente si parli del rischio che impari a migliorarsi in autonomia e di come questo potrebbe portare allo sterminio degli esseri umani entro la fine del

decennio. Un suo collega, **Evan Hubinger,** ha aggiunto che c'è una probabilità del 10 per cento che questo avvenga entro il prossimo decennio. Il primo a rispondere è stato **Sam Altman,** ceo di OpenAI - Coxon ha lavorato anche nei suoi laboratori - con **una richiesta ufficiale al Congresso statunitense di lavorare a una regolamentazione condivisa.** Poi ecco Amodei, che oltre a fare dichiarazioni ha anche accettato di rispondere alle domande di Jo Ling Kent.
Se è difficile distinguere tra reale allarmismo e mosse strategiche per sottolineare la potenza e l'efficacia dei prodotti da loro creati - **stiamo pur sempre parlando delle persone che l'AI la sviluppano e addestrano** e che dunque hanno in mano tutti i pulsanti per decidere quando andare veloce e quando fermarsi - **alcuni primi casi concreti di «incidenti» ci sono già stati**. OpenAI ha dovuto ammettere che un suo modello è stato responsabile di aver violato la piattaforma Hugging Face. Così come Anthropic, che ha denunciato episodi analoghi.
«Penso che il passo successivo sia, in qualche modo, che **le aziende collaborino per definire degli standard,** in particolare standard di sicurezza per il lancio dei prodotti, e magari anche qualcosa riguardo al ritmo di lancio», ha aggiunto Amodei. **Sia Altman sia Elon Musk** - la sua azienda xAI, oggi fusa con Tesla, ha sviluppato Grok - hanno già dichiarato di essere **d'accordo** su questa collaborazione. Altman ha addirittura deciso di r**imandare la quotazione in Borsa,** previsto entro la fine dell'anno: «Considerando tutto ciò che sta accadendo in materia di sicurezza, il momento non sarebbe opportuno e non sentiamo alcuna pressione in tal senso», ha spiegato a *Fortune*.
Anche **Demis Hassabis, cofondatore e presidente di Google DeepMind** e ricercatore capo di Alphabet (nonché premio Nobel per la chimica nel 2024), ha convenuto sul fatto che quella indicata da Amodei sia «la strada giusta da percorrere». Ma secondo lo stesso Amodei servono **linee guida e regolamentazioni a livello governativo** perché il «freno» abbia successo. Nell'intervista alla Cbs spiega: «Dobbiamo assicurarci che ogni generazione di modelli che rilasciamo sia stata adeguatamente testata».
Amodei lancia delle **proposte** su come - concretamente - limitare lo sviluppo dell'AI. Secondo lui, e gli altri concordano, bisogna utilizzare dei «**valutatori integrati**», ossia delle **persone specializzate e degli strumenti forniti da organizzazioni terze** (dunque neutrali) che possano **verificare che le aziende sviluppatrici stiano rispettando le regole** che verranno definite sul ritmo di velocità a cui procedere nello sviluppo dei modelli. Gli stessi strumenti dovrebbero anche garantire la pronta segnalazione di ogni incidente relativo alla sicurezza. «Ogni volta che qualcuno sviluppa un modello di IA, c'è un valutatore di terze parti in grado di verificare se l'azienda sta seguendo le pratiche di sicurezza a cui si è impegnata – **è come un ispettore alimentare**», precisa. Una sorta di sistema di *compliance* esterno. Questi valutatori dovranno però **sedere a fianco di ingegneri e ricercatori** e dovranno avere accesso a tutti i sistemi. Prima però, bisogna definire queste regole: le principali aziende «nei paesi democratici» dovranno, secondo Amodei, coordinarsi per definire «**standard di sicurezza comuni e limiti al progresso incontrollato dell'AI**». Qui serve l'intervento del governo: Amodei continua a sostenere la necessità di una legge federale, anche per arginare **i timori di possibili indagini Antitrust** qualora questa pausa dovesse effettivamente

verificarsi. Sono timori che Amodei non nasconde: «É utile che il governo degli Stati Uniti faccia da mediatore o almeno faciliti queste discussioni: non è necessario che partecipi, ma deve concedere una deroga limitata per determinati tipi di discussioni sulla sicurezza», aveva scritto nel post. Ma coinvolgere l'amministrazione statunitense sembra essere sempre più complicato. Parlando coi giornalisti nel suo golf club di Doonbeg, in Irlanda, il presidente **Donald Trump ha criticato le «forze negative» che vogliono rallentare lo sviluppo dell'AI**. «Possiamo introdurre delle **misure di salvaguardia**, possiamo fare questo e quello - ha affermato il tycoon -. Ma credo che al momento ci siano molte forze negative che sollevano la questione senza averne motivo, **tirando in ballo scenari che non si verificheranno**».

Altri timori sono legati all'innovazione cinese, che secondo il ceo di Anthropic si può frenare rifiutando al Paese - e alle sue aziende - la vendita dei chip e delle attrezzature più potenti. Questo potrebbe «rallentare i progressi della Cina abbastanza da ampliare il vantaggio degli Stati Uniti nei prossimi 3-5 anni». In ogni caso, sottolinea, serve un «**coordinamento globale**» che comprenda tutti, anche «i governi autoritari, nella misura in cui ciò sia possibile». Sempre alla Cbs, alla domanda riguardante la complessa sfida con le aziende asiatiche ha risposto: «Questo è il dilemma più difficile. È l'aspetto più complesso della nostra situazione».

Il tema è al centro del dibattito in tutto il mondo. L'intelligenza artificiale e una sua possibile regolamentazione è stata argomento di discussione anche al **18esimo Brics Summit,** al momento in corso a Nuova Delhi, in India. Alla conferenza che raduna le economie emergenti, proprio **la Cina ha dichiarato la volontà di guidare la creazione di una «zona di AI open source»** per gli 11 Stati membri (tra cui Brasile, Russia, India, e Sudafrica). É stato lo stesso presidente Xi Jinping a parlarne: le aziende cinesi, ha spiegato, si stanno concentrando sullo sviluppo di **modelli «aperti»**, dunque accessibili a tutti, consentendone la comprensione e la modifica, che si contrappongono a quelli **«chiusi» commercializzati dagli Stati Uniti**. La proposta rientra, spiega, in una «iniziativa per un'intelligenza artificiale più inclusiva», per «sostenere la cooperazione nello sviluppo e nell'applicazione di grandi modelli linguistici». Spostandosi verso Ovest, **l'Unione europea** può - ora più che mai - **vantare l'unico tentativo di regolamentare l'AI al momento esistente al mondo, l'AI Act**. Mentre negli Stati Uniti quelle istituzioni chiamate in causa sia da Altman sia da Amodei al momento non si sbilanciano. Al momento esiste solo **una proposta di legge bipartisan**, nota come «**AI Kill Switch Bill**», che prevede di inserire un interruttore di sicurezza per spegnere le AI qualora vadano fuori controllo. Ma anche altri grossi attori della corsa tecnologica hanno deciso - per ora - di non entrare nel dibattito. Tra questi: **Meta**. Qualche giorno prima che Coxon decidesse di scrivere su X le sue riflessioni, era trapelato il contenuto di **una telefonata tra Mark Zuckerberg e Trump**. Il fondatore di Facebook chiedeva al presidente statunitense di **abbandonare ogni piano per limitare lo sviluppo della AI**.