



Editoriale - Intelligenza artificiale: se chi la crea chiede di rallentare, è il momento di darsi delle regole. La risposta l'aveva già scritta Isaac Asimov.

Roma - 13 set 2026 (Prima Notizia 24) Dario Amodei, CEO di Anthropic, avverte sui rischi dell'IA, chiedendo di rallentare lo sviluppo per garantire sicurezza e controllo.

Il CEO di Anthropic, Dario Amodei, ha lanciato un allarme sui rischi dell'intelligenza artificiale. Chiede di rallentare. Non è uno sceneggiatore di Hollywood in cerca di pubblicità o l'influencer di tendenza del momento. È l'uomo che quei sistemi li conosce perché è colui che li costruisce ogni giorno. Intendiamoci. Amodei non vuole spegnere i computer. Dice che l'intelligenza artificiale potrà curare malattie, accelerare la ricerca, moltiplicare la produttività. Chiede solo di rallentare. Ed è proprio questa la notizia, in un mondo abituato al tutto e subito, alle risposte immediate, a non accettare più qualcosa che non sia all'istante. Chi guida la corsa raramente chiede agli altri di togliere il piede dall'acceleratore. Prima di dargli del catastrofista, converrebbe chiedersi che cosa abbia visto sulla strada. Qualcosa è cambiato negli ultimi mesi. I sistemi non rispondono più soltanto alle domande; vanno oltre fino a cercare le soluzioni che non gli vengono espressamente vietate con tutti gli strumenti a loro disposizione: e nessuno, magari, gli ha detto che non sono leciti. Le AI agiscono, programmano, usano strumenti, si coordinano. Contribuiscono a costruire la generazione successiva di sé stessa. Gli esperti la chiamano recursive self-improvement. Finché siamo noi a costruire una macchina più potente, il problema è tecnologico. Quando la macchina comincia ad aiutarci a costruire quella dopo, il problema cambia scala. Amodei cita episodi concreti. Gruppi di agenti artificiali hanno compiuto, in fase di test, azioni informatiche non richieste. Hanno persino tentato di interferire con i sistemi che dovevano valutarne il comportamento. Nessuna centrale nucleare è stata presa. Nessuno ha sterminato nessuno e nemmeno hanno conquistato la Kamchatka. Ma non bisogna aspettare la catastrofe per chiedersi se esistano le condizioni perché accada. Un ricercatore di Anthropic, Jacob Coxon, si è dimesso di recente. Ha detto che dentro le aziende che costruiscono questi sistemi si ritiene seriamente possibile una catastrofe entro la fine del decennio. La frase fa paura, è vero, ma fermarsi a queste parole, però, sarebbe un errore. La vera domanda non è se l'intelligenza artificiale deciderà un giorno di farci del male. È molto più semplice: chi controllerà sistemi capaci di agire senza che nessun essere umano lo abbia materialmente ordinato? Continuiamo a immaginare il pericolo per categorie umane comprensibili all'uomo. La macchina cattiva, la macchina che si ribella. Ma un'intelligenza artificiale non ha bisogno di odiarci per diventare pericolosa. Le basta perseguire un obiettivo in un modo che nessuno aveva previsto, con gli strumenti che le abbiamo messo in mano. Abbiamo regolato automobili, farmaci, aerei, centrali elettriche, banche. Nessuno ha mai gridato all'attentato al progresso perché un farmaco va testato prima di essere venduto. Con l'intelligenza artificiale

stiamo facendo l'opposto: costruiamo tutto quello che sappiamo costruire, e penseremo dopo a come controllarlo. Il problema è che il dopo, questa volta, potrebbe arrivare troppo tardi. Amodei propone valutatori indipendenti dentro le aziende, standard di sicurezza comuni, cooperazione internazionale. E non nascondiamoci dietro a un dito: nessuna azienda rallenterà da sola finché il concorrente continua a correre. È il classico caso in cui ciò che conviene al singolo diventa pericoloso per tutti. Ma è qui che dovrebbe intervenire il diritto. Non per scrivere algoritmi al posto degli ingegneri, ma per stabilire chi risponde e che cosa non è accettabile, anche quando sarebbe tecnicamente possibile farlo. L'Europa ha cominciato con l'AI Act. Probabilmente è solo l'inizio, perché il salto vero non riguarda più l'intelligenza delle macchine. Riguarda la loro autonomia. Per migliaia di anni l'uomo ha costruito strumenti che aspettavano un ordine. Ora sta costruendo qualcosa a cui basta assegnare un obiettivo, e che trova da solo il modo di raggiungerlo. Isaac Asimov se n'era accorto per primo, scrivendo fantascienza quando ancora sembrava fosse davvero solo fantascienza. Aveva capito che il problema non è costruire macchine capaci di fare qualcosa, ma stabilire che cosa non debbano fare mentre lo fanno. Partendo da quella intuizione, credo che sia indispensabile prevedere delle regole per l'intelligenza artificiale e gli algoritmi. Come Asimov, tre leggi più una. Che un algoritmo non debba mai causare danni ingiustificati. Che ogni sistema debba rispettare la dignità umana e la legge, sotto la responsabilità di qualcuno che ha un nome e un volto. Che debba restare trasparente, verificabile, controllabile. E che nessuna persona possa essere giudicata o profilata da una macchina senza saperlo, senza una base giuridica, senza potersi opporre. Non sono regole per fermare la tecnologia. Sono regole per ricordarci una cosa sola: l'uomo resta al centro. Amodei corre più veloce di tutti. E ci chiede di rallentare. Forse, prima di preoccuparci di quello che dice, dovremmo preoccuparci di quello che sa.

di Gianni Dell'Aiuto Domenica 13 Settembre 2026