Primo Piano - *Anthropic, i vertici lanciano l'allarme: "Fermare lo sviluppo globale dell'IA, rischiamo di perderne il controllo"*

Roma - 05 giu 2026 (Prima Notizia 24) **In un saggio, Marina Favaro e Jack Clark avvertono che l'umanità sta per perdere il controllo sulla tecnologia, evidenziando il fenomeno dell'auto-miglioramento ricorsivo dei network neurali e paragonano la proliferazione dei modelli linguistici a una corsa agli armamenti, sottolineando i rischi associati.**

I timori legati agli scenari fantascientifici sulla ribellione delle macchine si trasferiscono improvvisamente dai copioni cinematografici ai tavoli dei vertici dell'industria tecnologica mondiale. Un pesantissimo campanello d'allarme è stato suonato da Anthropic, il laboratorio di ricerca e sviluppo diventato la realtà aziendale più preziosa del pianeta nel campo dell'intelligenza artificiale e nota al pubblico come creatrice del chatbot Claude. La start-up statunitense ha formalizzato una richiesta drastica e trasparente rivolta a tutta la comunità internazionale: un congelamento planetario della progettazione di nuovi modelli computazionali, avvisando che l'umanità si sta avvicinando a grandi passi alla soglia critica oltre la quale gli esseri umani “perderanno il controllo” della tecnologia da loro stessi creata. La presa di posizione, che ricalca le dinamiche di una vera e propria profezia di Cassandra, è stata formalizzata all'interno di un saggio scritto a quattro mani da Marina Favaro, alla guida della divisione scientifica della società, e dal Ceo Jack Clark. Le conclusioni del documento poggiano su un archivio di metriche interne mai divulgate prima d'ora, focalizzate sul fenomeno dell' “recursive self-improvement”, espressione traducibile come auto-miglioramento ricorsivo. Questa funzione descrive l'abilità dei network neurali di scrivere in autonomia i propri upgrade evolutivi e di pianificare i parametri dei modelli destinati a succedergli. Nei laboratori di Anthropic, per esempio, una quota superiore all'80% delle righe di codice inserite nell'infrastruttura di produzione viene ormai generata direttamente dalle macchine. Il timore dei fondatori è che la transizione verso l'automazione totale della programmazione, in cui l'intervento umano non influirà più minimamente sulla catena evolutiva, si compia molto prima del previsto, cogliendo impreparate le istituzioni governative che si muovono con estrema lentezza nella stesura di leggi e barriere normative. Nell'esaminare lo scenario geopolitico attuale, Favaro e Clark hanno accostato la proliferazione dei grandi modelli linguistici alla storica “corsa agli armamenti” delle superpotenze mondiali. Il paragone non tiene conto soltanto delle immense risorse finanziarie iniettate ogni giorno dai fondi d'investimento nel comparto della Silicon Valley, ma riguarda l'intrinseca pericolosità legata alla diffusione di simili architetture digitali. Con un'aggravante logistica rispetto ai conflitti convenzionali: le mega-strutture di calcolo e i server impiegati per l'addestramento dei software sono enormemente più semplici da occultare rispetto ai silos missilistici sotterranei, i codici sono accessibili a chiunque e la spinta commerciale a muoversi in clandestinità è

gigantesca. Anthropic — che ha recentemente conquistato lo scettro di leader del mercato sorpassando OpenAI grazie a una capitalizzazione di mercato pari a 965 miliardi di dollari — ha deciso di dare il buon esempio sospendendo a tempo indeterminato la distribuzione di Mythos, il suo applicativo più avanzato, a causa dei timori del team che potesse essere utilizzato da attori malevoli per orchestrare attacchi informatici su scala globale. I massimi dirigenti dell'azienda si sono detti disposti a spegnere i propri supercomputer e a congelare ogni linea di ricerca, ponendo però come condizione vincolante il fatto che l'intero ecosistema dei concorrenti, sia in Occidente che in Asia, accetti di sottoscrivere la medesima tregua tecnologica: “Se fosse possibile rallentare efficacemente lo sviluppo di questa tecnologia per darci più tempo per affrontare le sue immense implicazioni, pensiamo che sarebbe probabilmente una cosa positiva”. Un invito alla prudenza motivato anche dai calcoli probabilistici sulla sicurezza condivisi dall'altro co-Ceo della compagnia, Dario Amodei, secondo il quale l'evoluzione incontrollata degli algoritmi nasconde insidie fatali per la civiltà: “C'è una possibilità su quattro che le cose vadano davvero, davvero male”.

(Prima Notizia 24) Venerdì 05 Giugno 2026

Sede legale : Via Costantino Morin, 45 00195 Roma
E-mail: redazione@primanotizia24.it