

Primo Piano - *La mente umana batte l'IA: crollano i modelli in un difficile test di matematica del progetto First Proof*

Roma - 16 giu 2026 (Prima Notizia 24) **Quattro software di ultima generazione falliscono l'obiettivo del punteggio pieno.**

I rapidi progressi nel campo dei sistemi computazionali avanzati registrano una significativa battuta d'arresto quando messi alla prova sul terreno del ragionamento logico puro e della formulazione teorica. All'interno della prova d'esame più complessa e strutturata mai ideata per saggiare le competenze scientifiche dei software, la mente umana ha riaffermato la propria superiorità nei confronti delle macchine. Il banco di prova ha coinvolto quattro differenti architetture digitali, tra le quali figuravano le versioni più aggiornate dei prodotti commerciali di OpenAI, e nessuna di esse ha dimostrato le capacità necessarie per risolvere la totalità dei dieci problemi proposti. Come documentato in un focus pubblicato sulla versione online della prestigiosa rivista accademica Nature, la prestazione più convincente è stata quella registrata dall'algoritmo sviluppato dal Politecnico di Zurigo (Eth), capace di fornire la soluzione esatta per sei quesiti su dieci. Alle spalle dell'ateneo svizzero si è posizionato il sistema messo a punto dall'Università della California a Los Angeles, mentre l'ultimo gradino della graduatoria è stato occupato dalla piattaforma dell'Università di Princeton, la quale si differenziava dalle concorrenti per l'utilizzo del motore di calcolo Gemini strutturato da Google. L'esperimento si inserisce nelle attività di First Proof, un programma di monitoraggio autonomo nato con lo scopo di tracciare l'evoluzione delle abilità sintetiche applicate alle scienze matematiche. Per garantire l'assoluta validità scientifica del test ed escludere che le macchine potessero semplicemente attingere a nozioni già assimilate durante le fasi di programmazione e addestramento, i quesiti somministrati sono stati creati ex novo, evitando qualsiasi precedente pubblicazione sul web o all'interno di saggi accademici. L'esattezza e la coerenza delle soluzioni elaborate dai computer sono state successivamente analizzate e validate da una commissione di trenta matematici professionisti. Le linee guida del progetto imponevano che i software candidati fossero accessibili alla collettività; tale vincolo ha fatto sì che OpenAI risultasse l'unico colosso del settore tecnologico a figurare in modo diretto nella competizione. Nonostante ciò, due delle restanti tre configurazioni in gara sfruttavano indirettamente la medesima tecnologia di base della società americana, impiegando moduli automatizzati nei quali i testi generati venivano progressivamente perfezionati mediante un confronto incrociato con altri assistenti digitali. Gli accademici di Zurigo hanno infine avviato un'analisi preliminare per fare luce sulle cause dei fallimenti sistematici registrati su alcune specifiche tracce, rimaste insolute per tutti i partecipanti. Dai primi riscontri è emerso come in determinate circostanze gli algoritmi siano rimasti privi di quell'impulso intuitivo decisivo per sbloccare i passaggi logici finali; in altri frangenti, pur avendo impostato correttamente la strategia risolutiva generale, i sistemi non sono stati in grado di gestire con la dovuta accuratezza tutti i singoli dettagli

analitici richiesti dal problema.

(Prima Notizia 24) Martedì 16 Giugno 2026

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