

Intelligenza artificiale, il ritorno di “Oppenheimer”

Pubblicato: Domenica 19 Maggio 2024



Jan Leike, ex-responsabile della sicurezza di OpenAI, si è dimesso. Ecco la sua lettera di addio, pubblicata su X.

“Ieri è stato il mio ultimo giorno come capo dell’allineamento, leader del super-allineamento e dirigente di OpenAI. È stato un viaggio incredibile negli ultimi 3 anni. Il mio team ha lanciato il primo LLM RLHF con InstructGPT, ha pubblicato la prima supervisione scalabile sugli LLM, ha pionierizzato l’interpretabilità automatizzata e la generalizzazione da debole a forte. Presto usciranno altre cose emozionanti. Sono così grato per le tante persone straordinarie con cui ho avuto il privilegio di lavorare, sia all’interno che all’esterno del team di super-allineamento. OpenAI ha un talento eccezionalmente intelligente, gentile ed efficace.

Lasciare questo lavoro è stata una delle decisioni più difficili che abbia mai preso, perché abbiamo urgentemente bisogno di capire come indirizzare e controllare sistemi di intelligenza artificiale molto più intelligenti di noi. Mi sono unito a OpenAI perché pensavo che sarebbe stato il miglior posto al mondo per fare questa ricerca. Tuttavia, sono stato in disaccordo con la leadership di OpenAI sulle priorità fondamentali dell’azienda per un bel po’ di tempo, fino a quando non siamo arrivati a un punto di rottura.

Credo che molto più del nostro tempo e delle nostre risorse dovrebbero essere dedicati a prepararci per le prossime generazioni di modelli, alla sicurezza, al monitoraggio, alla preparazione, alla robustezza contro gli attacchi, al super-allineamento, alla riservatezza, all’impatto sociale e ad argomenti correlati. Questi problemi sono piuttosto difficili da risolvere e sono preoccupato che non siamo sulla strada giusta per riuscirci.

Negli ultimi mesi il mio team ha navigato controvento. A volte lottavamo per ottenere risorse di calcolo e diventava sempre più difficile portare avanti questa ricerca cruciale. Costruire macchine più intelligenti degli esseri umani è un'impresa intrinsecamente pericolosa. OpenAI si assume una responsabilità enorme a nome di tutta l'umanità.

Ma negli ultimi anni, la cultura e i processi di sicurezza hanno ceduto il passo alla priorità di avere prodotti sempre più brillanti. Siamo molto in ritardo nel prendere incredibilmente sul serio le implicazioni dell'AGI (intelligenza artificiale generale). Dobbiamo prioritizzare la preparazione per queste sfide al meglio delle nostre capacità. Solo allora possiamo garantire che l'AGI benefici tutta l'umanità. OpenAI deve diventare un'azienda di AGI che mette la sicurezza al primo posto.

A tutti i dipendenti di OpenAI, voglio dire: Imparate a percepire l'AGI. Agite con la gravitas appropriata per ciò che state costruendo. Credo che possiate "realizzare" il cambiamento culturale necessario. Conto su di voi."

Le dimissioni di Leike sono arrivate alla fine della settimana in cui **OpenAI ha lanciato GLT 4o**, la sua versione più potente di sempre (vedi articolo <https://www.varesenews.it/2024/05/gpt-4o-rivoluzione-nelleducazione-e-nella-comunicazione-interpersonale/1934979/>).

Il clamore del lancio è offuscato da questa uscita altrettanto clamorosa e ci induce ad alcune considerazioni sulla gestione di questo cambiamento trasformativo per la nostra epoca.

General

1. Trasformazione e integrazione educativa. GPT-4o offre tutoraggio personalizzato in formati testuali, audio e visivi in tempo reale, presentando un cambiamento significativo nei metodi educativi. Le scuole e le istituzioni devono integrare attentamente questa tecnologia per migliorare l'apprendimento senza oscurare le pratiche educative tradizionali. L'obiettivo è arricchire l'esperienza educativa, garantendo che l'AI serva come strumento di miglioramento piuttosto che come sostituto degli educatori umani.

2. Bilanciamento tra tecnologia e interazione umana. La capacità di GPT-4o di facilitare conversazioni e apprendimento presenta opportunità per arricchire la comunicazione. Tuttavia, gestire l'equilibrio tra interazioni AI e connessioni umane è cruciale. Mentre l'AI può supportare sessioni di brainstorming virtuale e la preparazione a interviste, è essenziale preservare il valore degli incontri faccia a faccia. Le organizzazioni devono sviluppare strategie che incorporino i benefici dell'AI mantenendo l'integrità delle interazioni personali.

3. Affrontare le implicazioni etiche e psicologiche. L'uso crescente dell'intelligenza artificiale come GPT-4o porta le considerazioni etiche e psicologiche in primo piano. Il tempo prolungato davanti agli schermi e la dipendenza dall'IA potrebbero influenzare la capacità di attenzione e la salute mentale. La gestione efficace del cambiamento dovrebbe includere misure per mitigare questi rischi, come promuovere un tempo bilanciato davanti agli schermi e incoraggiare attività in presenza. Inoltre, garantire l'uso etico dell'IA attraverso una gestione sicura dei dati sarà vitale per mantenere la fiducia e proteggere veramente le informazioni personali (privacy).

Solo affrontando questi elementi critici, potremo sfruttare il potenziale di GPT-4o in modo responsabile, aprendo la strada a un futuro equilibrato e innovativo nell'educazione e nella comunicazione.

Questa sfida mi ricorda il film #Oppenheimer. Gli ingegneri di ieri sono i "programmatori" di oggi. I rischi sono gli stessi.

"Lei è l'uomo che ha dato loro il potere di distruggere sé stessi. E il mondo non è pronto", Niels Bohr nel film Oppenheimer.

di [Giuseppe Geneletti](#)

