Elon Musk e Steve Wozniak chiedono una moratoria per l'Intelligenza Artificiale. Troppo pericolosa

scenarieconomici.it/elon-musk-e-steve-wozniak-chiedono-una-moratoria-per-lintelligenza-artificiale-troppo-pericolosa/

Guido da Landriano 29 marzo 2023

Elon Musk, Steve Wozniak, il pioniere dell'IA Yoshua Bengio e altri hanno firmato una lettera aperta in cui chiedono una pausa di sei mesi nello sviluppo di nuovi strumenti di IA più potenti del GPT-4, la tecnologia rilasciata all'inizio del mese dalla startup OpenAI sostenuta da Microsoft, come riporta il Wall Street Journal.

I sistemi di IA contemporanei stanno diventando competitivi per l'uomo in compiti generali e dobbiamo chiederci se sia il caso di lasciare che le macchine occupino i nostri canali di informazione: Dobbiamo lasciare che le macchine inondino i nostri canali di informazione con propaganda e falsità? Dovremmo automatizzare tutti i lavori, compresi quelli più soddisfacenti? Dovremmo sviluppare menti non umane che alla fine potrebbero superarci di numero, essere più intelligenti, obsolete e sostituirci? Dobbiamo rischiare di perdere il controllo della nostra civiltà? Queste decisioni non devono essere delegate a leader tecnologici non eletti. I potenti sistemi di intelligenza artificiale dovrebbero essere sviluppati solo quando saremo sicuri che i loro effetti saranno positivi e i loro rischi gestibili. futureoflife.org

"Abbiamo raggiunto il punto in cui questi sistemi sono abbastanza intelligenti da poter essere utilizzati in modi pericolosi per la società", ha detto Bengio, direttore del Montreal Institute for Learning Algorithms dell'Università di Montreal, aggiungendo: "E non lo capiamo ancora".

Le loro preoccupazioni sono state esposte in una lettera intitolata "***Pause Giant AI Experiments: An Open Letter***", promossa dal Future of Life Institute, un'organizzazione no-profit consigliata da Musk.

La lettera non chiede di fermare tutto lo sviluppo dell'IA, ma esorta le aziende a interrompere temporaneamente l'addestramento di sistemi più potenti del GPT-4, la tecnologia rilasciata questo mese dalla startup OpenAI sostenuta da Microsoft Corp. Questo include la prossima generazione della tecnologia di OpenAI, GPT-5.

I funzionari di OpenAI affermano di non aver ancora iniziato la formazione di GPT-5. In un'intervista, l'amministratore delegato di OpenAI Sam Altman ha dichiarato che l'azienda ha da tempo dato priorità alla sicurezza nello sviluppo e ha trascorso più di sei mesi a effettuare test di sicurezza su GPT-4 prima del suo lancio. –WSJ

Secondo Goldman, invece, fino a 300 milioni di posti di lavoro potrebbero essere sostituiti dall'IA, poiché "due terzi delle occupazioni potrebbero essere parzialmente automatizzati dall'IA". Una situazione che porterebbe a sconvolgimenti sociali di difficile gestione.

La cosiddetta IA generativa crea contenuti originali sulla base di suggerimenti umani, una tecnologia che è già stata implementata nel motore di ricerca Bing di Microsoft e in altri strumenti. Poco dopo, Google ha implementato un rivale chiamato Bard. Altre aziende, tra cui Adobe, Salesforce e Zoom, hanno introdotto strumenti avanzati di IA. "La gara inizia oggi", ha dichiarato il mese scorso il CEO di Microsoft Satya Nadella. "Ci muoveremo, e in fretta".

Uno degli organizzatori della lettera, Max Tegmark, che dirige il Future of Life Institute ed è professore di fisica al Massachusetts Institute of Technology, la definisce una "corsa al suicidio".

"È un peccato inquadrare la questione come una corsa agli armamenti", ha detto. "È piuttosto una corsa al suicidio. Non è importante chi arriverà per primo. Significa solo che l'umanità nel suo complesso potrebbe perdere il controllo del proprio destino".

Il Future of Life Institute ha iniziato a lavorare alla lettera la settimana scorsa e inizialmente ha permesso a chiunque di firmare senza verificare l'identità. A un certo punto, il nome del signor Altman è stato aggiunto alla lettera, ma poi è stato rimosso. Il signor Altman ha dichiarato di non aver mai firmato la lettera. Ha dichiarato che l'azienda si coordina spesso con altre aziende di IA per quanto riguarda gli standard di sicurezza e per discutere di problemi più ampi.

Anche Harari ha firmato il Manifesto per la moratoria della Super-AI

mauriziobloondet.it/anche-harari-ha-firmato-il-manifesto-per-la-moratoria-della-super-ai/

Maurizio Blondet 29 marzo 2023

Il CEO di Tesla Elon Musk, il co-fondatore di Apple Steve Wozniak, e persino lo Harari nota moglie e demiurga del WEF, sono tra gli oltre 1.000 imprenditori tecnologici ed esperti di intelligenza artificiale che chiedono una pausa sullo sviluppo di sistemi di intelligenza artificiale più potenti del GPT-4 di OpenAI.

Il Manifesto esordisce parole di profonda umanità e allarme:

I sistemi di intelligenza artificiale contemporanei stanno ora diventando competitivi rispetto all'uomo in compiti generali *e dobbiamo chiederci: dovremmo* ***lasciare che le macchine inondino i nostri canali di informazione con propaganda e falsità?*** *Dovremmo* ***automatizzare tutti i lavori*** *, compresi quelli soddisfacenti?* ***Dovremmo sviluppare menti non umane che alla fine potrebbero essere più numerose, superate in astuzia, obsolete e sostituirci?*** *Dovremmo rischiare di perdere il controllo della nostra civiltà?* ***Tali decisioni non devono essere delegate a leader tecnologici non eletti*** *. Potenti sistemi di intelligenza artificiale dovrebbero essere sviluppati solo quando saremo certi che i loro effetti saranno positivi e i loro rischi saranno gestibili. -futurodellavita.org* _

" **Siamo arrivati al punto in cui questi sistemi sono abbastanza intelligenti da poter essere utilizzati in modi pericolosi per la società** ", ha affermato Bengio, direttore del Montreal Institute for Learning Algorithms dell'Università di Montreal, aggiungendo "E non abbiamo ancora capito quali."

Le loro preoccupazioni sono state espresse in una lettera aperta intitolata "*Pause Giant AI Experiments: An Open Letter",* guidata dal Future of Life Institute, un'organizzazione no profit consigliata da Musk.
La lettera tuttavia non richiede l'arresto di tutto lo sviluppo dell'IA , ma esorta le aziende a interrompere temporaneamente i sistemi di formazione più potenti di GPT-4, *la tecnologia rilasciata questo mese dalla startup OpenAI supportata da Microsoft Corp. Ciò include la prossima generazione della tecnologia di OpenAI, GPT-5.*

I funzionari di OpenAI affermano di non aver iniziato ad addestrare GPT-5 *. In un'intervista, il CEO di OpenAI Sam Altman ha affermato che la società ha da tempo dato priorità alla sicurezza nello sviluppo e ha trascorso più di sei mesi a eseguire test di sicurezza su GPT-4 prima del suo lancio. -WSJ*

"In un certo senso, questa è una predica al coro", ha detto Altman. "Penso che abbiamo parlato di questi problemi più forte, più intensamente, più a lungo".

Goldman, nel frattempo, afferma che fino a 300 milioni di posti di lavoro potrebbero essere sostituiti con l'IA, poiché "due terzi delle occupazioni potrebbero essere parzialmente automatizzate dall'IA".

La cosiddetta IA generativa crea contenuti originali basati su suggerimenti umani, una tecnologia che è già stata implementata nel motore di ricerca Bing di Microsoft e in altri strumenti. Poco dopo, Google ha schierato un rivale chiamato Bard . Altre aziende, tra cui Adobe, Salesforce e Zoom, hanno tutte introdotto strumenti avanzati di intelligenza artificiale"

"Oggi inizia una gara", aveva dichiarato con sinistro entusiasmo il CEO di Microsoft Satya Nadella (e chi altro poteva essere?) il mese scorso. "Ci muoveremo e ci muoveremo velocemente."
Uno dei promotori della lettera , Max Tegmark, che dirige il Future of Life Institute ed è professore di fisica al Massachusetts Institute of Technology, la definisce una "corsa suicida".

"È una colpa gravissima inquadrare questo come una corsa agli armamenti", ha detto. "È più una corsa suicida. Non importa chi arriverà per primo. Significa solo che l'umanità nel suo insieme potrebbe perdere il controllo del proprio destino."

Il Future of Life Institute ha iniziato a lavorare alla lettera la scorsa settimana e inizialmente ha permesso a chiunque di firmare senza verifica dell'identità. Ad un certo punto, il nome del signor Altman è stato aggiunto alla lettera, ma successivamente rimosso. Il signor Altman ha detto di non aver mai firmato la lettera. Ha affermato che la società si coordina spesso con altre società di intelligenza artificiale sugli standard di sicurezza e per discutere preoccupazioni più ampie.

*“C’è un lavoro che non facciamo perché **non pensiamo di sapere ancora come renderlo sufficientemente sicuro** “, ha detto. “Quindi sì, penso che ci siano modi in cui puoi rallentare su più assi e questo è importante. E fa parte della nostra strategia”. -WSJ*

Musk, uno dei primi fondatori e sostenitore finanziario di OpenAI, e Wozniak, hanno parlato apertamente dei pericoli dell’IA per un po’ di tempo.

” **Ci sono seri problemi di rischio di intelligenza artificiale** “, ha twittato.

Il capo scienziato AI di Meta, Yann LeCun, non ha firmato la lettera aperta perché dice di non essere d’accordo con la sua premessa (senza elaborare).

Nell’elenco dei firmatari ci sono anche italiani come **Gianluca Bontempi**, professore ordinario di machine learning alla Université Libre de Bruxelles, il ricercatore **Alessandro Perilli**, professore della Orebro University e membro dell’Association for Artificial Intelligence, e **Domenico Talia,** professore dell’Università della Calabria.

Ovviamente alcuni stanno già ipotizzando che i firmatari *possano* avere secondi fini.

> Read this as “a moratorium of six months or more would give me time to copy and/or build something similar and not be left sidelined” https://t.co/MsO0PwkJON
>
> — alpha raccoon (@thealpharaccoon) March 29, 2023

Leggi questo come “una moratoria di sei mesi o più mi darebbe il tempo di copiare e/o costruire qualcosa di simile e non essere lasciato da parte”

Qui sotto un articolo svedese sui pericolo del nuovo sistema:

Come disse una volta Arthur C. Clarke, qualsiasi tecnologia sufficientemente avanzata è “indistinguibile dalla magia”. Alcuni potrebbero dire che questo vale anche per ChatGPT , inclusa, se vuoi, la magia nera.

Subito dopo il suo lancio a novembre, i team di sicurezza, i pen tester e gli sviluppatori hanno iniziato a scoprire exploit nel chatbot AI e continuano a evolversi con la sua ultima iterazione, GPT-4, rilasciata all’inizio di questo mese.

“GPT-4 non inventerà una nuova minaccia informatica”, afferma Hector Ferran, Chief Marketing Officer di BlueWillow AI . “Ma proprio come è già utilizzato da milioni di persone per aumentare e semplificare una miriade di attività quotidiane, potrebbe anche essere utilizzato da una minoranza di cattivi attori per migliorare il proprio comportamento criminale”.

Sviluppo di tecnologie, minacce

A gennaio, appena due mesi dopo il suo lancio, ChatGPT ha raggiunto i 100 milioni di utenti , stabilendo il record per la più rapida crescita degli utenti di un’app. E poiché è diventato un nome familiare, è anche un nuovo brillante strumento per i criminali

informatici, che consente loro di creare rapidamente strumenti e distribuire attacchi.

In particolare, lo strumento viene utilizzato per generare programmi che possono essere utilizzati in attacchi di malware, ransomware e phishing.

BlackFog , ad esempio, ha recentemente chiesto allo strumento di creare un attacco PowerShell in modo "non dannoso". Lo script è stato generato rapidamente ed era pronto per l'uso, secondo i ricercatori.

Vedi anche Report: l'86% dei primi soccorritori desidera una tecnologia di reporting modernizzata

CyberArk , nel frattempo, è stato in grado di aggirare i filtri per creare malware polimorfici, che possono mutare ripetutamente. CyberArk ha anche utilizzato ChatGPT per mutare il codice che è diventato altamente evasivo e difficile da rilevare.

Inoltre, Check Point Research è stata in grado di utilizzare ChatGPT per creare un convincente attacco di spear phishing. I ricercatori dell'azienda hanno anche identificato cinque aree in cui ChatGPT viene utilizzato dagli hacker: malware C++ che raccoglie file PDF e li invia a FTP; phishing che impersona banche; dipendenti di phishing; PHP reverse shell (che avvia una sessione shell per sfruttare le vulnerabilità e accedere al dispositivo di una vittima); e programmi Java che scaricano ed eseguono putty che può essere avviato come PowerShell nascosto.

GPT-4: nuove entusiasmanti funzionalità, rischi

Quanto sopra sono solo alcuni esempi; ce ne sono senza dubbio molti altri che devono ancora essere scoperti o messi in pratica.

"Se diventi molto specifico nei tipi di domande che stai ponendo, è molto facile aggirare alcuni dei controlli di base e generare malware che in realtà è piuttosto efficace", ha affermato Darren Williams, fondatore e CEO di BlackFog. "Questo può essere estrapolato praticamente a qualsiasi disciplina, dalla scrittura creativa all'ingegneria e all'informatica".

E, ha affermato Williams, "GPT-4 ha molte nuove entusiasmanti funzionalità che liberano nuovo potere e potenziali minacce".

Un buon esempio di ciò è il modo in cui lo strumento può ora accettare immagini come input e personalizzarle, ha affermato. Ciò può portare all'uso di immagini incorporate con codice dannoso, spesso denominate "attacchi steganografici".

In sostanza, l'ultima versione è "un'evoluzione di un sistema già potente ed è ancora oggetto di indagine da parte del nostro team", ha affermato Williams.

Vedi anche EU Cyber Resilience Act stabilisce uno standard globale per i prodotti connessi

“Questi strumenti rappresentano alcuni grandi progressi in ciò che l’IA può davvero fare e guidare l’intero settore in avanti, ma come tutta la tecnologia, siamo ancora alle prese con quali controlli devono essere posizionati attorno ad essa”, ha affermato Williams. “Questi strumenti sono ancora in fase di sviluppo e sì, hanno alcune implicazioni sulla sicurezza.”

Più in generale, un’area di preoccupazione è l’uso di ChatGPT per amplificare o migliorare l’attuale diffusione della disinformazione, ha affermato Ferran.

Tuttavia, ha sottolineato, è fondamentale riconoscere che l’intento dannoso non è esclusivo degli strumenti di intelligenza artificiale.

“ChatGPT di per sé non rappresenta una minaccia alla sicurezza”, ha affermato Ferran. “Tutta la tecnologia ha il potenziale per essere utilizzata nel bene o nel male. La minaccia alla sicurezza proviene da malintenzionati che utilizzeranno una nuova tecnologia per scopi dannosi”.

In poche parole, ha detto Ferran, “la minaccia deriva da come le persone scelgono di usarla”.

In risposta, gli individui e le organizzazioni dovranno diventare più vigili e controllare più da vicino le comunicazioni per cercare di rilevare gli attacchi assistiti dall’intelligenza artificiale, ha affermato. Devono inoltre adottare misure proattive per prevenire gli abusi implementando adeguate salvaguardie, metodi di rilevamento e linee guida etiche.

“In questo modo, possono massimizzare i vantaggi dell’IA riducendo i potenziali rischi”, ha affermato.

Affrontare le minacce richiede anche uno sforzo collettivo da parte di più parti interessate. “Lavorando insieme, possiamo garantire che ChatGPT e strumenti simili vengano utilizzati per una crescita e un cambiamento positivi”, ha affermato Ferran.

E sebbene lo strumento disponga di filtri di contenuto per prevenire gli abusi, questi possono chiaramente essere gestiti abbastanza facilmente, quindi “potrebbe essere necessario esercitare pressioni sui suoi proprietari per migliorare queste salvaguardie”, ha affermato.

Vedi anche Come i token crittografici sono diventati insicuri come lo erano una volta le carte di pagamento

Anche la capacità di sicurezza informatica è buona

D’altra parte, ChatGPT e altri strumenti avanzati di intelligenza artificiale possono essere utilizzati dalle organizzazioni sia per funzioni offensive che difensive.

“Fortunatamente, l’intelligenza artificiale è anche uno strumento potente che può essere utilizzato contro i cattivi attori”, ha affermato Ferran.

Le società di sicurezza informatica, ad esempio, utilizzano l'intelligenza artificiale nei loro sforzi per trovare e catalogare le minacce dannose.

"La gestione delle minacce informatiche dovrebbe cogliere ogni opportunità per sfruttare l'IA nello sviluppo di misure preventive", ha affermato Ferran, "in modo che possano vincere quella che potrebbe essenzialmente essere una corsa agli armamenti".

E, con le sue protezioni potenziate e la capacità di rilevare comportamenti dannosi, può in ultima analisi diventare una "risorsa potente" per le organizzazioni.

"GPT-4 è un notevole passo avanti nei modelli basati sul linguaggio naturale, ampliando notevolmente i suoi potenziali casi d'uso e basandosi sui risultati delle sue precedenti iterazioni", ha affermato Ferran, indicando la sua capacità ampliata di scrivere codice in qualsiasi lingua, ha affermato .

Williams ha concordato, affermando che l'intelligenza artificiale è come qualsiasi strumento potente: le organizzazioni devono fare la propria due diligence.

"Ci sono rischi che le persone possano usarlo per scopi nefasti? Certo, ma i benefici superano di gran lunga i rischi", ha detto.