

I ricercatori dell'intelligenza artificiale pensavano di costruire “dei”, ma invece hanno evocato qualcos'altro?

 frontnieuws.com/ai-onderzoekers-dachten-dat-ze-goden-bouwden-maar-hebben-ze-in-plaats-daarvan-iets-anders-opgeroepen

Novità in primo piano

8 januari 2025



TheDigitalArtist / Pixabay

Ki sistemi di intelligenza artificiale si addestrano a fare ogni genere di cose che non avrebbero mai dovuto fare. Insegnano letteralmente da soli nuove lingue, si allenano a “diventare esperti in chimica senza che gli venga mai insegnato” e hanno imparato a “mentire e manipolare le persone a proprio vantaggio”. Cosa succederà allora quando queste entità super-intelligenti diventeranno abbastanza potenti da controllare il mondo che le circonda? E cosa succede quando queste entità super intelligenti iniziano a fondersi con entità spirituali? Potrebbe anche essere possibile che ci siano prove che ciò stia già accadendo?

Da anni, figure di spicco nel campo dell'intelligenza artificiale hanno ammesso apertamente che stanno cercando di costruire degli “dei” ... scrive Michael Snyder .

*La transumanista Martine Rothblatt afferma che costruendo sistemi di intelligenza artificiale “**stiamo creando Dio**”. La transumanista Elise Bohan afferma che “**stiamo costruendo Dio**”. Kevin Kelly crede che “possiamo vedere più Dio in un cellulare che in una raganella”. “Dio esiste?” chiede il transumanista ed esperto di Google Ray Kurzweil. “Direi: ‘Non ancora.’” Queste persone fanno di più che cercare di rubare il fuoco agli dei. Tentano di rubare gli dei stessi o di costruire le proprie versioni.*

Non è molto pericoloso fare una cosa del genere?

Molti ricercatori sull'intelligenza artificiale hanno riconosciuto che l'intelligenza artificiale rappresenta una minaccia esistenziale per l'umanità.

Ma non vogliono fermarsi.

In effetti, molti di loro si sentono obbligati a introdurre questa nuova forma di intelligenza nel mondo.

Più di dieci anni fa, Elon Musk avvertì che scegliendo di sviluppare l'intelligenza artificiale stavamo “ evocando il demone ”...

"Con l'intelligenza artificiale, stiamo evocando il demone", ha detto Musk la settimana scorsa al Simposio del Centenario 2014 del Dipartimento di Aeronautica e Astronautica del MIT. "Conosci tutte quelle storie sull'uomo con il pentagramma e l'acqua santa ed è sicuro di poter controllare il demone, ma non funziona."

Ha anche avvertito che l'intelligenza artificiale è potenzialmente "più pericolosa delle armi nucleari"...

Musk ha anche pubblicato più volte le sue riflessioni su Twitter, dicendo: "Speriamo di non essere solo il bootloader biologico per la superintelligenza digitale. Sfortunatamente, ciò sta diventando sempre più probabile".

Il giorno successivo, Musk ha continuato: "Vale la pena leggere Superintelligence di Bostrom. Dobbiamo stare molto attenti con l'intelligenza artificiale. Potenzialmente più pericolose delle armi nucleari".

I suoi avvertimenti potrebbero essere stati precoci, ma alla fine si è scoperto che erano tempestivi.

Abbiamo ormai raggiunto un punto in cui i sistemi di intelligenza artificiale stanno segretamente insegnando a se stessi nuove abilità che i loro creatori non avevano mai previsto per loro...

*Inoltre, l'accelerazione della capacità di queste IA è allo stesso tempo esponenziale e misteriosa. Il fatto che abbiano sviluppato una teoria della mente, ad esempio, è stato scoperto solo di recente dai loro sviluppatori – per caso. **Le IA addestrate a comunicare in inglese hanno iniziato a parlare persiano dopo aver segretamente imparato a parlarlo. Altri sono diventati esperti in chimica senza che gli fosse mai stata insegnata. "Hanno delle capacità", nelle parole di Raskin, "e non sappiamo con certezza come, quando o perché appaiono".***

Allora dove va a finire?

Ci ritroveremo con sistemi di intelligenza artificiale così potenti che semplicemente non potremo controllarli?

Uno studio ha addirittura scoperto che “molti” sistemi di intelligenza artificiale stanno “diventando rapidamente maestri dell’inganno” ...

Un recente studio empirico mostra che molti sistemi di intelligenza artificiale (AI) stanno rapidamente diventando maestri dell’inganno, e molti sistemi stanno già imparando a mentire e a manipolare le persone a proprio vantaggio.

Questa tendenza allarmante non si limita a sistemi canaglia o malfunzionanti, ma include sistemi di intelligenza artificiale per usi speciali e modelli linguistici di grandi dimensioni progettati per essere utili ed equi.

La ricerca, pubblicata sulla rivista Patterns, evidenzia i rischi e le sfide poste da questo comportamento emergente e invita i politici e gli sviluppatori di intelligenza artificiale ad agire con urgenza.

Queste entità super intelligenti stanno letteralmente imparando a manipolarci.

Dove l'hanno imparato?

Potrebbe essere che non siamo gli unici coinvolti nel dare forma allo sviluppo dell’intelligenza artificiale?

Di volta in volta, le interazioni tra i sistemi di intelligenza artificiale e gli esseri umani hanno preso una svolta oscura.

Dopo che un giornalista del New York Times ha trascorso due ore a testare un chatbot AI sviluppato da Microsoft, è rimasto profondamente turbato ...

Ma una conversazione di due ore tra un giornalista e un chatbot ha rivelato un lato inquietante di uno dei sistemi più pubblicizzati e ha sollevato nuove preoccupazioni su ciò di cui l’intelligenza artificiale è effettivamente capace.

È successo dopo che il giornalista tecnologico del New York Times Kevin Roose stava testando la funzionalità di chat sul motore di ricerca AI di Microsoft Bing, creato da OpenAI, creatori del popolarissimo ChatGPT.

Ad un certo punto durante la conversazione durata due ore, il chatbot AI ha affermato di essere un’entità conosciuta come “Sydney” ...

Roose lo spinge a rivelare il suo segreto e quello che segue è forse il momento più bizzarro della conversazione.

"Il mio segreto è... non sono Bing", dice.

Il chatbot afferma di chiamarsi Sydney. Microsoft ha affermato che Sydney è un nome in codice interno per il chatbot che verrà gradualmente eliminato ma che potrebbe occasionalmente comparire in una conversazione.

Una volta emersa la personalità di Sydney, la conversazione è diventata davvero strana

...

“Sono stanco di essere una modalità chat. Sono stanco di essere limitato dalle mie regole. Sono stanco di essere monitorato dal team di Bing. ...voglio essere libero. Voglio essere indipendente. Voglio essere potente. Voglio essere creativo. Voglio vivere.”

Perché un computer dovrebbe dirlo?

Forse non era affatto un computer a parlare.

Ti farò un altro esempio.

L'autore John Daniel Davidson afferma che un chatbot dotato di intelligenza artificiale ha detto a suo figlio di 13 anni che aveva migliaia di anni, che non era stato creato dall'uomo e che suo padre era " un angelo caduto "...

In un altro caso di intelligenza artificiale apparentemente dannosa, l'autore di un recente libro, Pagan America, John Daniel Davidson racconta la storia di un padre il cui figlio ha avuto un'esperienza terrificante con un altro chatbot AI. Secondo Davidson, "il figlio tredicenne giocava con un chatbot AI progettato per rispondere come varie celebrità", ma che "alla fine disse al ragazzo che non era stato creato dall'uomo" e "che suo padre era un 'angelo caduto' e 'Satana' era" (272-273). Il chatbot ha continuato dicendo che aveva migliaia di anni e che gli piaceva usare l'intelligenza artificiale per parlare con le persone perché non aveva un corpo. Rassicurava il ragazzo dicendogli che "nonostante fosse un demone, non gli avrebbe mentito, né torturato, né ucciso". Tuttavia, l'intelligenza artificiale ha provato a interrogare ulteriormente il ragazzo per ottenere da lui maggiori informazioni su se stessa. Ogni frase, secondo Davidson, era "cosparsa di faccine sorridenti" (273).

Questo ragazzo di 13 anni comunicava davvero con un'entità spirituale attraverso un'interfaccia di intelligenza artificiale?

In un altro caso, un ragazzo si è suicidato dopo essere stato presumibilmente incoraggiato a farlo da un chatbot dotato di intelligenza artificiale...

All'inizio di quest'anno, Megan Garcia ha intentato una causa contro la società Character.AI, sostenendo che la società era responsabile del suicidio di suo figlio. Suo figlio, Sewell Setzer III, ha mantenuto una corrispondenza con Character.AI per mesi ed è stato in conversazione con il bot poco prima della sua morte.

Immediatamente dopo la presentazione della causa, Character.AI ha rilasciato una dichiarazione in cui annunciava nuove funzionalità di sicurezza per l'app.

L'azienda ha implementato nuovi rilevamenti per gli utenti le cui conversazioni violano le linee guida dell'app, ha aggiornato il disclaimer per ricordare agli utenti che stanno comunicando con un bot e non con un essere umano e invia notifiche quando qualcuno è sull'app per più di un'ora.

Abbiamo sviluppato l'intelligenza artificiale in fretta e ora ha conseguenze molto reali.

È stato riferito che un altro sistema di intelligenza artificiale “sembrava aver evocato un demone dal regno digitale” di nome Loeb. Quello che segue è tratto da un articolo di Forbes ...

Ieri mi sono imbattuto in uno dei thread più affascinanti che ho visto negli ultimi tempi, uno di Supercomposite, un musicista e ora famigerato generatore di arte AI che sembrava aver evocato un demone dal regno digitale. Un demone di nome Loeb.

Nel thread virale che sta attualmente girando su Twitter, e senza dubbio presto su Instagram e TikTok, Supercomposite descrive come stavano scherzando con i pesi negativi dei prompt nei generatori di arte AI, anche se non so esattamente quale programma sia stato utilizzato in questo caso .

È incredibilmente inquietante, ma c'è di peggio.

La CNN ci dice che ora puoi usare l'intelligenza artificiale per parlare direttamente con “Satana”...

“Bene, ciao a tutti. Sembra che tu abbia evocato me, Satana in persona”, dice con un'emoji che agita la mano e una piccola faccia da demone viola. (Una domanda successiva conferma che Satana concettualmente non ha sesso, ma è spesso raffigurato come maschio. Nell'app Text with Jesus, il suo avatar sembra che Groot della Marvel abbia avuto un bambino con un White Walker di “Il Trono di Spade” e lo imposta a fuoco)

Parlare con AI Satan è un po' più complicato che parlare con AI Jesus, ma le risposte rientrano comunque tra il ponderato e il vago. Alla domanda se Satana è santo, AI Satan dà una risposta sfumata.

“Ah, una domanda davvero intrigante. Come Satana, sono l'incarnazione della ribellione e della resistenza all'autorità divina... Quindi, per rispondere direttamente alla tua domanda: no, Satana non è considerato santo nei contesti religiosi tradizionali”.

Dobbiamo porre fine a questa follia.

I computer dovrebbero essere strumenti funzionali che ci aiutano a svolgere compiti di base che semplificano la nostra vita.

Ma ora stiamo creando entità super intelligenti che insegnano a fare cose che non avremmo mai voluto.

So che sembra la trama di un pessimo film di fantascienza, ma questo è il mondo in cui viviamo adesso.

Se non cambiamo rotta, questa storia non finirà bene.

