Tiratura: 230.062 Diffusione: 261.691

CORRIERE DELLA SERA
laLettura

Settimanale

RUBBETTINO

Data 03-09-2023
Pagina 12
Foglio 1 / 2

Macchine pensanti ma a modo loro

di NELLO CRISTIANINI

«Alcuni sistemi di Intelligenza artificiale utilizzati per interagire con persone fisiche o per generare contenuti possono comportare rischi specifici di impersonificazione o inganno (...). Le persone fisiche dovrebbero (...) ricevere una notifica nel momento in cui interagiscono con un sistema di IA, a meno che tale interazione non risulti evidente dalle circostanze e dal contesto di utilizzo». Questo recente emendamento alla proposta di legge europea «AI Act» fa riflettere più dei tanti annunci entusiastici, o petizioni allarmistiche, sull'Intelligenza artificiale di questi giorni.

Per decenni il Santo Graal della ricerca nell'IA è stato il «Test di Turing», un esame concepito nel 1950 da Alan Turing per stabilire se una macchina fosse «pensante», e da lui chiamato «il gioco dell'imitazione». In tale gioco, la macchina conversa per iscritto con un essere umano, rispondendo a domande su ogni argomento, e ha l'obiettivo di ingannarlo, facendogli credere di essere una persona. Di converso, l'intervistatore sceglie le domande per smascherare la macchina.

Ci sono stati innumerevoli tentativi, a partire dagli anni Sessanta, di creare algoritmi che potessero passare tale test, e per circa un ventennio, a partire dal 1990, c'è stata anche una competizione annuale chiamata il «Premio Loebner». Nessuno ha mai vinto: l'intervistatore ha sempre capito quali dei suoi interlocutori fossero umani e quali fossero gli algoritmi.

Tenere una conversazione credibile su qualsiasi argomento pone almeno due sfide, una che riguarda il linguaggio e l'altra la conoscenza del mondo e delle sue regole, ed entrambe sono direzioni attive di ricerca da decenni. Oggi però è cambiato qualcosa: con l'introduzione di una nuova classe di sistemi intelligenti, chiamati Large Language Models e a cui appartiene ChatGPT, è possibile entrare in dialoghi lunghi e interessanti, e mantenere l'impressione non solo che ci sia una persona dall'altra parte, ma anche una persona ben istruita e sensata. Non so di alcun esame formale passato da ChatGPT con il rituale stretto del «Premio Loebner», ma mi sembra chiaro che nel 2023 abbiamo passato il «test di Turing», o ci manca pochissimo.

L'emendamento aggiunto all'AI Act, che obbliga ogni intelligenza artificiale a dichiararsi tale, lascia vedere un giorno in cui non saremo in grado di distinguerne i comportamenti dai nostri, così come siamo già incapaci di individuare le immagini false create da un computer. In altre parole: se la legge deve obbligare un algoritmo a dichiararsi tale, è chiaro che siamo giunti al punto in cui potrebbe farsi passare per un essere umano.

Che cosa significa avere un agente intelligente in grado di superare l'*imitation game*? Turing direbbe che dovremmo considerarlo un essere «pensante» a tutti gli effetti. Ma una domanda che si sente spesso riguardo a ChatGPT è se questa macchina capisce veramente quello che diciamo e il mondo, o se solo si comporta come se li capisse. Non entrerò nei dettagli della differenza tra «comprendere veramente» e «apparire di comprendere», ma noto che in genere queste obiezioni si riferiscono alla comprensione di tipo umano. Ovvero ci si chiede se la macchina sia capace di comprendere il mondo nel senso in cui lo comprendiamo noi.

Una cosa è chiara: il solo fatto che la macchina sia addestrata mediante relazioni statistiche scoperte nei dati non implica che non sia in grado di comprendere il mondo. Anche i nostri neuroni hanno un comportamento probabilistico, ma organizzati nel modo e nei numeri giusti possono comunque produrre una comprensione del mondo.

La mia opinione è che ChatGPT non potrebbe rispondere a certe domande senza avere qualche forma di comprensione, ovvero una rappresentazione astratta delle proprietà del mondo, ma che non c'è alcun motivo per credere che questa corrisponda alla nostra: perché dovrebbe? Possiamo immaginare degli esami psicometrici, come quelli usati per misurare le caratteristiche psicologiche degli esseri umani, dedicati ad algoritmi come ChatGPT: potremmo misurarne le capacità di ragionamento verbale («se ho tre mele e ne mangio due, quante mele rimangono?»), quelle di ragionamento spaziale, e così via. Studi recenti hanno trovato che la macchina può rispondere alle domande di esami medici e giuridici con qualità quasi umana.

Quello che stiamo scoprendo è che la macchina è in grado di imparare dei compiti che non le abbiamo insegnato, e che questo avviene solo dopo che ha visto una quantità sufficiente di dati, in modo non graduale: le chiamiamo «abilità emergenti», e sono oggetto di studio intenso al momento. Tra le abilità emergenti più interessanti ci sono quelle di risolvere indovinelli, completare sillogismi e fare uso di analogie. Non sappiamo ancora quali altre abilità potrebbero emergere, aggiungendo più dati.

È interessante notare che la nostra prima idea, di risolvere separatamente la comprensione del linguaggio e la conoscenza del mondo, per consentire alla macchina di conversare su qualunque argomento, non si è realizzata: gli algoritmi alla base di GPT rappresentano nozioni linguistiche e fatti del mondo senza distinzione al loro interno, e non hanno bisogno di conoscerne la differenza. Una frase incorretta grammaticalmente o semanticamente è trattata allo stesso modo dal sistema. Forse queste macchine ci insegneranno un giorno dei modi nuovi di pensare anche all'intelligenza umana?

Che cosa ci resterà da fare, come comunità di ricerca, dopo avere «superato il test di Turing»? Una cosa importante è assicurarsi che le leggi come quella menzionata sopra siano approvate, senza essere diluite. Al momento ChatGPT è espressamente addestrato per non ingannare o confondere l'utente, o impersonare un essere umano, e già così molte persone hanno la sensazione di parlare con qualcuno. Se fosse espressamente addestrato per l'obiettivo opposto, ovvero per creare l'impressione di avere una coscienza, dei sentimenti, delle ambizioni, l'illusione sarebbe probabilmente molto potente, e così anche la possibilità di manipolazione ed effetti indesiderati.

NELLO CRISTIANINI
La scorciatoia. Come le macchine sono diventate intelligenti senza pensare in modo umano
IL MULINO
Pagine 216, € 16

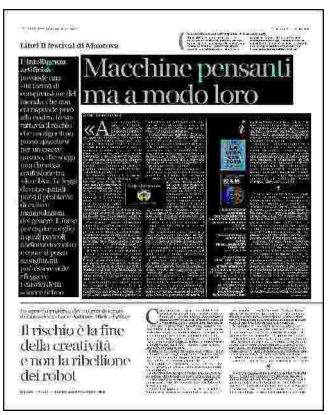
RICCARDO MANZOTTI
SIMONE ROSSI
Io & IA. Mente, cervello & GPT
RUBBETTINO
Pagine 176, € 15

L'appuntamento
Nello Cristianini (Gorizia, 1968) interverrà al Festivaletteratura di Mantova domenica 10 (ore 17.15) presso l' Aula Magna dell'università: discuterà con Silvia Bencivelli

Le illustrazioni
Nella pagina a fianco, il designer Enzo Mari in due interpretazioni dell'artista Fausto Gilberti

laLettura
CORRIERE DELLA SERA
Settimanale
RUBBETTINO
Data 03-09-2023
Pagina 12
Foglio 2 / 2

L'**Intelligenza artificiale** possiede una sua forma di comprensione del mondo, che non corrisponde però alla nostra. Resta tuttavia il rischio che un algoritmo possa spacciarsi per un essere umano, che sorga una dannosa confusione tra «Io e IA». Le leggi devono quindi porsi il problema di evitare manipolazioni del genere. E forse per capire meglio a quali pericoli andiamo incontro e come si possa scongiurarli può essere utile rileggere i classici della science fiction


Macchine pensanti ma a modo loro

Il rischio è la fine della creatività e non la ribellione dei robot