Avvenire

diffusione:94524
tiratura:125758

Dominio IA, come evitare un futuro poco desiderabile

ANDREA **LAVAZZA**

L'intelligenza artificiale generativa fa miracoli, ma il suo uso poco accorto comincia a produrre guai seri. Ne sanno qualcosa i magistrati americani. Le allucinazioni dell'IA stanno comparendo sempre più spesso nei documenti legali. Qualche settimana fa, come riferisce la Mit Technology Review, il giudice della California Michael Wilner è rimasto incuriosito da una serie di argomentazioni presentate da alcuni avvocati. Ha cercato di saperne di più seguendo gli articoli citati, ma ha scoperto che quegli articoli non esistevano. Ha chiesto maggiori dettagli al noto e finora prestigioso studio legale coinvolto, che ha risposto con una nuova memoria contenente ancora più errori della prima. Wilner ha ordinato così di fornire testimonianze giurate che spiegassero l'accaduto. Ha quindi appreso che era stato utilizzato Google Gemini insieme con modelli di IA specifici per il diritto al fine di redigere la memoria difensiva, generando informazioni false. Il giudice ha multato lo studio legale per 31.000 dollari.

Torna d'attualità, con casi come questo, la critica di Gerd Gigerenzer all'eccessiva fiducia nell'intelligenza artificiale, sviluppata nel libro *Perché l'intelligenza umana batte ancora gli algoritmi* (Cortina Editore, pagine 376, euro 26,00). Lo psicologo tedesco contesta l'idea diffusa che l'IA, in particolare il machine learning (imparare dai dati senza addestramento umano diretto), possa essere applicato in modo efficace in contesti non strutturati o incerti. A suo parere, molti successi delle macchine sono limitati a contesti ristretti, ben definiti, come giochi (scacchi, Go) o il riconoscimento di modelli in grandi quantità di informazioni quantitative. In ambienti reali e incerti, dove i dati sono scarsi o imprecisi, l'intelligenza umana e le euristiche semplici (regole intuitive e veloci, imperfette ma razionali) sono spesso più efficaci.

Molti algoritmi operano assumendo che il mondo sia prevedibile e che le decisioni si possano ottimizzare con sufficiente potenza computazionale. Gigerenzer, invece, distingue tra rischio (dove le probabilità sono note) e incertezza (dove le probabilità non sono conoscibili in anticipo). Dato che l'incertezza domina la maggior parte delle decisioni umane importanti, in tali casi l'IA fallisce, mentre gli esseri umani possono ricorrere a strategie euristiche robuste e adattive. Gigerenzer mette anche in guardia contro la "trasparenza ingannevole" di molti sistemi artificiali, che non spiegano i propri criteri decisionali. Questo porta a una dipendenza cieca dagli algoritmi, con le relative conseguenze.

> Ecco come prendere sul serio i sistemi algoritmici: senza accodarsi all'ottimismo ma sottolineando invece i problemi (economici, sociali e politici) sollevati dalla rapidissima diffusione dei chatbot di IA generativa

Le riserve circostanziate di Gigerenzer nei confronti dell'IA risultano oggi più persuasive agli occhi degli esperti rispetto a quelle classiche mosse dal filosofo americano John Searle, utilmente riepilogate nei saggi contenuti nel volume *Intelligenza artificiale e pensiero umano* (pagine 210, euro 20,00) pubblicato da Castelvecchi. In questi sei noti testi, Searle (oggi novantaduenne) espone la sua critica all'idea che l'intelligenza artificiale possa replicare la mente umana. La tesi centrale è che i computer, pur potendo simulare comportamenti intelligenti, non possiedono coscienza né afferrano i significati. Se si distingue tra sintassi (la manipolazione formale di simboli, tipica dei computer) e semantica (il significato e la comprensione, proprie della mente umana), sostiene il grande filosofo, l'esecuzione di un programma non è sufficiente per generare stati mentali dotati di significato. Ciò è illustrato nel celebre esperimento mentale della Stanza Cinese. Searle immagina una persona chiusa in una stanza che, grazie a istruzioni nella sua lingua, manipola correttamente simboli cinesi senza comprenderli. Da fuori sembra sapere il cinese, in realtà non lo capisce.

La mente umana, secondo tale approccio, è caratterizzata da intenzionalità, ovvero la capacità di riferirsi a contenuti e significati, capacità che i programmi informatici non possono riprodurre e che viene interpretata come fenomeno biologico emergente dalle proprietà fisiche del cervello. Sebbene Searle abbia ancora un ruolo da "coscienza critica" nel dibattito sull'IA, le sue tesi non offrono più un fondamento solido per negare in assoluto la possibilità di menti artificiali. I progressi dell'IA e le nuove teorie


AGORÀ
Il messaggio evangelico nella cultura digitale

Avvenire
diffusione:94524
tiratura:125758

della mente hanno infatti spostato rapidamente il terreno del confronto.

Bisogna quindi calarsi in una discussione puntuale dei nuovi agenti artificiali, come fa anche l'agile volume *Critica di ChatGPT*, scritto da Antonio Santangelo, Alberto Sissa e Maurizio Borghi (Elèuthera, pagine 160, euro 15,00) quale esito di un progetto di ricerca promosso dal Politecnico e dall'Università di Torino. L'idea degli autori è di prendere sul serio i sistemi algoritmici senza però accodarsi all'ottimismo diffuso e, anzi, sottolineando i problemi economici, sociali e politici sollevati dalla rapidissima diffusione dei chatbot di IA generativa. Il loro impatto va dal depauperamento ambientale a causa del forte bisogno di energia ai temi della diversità e dell'equità, passando dagli effetti psicologici sugli utilizzatori. Ampio spazio agli aspetti legali, dalla protezione dei dati al diritto d'autore e le regolamentazioni nazionali e sovranazionali. Per concludere che bisogna agire, per evitare "un futuro poco desiderabile".