CHATGPT: NON UN ORACOLO MA UN ASSISTENTE SGOBBONE

ChatGpt e gli altri sistemi basati su Large language quando non trovano l'informazione richiesta inventano di sana pianta: dati scientifici, diagnosi mediche, resoconti storici, trame letterarie e informazioni biografiche.

Fabrizio Tonello

Ho scritto a ChatGPT: “La mia gattina Cip è sdraiata: dove?” e ChatGPT ha correttamente risposto "sul pavimento". Meraviglie dell'intelligenza artificiale o sorveglianza onnipresente, per cui ChatGPT sa non solo dove sono io ma anche dov'è la mia gattina?

Niente di tutto questo: il suo modello di funzionamento è statistico, non "intelligente" (differenza che, a quanto pare, politici e giornalisti italiani non comprendono). Quindi ChatGPT prevede che le parole che hanno maggiore probabilità di venire dopo "sdraiaita" sono "sul" oppure “per” e che quelle che probabilmente seguono sono “terra” o "pavimento". Non sa però che cosa sia un gatto, un pavimento o l’essere sdraiati perché fa un caldo boia. Quindi se nel calcolo statistico qualcosa non funziona non ha modo di capirlo autonomamente.

I *Large language model* si limitano a prevedere statisticamente quale parola abbia la maggior probabilità di essere coerente con quelle che l’hanno preceduta, basandosi sui database usati per l'addestramento ma senza ciò che noi *homo sapiens* definiremmo "vera conoscenza" dell'argomento. (Il che fa anche pensare che prima che AI decida di eliminare gli umani e dominare la terra ci voglia un po' di tempo: qualcuno lo spieghi a Valditara).

L'ultimo numero della *MIT Technology Review* del Massachusetts Institute of Technology ha segnalato il fatto che alcuni mesi fa l’Organizzazione mondiale della sanità aveva lanciato il *chatbot* Sarah, basato sul Gpt 3.5 di OpenAI (quello più vecchio, ce ne sono di più recenti). Lo scopo di Sarah (Smart AI Resource Assistant for Health) era di fornire consigli in diverse lingue su temi importanti per la sanità, in particolare attività fisica, mangiare sano, fumo e salute mentale.

Purtroppo, nel giro di poche settimane sono iniziate le proteste per i pessimi consigli ricevuti da Sarah, che conosceva l'Amazon di Jeff Bezos ma non l'Amazon foresta brasiliana e quindi non capiva il problema del disboscamento dell’Amazzonia. Ignara della situazione della sanità a San Francisco (e di come si usa un elenco telefonico), Sarah ha poi fornito nomi e numeri di telefono di cliniche inesistenti in città. Il chatbot scientifico *Galactica* di Facebook ha avuto una vita ancora più breve perché scriveva articoli accademici e pagine wiki sulla storia degli orsi nello spazio.

ChatGpt e gli altri sistemi basati su *Large language model* forniscono risposte a richieste poste in linguaggio naturale pescando dai database a loro disposizione. Quando non trovano l'informazione richiesta inventano di sana pianta: dati scientifici, diagnosi mediche, resoconti storici, trame letterarie e informazioni biografiche. Di tutti questi problemi OpenAi, Meta, Microsoft e gli altri protagonisti del settore sono ben al corrente: il sito di ChatGPT avverte: "Può commettere errori. Verifica le informazioni importanti". Un avvertimento che la *MIT Technology Review* illustra in questo modo: "inventare cose è esattamente ciò per cui questi modelli sono progettati".

Finché il problema delle invenzioni (che nel gergo degli esperti del settore vengono chiamate "allucinazioni") non sarà risolto, integrare i *Large language model* nei motori di ricerca continuerà a esporre gli utenti a informazioni sbagliate, inaccurate o addirittura pericolose. Gli ottimisti giurano che si tratta solo di aspettare i prossimi miglioramenti ma è davvero così? In realtà ci sono già vari studi accademici che dimostrano come le allucinazioni siano un limite intrinseco dei modelli linguistici."Fare in modo che un *chatbot* sia corretto nel 90% dei casi è abbastanza facile, ma fare in modo che sia corretto nel 99% dei casi è un enorme problema di ricerca non ancora risolto" dice, per esempio, Yonadav Shavit di Harvard. E, come spiega Daniel Andler, "Inserire tecnologie i cui processi non controlliamo nel nostro spazio pubblico significa rinunciare a valutare il loro impatto".

"Nonostante i tentativi di Big Tech di convincerci del contrario, i *Large language model* non sono e forse non saranno mai degli oracoli in grado di rispondere correttamente alle nostre domande. Sono invece molto più simili a degli assistenti sgobboni ma assolutamente inaffidabili. Il cui lavoro va verificato con estrema attenzione" ha scritto Andrea Signorelli sul *Domani*, qualche tempo fa.

Letture utili

Will Douglas Heaven, “Why Does AI hallucinate?”, *MIT Technology Review*, 18 giugno 2024.
Daniel Andler, *Il duplice enigma. Intelligenza artificiale e intelligenza umana*, Einaudi, 2023.
Gino Roncaglia, *L’architetto e l’oracolo. Forme digitali del sapere da Wikipedia a ChatGPT*, Laterza, 2023.

FABRIZIO TONELLO

è docente di Scienza Politica presso l’Università di Padova, dove insegna, tra l’altro, un corso sulla politica estera americana dalle origini ad oggi. Ha insegnato alla University of Pittsburgh e ha fatto ricerca alla Columbia University, oltre che in Italia (alla SISSA di Trieste e all’Università di Bologna). Ha scritto *Democrazie a rischio. La produzione sociale dell’ignoranza (Pearson, 2019), L’età dell’ignoranza* (Bruno Mondadori 2010), *Il Nazionalismo americano* (Livian, 2007), *La politica come azione simbolica (Franco Angeli, 2003). Da molti anni collabora alle pagine culturali del Manifesto.*