

Le falsità tutte umane dell'intelligenza artificiale

Gilberto Corbellini

Andrea Bassi, ieri sul Messaggero, ha raccontato con quanta falsificazione Copilot, l'intelligenza Artificiale di Microsoft, ricostruisce e interpreta la politica italiana sul nucleare a valle del referendum abrogativo del 2011, dimostrando per l'ennesima volta che queste macchine possono mentire in quanto usano documenti prodotti dall'uomo. La macchina intelligente commenta che le ricadute del rifiuto referendario del nucleare in Italia sono state positive. Salvo scoprire che tra le basi dati usate dal generatore (...)

Continua a pag. 16

stralciato da
IL
MESSAGGERO

pagg. 1 e 16
del
05/05/24

Gilberto Corbellini*

segue dalla prima pagina

(...) di linguaggi ci sono i testi di Greenpeace, ovvero dell'organizzazione che più si è spesa per la propaganda antinucleare. Perché le intelligenze artificiali possono dire cose false, inventare informazioni inesistenti o difendere tesi ideologiche, anche se in genere quando sono interrogate sulle questioni controverse tendono a esprimersi in modi cerchiobottisti?

Le ultime generazioni di intelligenza artificiale sono macchine statistiche che non hanno contezza di cosa dicono, e sono molto affidabili solo in campi molto specialistici o a basso tasso di ambiguità semantica del lessico. Le chatbot con cui dialoghiamo estraggono dati fatti loro digerire, informazioni, nozioni, frasi, parole per interagire con noi facendoci credere che stiamo parlando con qualcuno che ha il controllo di quello che dice. I bias e gli incidenti poco politicamente corretti che queste macchine causano sono quasi memorabili, come quando le applicazioni di una IA al sistema giudiziario americano produsse il risultato di negare ai detenuti di colore la libertà condizionale o stimarne in eccesso la pericolosità con una frequenza altissima, e non giustificata dai dati contenuti nel casellario. Le cose migliorano nelle prestazioni di queste intelligenze artificiali quanto più ampia è la quantità di testi con cui lavorano (centinaia di miliardi) e quanto più numerosi sono gli strati della rete neurale (ChatGPT ne ha circa 100).

L'intelligenza artificiale lavora comunque con dati che sono stati raccolti da persone umane, le quali li hanno categorizzati usando regolarmente dei pregiudizi. Come ha dimostrato Daniel Kahnemann, psicologo e premio Nobel morto da poco, larga parte delle nostre decisioni sono il risultato di ragionamenti non ottimizzati, ovvero di inclinazione spontanee che portano a decidere o scegliere, usando come criterio principale non la logica, le prove, ma il soddisfacimento psicologico. Si consideri il sondaggio effettuato in Russia, di cui parla il Corriere della Sera di ieri, per cui solo l'11% dei Russi pensa che l'attentato al Crocus City Hall sia stata opera dell'Isis, malgrado la rivendicazione dell'organizzazione terroristica islamica, mentre il 50% di loro ritiene che si sia trattato di un attentato organizzato dai servizi segreti ucraini.

Non è che le falsità o le disinformazioni circolano perché non le sappiamo riconoscere, ma perché le preferiamo e quindi le scambiamo più rapidamente. Noi preferiamo una bugia rassicurante a una verità scomoda, un concetto che un predicatore britannico di metà Ottocento esprimeva dicendo che «mentre la verità di allaccia le scarpe, la bugia ha girato mezzo mondo».

Uno studio pubblicato su Science nel 2018 ha dimostrato che sulla piattaforma ex-Twitter, le informazioni false circolavano più rapidamente di quelle vere. Le persone, nei social media, inoltrano le falsità più delle verità e ciò causa la diffusione di false notizie, malgrado tutte le dichiarazioni o i tentavi sociali e individuali di pri-

vilegiare la verità. Inoltre, non sono i robot, ma i comportamenti umani che contribuiscono al differenziale di diffusione di falsità e verità. L'articolo concludeva che servono interventi per dissuadere soprattutto la diffusione della misinformazione, vale a dire l'informazione appositamente creata per fuorviare i destinatari. Non dimentichiamo che nel 2018, l'American Journal of Public Health dimostrava che i bot di twitter e i troll russi avevano negli anni reso conflittuale con disinformazioni mirate la comunicazione sui vaccini negli Stati Uniti.

Gli scienziati, gli esperti, gli intellettuali e i politici ma anche le persone comuni di norma credono che i modi tradizionalmente proposti per combattere la disinformazione debbano basarsi sul presupposto che, se le persone incontrano le informazioni corrette su un argomento, prenderanno decisioni razionali. Non è così, neppure sul fronte degli esperti, come ha tristemente dimostrato lo spettacolo mediatico a cui abbiamo assistito durante la recente pandemia. Noi non siamo spontaneamente portati a ragionare cercando di confutare le ipotesi che ci facciamo o che troviamo, ma facciamo il possibile e manipoliamo anche i dati per trovare conferma di quello in cui già crediamo. La razionalità umana è davvero limitata, ma la storia e i fatti dimostrano che se si viene addestrati nei tempi giusti e con un'educazione appropriata si possono prevenire gli errori sistematici e rendere più cooperativa la socializzazione.

** Epistemologo, professore ordinario di storia della medicina all'Università La Sapienza*