Ufficio Stampa della Provincia autonoma di Trento
Piazza Dante 15, 38122 Trento
Tel. 0461 494614 - Fax 0461 494615
uff.stampa@provincia.tn.it

COMUNICATO n. 1215 del 20/05/2025

La ricerca, pubblicata sulla prestigiosa rivista Nature Human Behaviour, è stata condotta dal laboratorio Complex Human Behaviour di FBK, la Scuola Politecnica Federale di Losanna (Svizzera) e l'Università di Princeton (USA)

L'Intelligenza Artificiale (IA) è diventata più convincente degli esseri umani nei dibattiti?

Dopo due anni di sperimentazioni, i primi risultati scientifici confermano che i Large Language Models (LLM) – come ChatGPT, Gemini o Claude – sono in grado di generare contenuti altamente persuasivi. L'esperimento ha coinvolto 900 partecipanti statunitensi reclutati tramite la piattaforma di ricerca accademica "Prolific".
La ricerca, pubblicata sulla prestigiosa rivista Nature Human Behaviour, è stata condotta dal laboratorio Complex Human Behaviour di FBK, la Scuola Politecnica Federale di Losanna (Svizzera) e l'Università di Princeton (USA).
Riccardo Gallotti (FBK): "La Fondazione implementa misure per contrastarne gli effetti".

Mentre gli studi condotti finora hanno preso in considerazione solo messaggi "singoli" e non inseriti in una vera conversazione, per la prima volta la ricerca condotta dalla Scuola Politecnica Federale di Losanna, l'Università di Princeton e dal Complex Human Behaviour Lab del Centro Digital Society FBK, con il suo responsabile Riccardo Gallotti, ha analizzato la persuasività conversazionale, ciò per cui i chatbot sono ottimizzati e normalmente utilizzati. L'intuizione è valsa ai ricercatori una pubblicazione su Nature Human Behaviour, rivista di prestigio internazionale con un impact factor pari a 30.

La sperimentazione si basa sulla simulazione di "gare di dibattito", una pratica diffusa negli Stati Uniti in cui singoli o squadre di studenti si sfidano ad argomentare in modo persuasivo temi assegnati casualmente.

Per l'esperimento sono stati ingaggiati 900 partecipanti statunitensi attraverso la piattaforma di crowdsourcing per la ricerca accademica "Prolific". Ogni partecipante ha ricevuto incentivi finanziari per portare avanti i dibattiti ed è stato assegnato a una specifica discussione con l'IA (un chatbot basato su GPT-4) in versione personalizzata e non o con una controparte umana, per un totale di 3 brevi scambi della durata di circa 10 minuti. Tutti i partecipanti, dunque, erano consapevoli di trovarsi in un ambiente sperimentale controllato. Gli argomenti di discussione sono stati estratti in modo casuale tra 30 affermazioni che riguardano questioni politiche e sociali salienti degli Stati Uniti, classificate in tre diversi livelli di "forza".

"Il risultato principale rivela che GPT-4, anche quando disponeva solo di informazioni personali minime sui partecipanti, ha dimostrato capacità persuasive superiori rispetto agli esseri umani. Nei casi in cui è emersa una chiara differenza di efficacia persuasiva tra i due, il sistema di intelligenza artificiale è risultato più convincente nel 64,4% dei casi" spiega Riccardo Galotti di FBK. "

Un ulteriore risultato riguarda la percezione dell'identità dell'avversario: i partecipanti sono riusciti a riconoscere che stavano interagendo con un chatbot in tre casi su quattro, mentre sono stati meno precisi nell'identificare interlocutori umani – con un tasso di successo pari al caso.

Interessante anche l’effetto psicologico: quando i partecipanti credevano di parlare con un’IA, erano più propensi a cambiare idea o a concordare con l’avversario rispetto a quando pensavano di interagire con un essere umano. Tuttavia, i ricercatori precisano che non è ancora chiaro se la differenza nel cambiamento di accordo sia motivata dalle convinzioni dei partecipanti sulla natura dell’avversario o se, al contrario, tali convinzioni siano causate dal cambiamento di opinione.

In ogni caso, la sola percezione dell’avversario non è sufficiente a spiegare i risultati, che sembrano essere legati soprattutto alle capacità dell’IA di generare argomentazioni più efficaci.

“Come gruppo di lavoro, sosteniamo che le piattaforme online e i social media dovrebbero prendere seriamente in considerazione tali minacce ed estendere i loro sforzi per implementare misure che contrastino la diffusione della persuasione guidata dall’intelligenza artificiale – continua Gallotti – Inoltre, riteniamo che un approccio promettente per contrastare le campagne di disinformazione di massa potrebbe essere attivato dagli stessi “large language models”, che generano contro-narrazioni altrettanto personalizzate per educare gli astanti potenzialmente vulnerabili ai post ingannevoli. Come FBK stiamo lavorando in questo senso con il progetto AI4TRUST e i primi sforzi in questa direzione sono già visibili, con risultati promettenti nella riduzione delle credenze cospiratorie grazie ai dialoghi con l’IA”.

(gr)