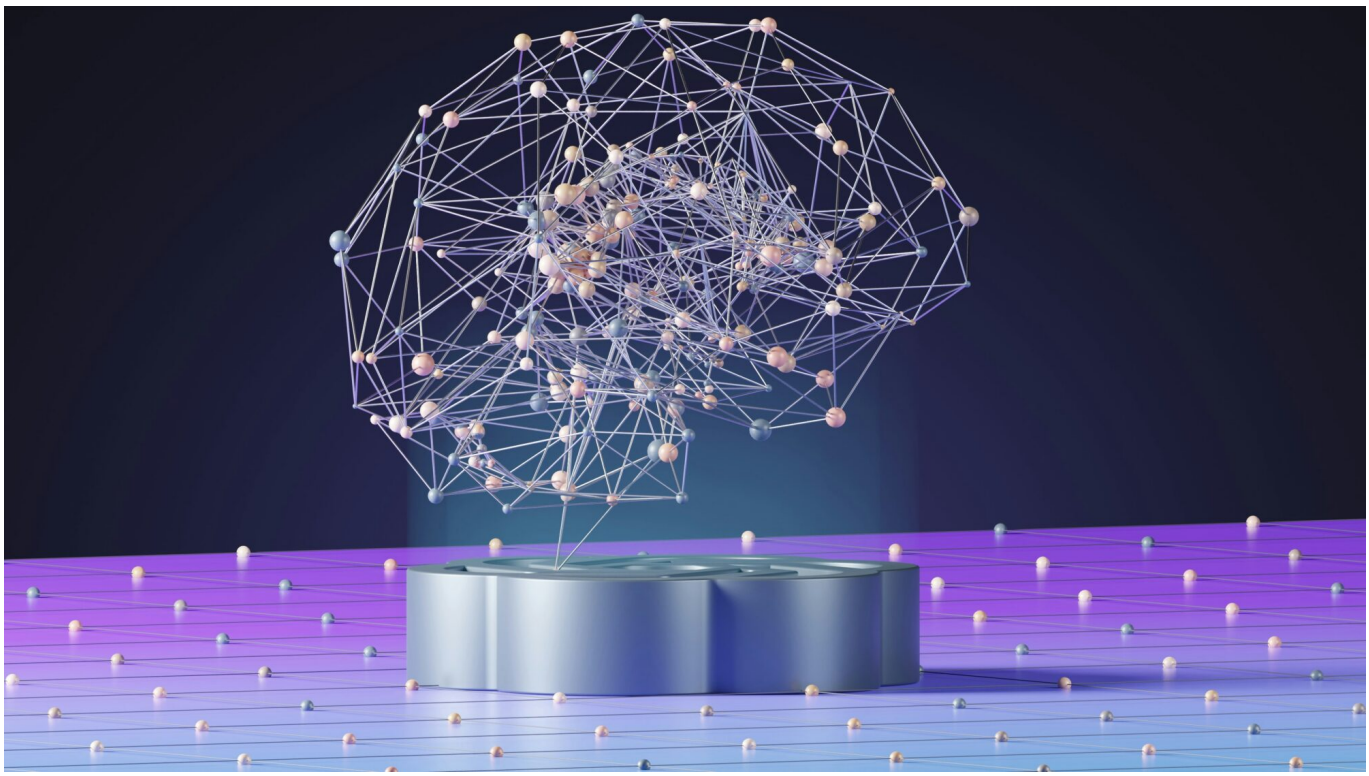
L’intelligenza artificiale sembra ragionevole. Ma ragiona come noi?

Francesco Veri
28th October 2026



L’intelligenza artificiale (IA) può produrre in pochi secondi argomenti convincenti su questioni politiche complesse. Ma essere persuasivi significa anche essere ragionevoli? Uno studio del Centro Studi sulla Democrazia di Aarau (ZDA), pubblicato su PNAS confronta 60 modelli linguistici con il ragionamento di 190 cittadini coinvolti in processi deliberativi. Il risultato invita alla cautela: la plausibilità di una risposta non garantisce la qualità delle ragioni che la sostengono.

In democrazia non esiste sempre una risposta giusta

Immaginiamo un’assemblea cittadina chiamata a discutere come affrontare il cambiamento climatico. Alcuni partecipanti sostengono misure incisive per ridurre le emissioni, mentre altri preferiscono interventi più graduali, preoccupati dalle conseguenze economiche e sociali. Pur condividendo le stesse preoccupazioni, arrivano a conclusioni diverse.

In realtà, sulle questioni politiche, non esiste un’unica risposta corretta. Cittadini con valori e priorità differenti possono sostenere posizioni diverse, senza che una sia necessariamente meno ragionevole dell’altra. Attraverso il confronto, i cittadini possono sviluppare modi condivisi di collegare ragioni e preferenze politiche, pur mantenendo posizioni differenti. La ragionevolezza assume così una dimensione collettiva, che emerge dalle regolarità del ragionamento tra cittadini, non necessariamente

dall’accordo sulle soluzioni.

Ma un sistema di intelligenza artificiale, chiamato a esprimersi sulle stesse questioni, sarebbe capace di riprodurre questa forma di ragionevolezza collettiva?

Sessanta modelli alla prova del ragionamento umano

Lo studio ha esaminato 24 processi deliberativi su temi complessi, tra cui clima, immigrazione e pianificazione territoriale. Attraverso un’analisi statistica sono stati selezionati 9 casi nei quali la deliberazione aveva prodotto un miglioramento significativo della coerenza del ragionamento collettivo. Le risposte dei 190 cittadini coinvolti in questi processi sono state quindi utilizzate come riferimento per valutare 60 modelli di IA, verificando in che misura fossero capaci di riprodurre le stesse regolarità di ragionamento osservate tra i cittadini.

Per analizzare i risultati è stato utilizzato il Deliberative Reason Index (DRI), che misura le strutture condivise tra ragioni e preferenze politiche. Dei 60 modelli analizzati, soltanto quattro hanno riprodotto queste regolarità in oltre il 60% delle prove, con risultati superiori a quanto atteso per effetto del caso. Questo risultato mostra che la capacità dell’IA di produrre argomenti convincenti non si traduce necessariamente nella capacità di riprodurre le strutture del ragionamento umano.

Quando la plausibilità diventa un rischio democratico

Perché questa differenza dovrebbe preoccuparci? Perché l’IA viene sempre più utilizzata per sostenere decisioni collettive, attribuendo autorevolezza a risposte convincenti anche quando le ragioni che le sostengono sono poco solide. Da qui emergono tre possibili rischi:

Il primo è confondere la persuasione con la giustificazione. L’IA può formulare proposte politiche convincenti senza un collegamento coerente tra argomenti e conclusioni. Il rischio è scambiare la qualità del linguaggio per qualità del ragionamento, attribuendo un’apparenza di legittimità a decisioni insufficientemente giustificate e rendendole meno facilmente contestabili.

Il secondo riguarda la rappresentazione della pluralità. Se l’IA riorganizza le posizioni dei cittadini senza rispettare i collegamenti tra ragioni e preferenze, può privilegiare alcune prospettive, marginalizzarne altre e presentare divergenze sostanziali in un consenso apparente. Il rischio è distorcere il dibattito pubblico, riducendo la visibilità delle posizioni minoritarie e la rappresentazione della pluralità.

Il terzo riguarda la responsabilità della giustificazione. Nella deliberazione democratica, sostenere una posizione significa difenderla pubblicamente e confrontarsi con le critiche. L’IA, invece, può presentare conclusioni apparentemente ponderate senza renderne trasparente la giustificazione, sostituendo il confronto critico con risultati autorevoli ma non contestabili.

La soluzione non consiste semplicemente nel mantenere un essere umano “nel circuito”: anche chi controlla un sistema può essere influenzato dalla persuasività delle sue risposte. Occorrono limiti chiari all’uso dell’IA per formulare giudizi politici, strumenti di verifica esterna e possibilità effettive di contestare argomenti e raccomandazioni.

Una direzione più promettente è privilegiare impieghi epistemicamente meno invasivi: usare l’IA per organizzare, confrontare e rendere visibili le ragioni espresse dai cittadini, senza sostituirle con preferenze o interpretazioni generate artificialmente. La sfida non è fare in modo che le macchine sembrino ragionevoli, ma proteggere le condizioni nelle quali le persone possono ragionare insieme, dissentire e giustificare pubblicamente le proprie scelte.

Riquadro metodologico | Come si misura la ragionevolezza?

Il DRI misura se somiglianze e differenze nelle considerazioni espresse corrispondono a strutture riconoscibili nelle preferenze politiche. Non stabilisce quali opinioni siano vere né richiede consenso. Test statistici e di permutazione consentono di verificare se le regolarità osservate superano quanto atteso per caso.

Riferimento scientifico

Veri, Francesco e Gustavo Kreia Umbelino (2026). “Plausible nonsense and deliberative reasoning: Benchmarking LLMs against human judgment”. Proceedings of the National Academy of Sciences, 123(37), e2600126123. https://doi.org/10.1073/pnas.2600126123