Un modello, due menti: l’IA imita i nostri pensieri

Data 05 ottobre 2025
Categoria Medicinadigitale

Un nuovo modello di IA dotato di una mente “veloce” e di una mente “lenta” e in grado di scegliere quale di volta in volta usare

Negli ultimi anni l’intelligenza artificiale sta lavorando verso un traguardo tanto complesso quanto cruciale: sviluppare una vera Teoria della mente (Theory of Mind), ossia la capacità delle macchine di comprendere e anticipare gli stati mentali degli altri, come credenze, intenzioni e desideri. Questa abilità è fondamentale per la collaborazione con gli esseri umani, per la robotica sociale e per la gestione di sistemi multi-agenti. I modelli attuali, però, mostrano forti limiti: si comportano bene in situazioni standardizzate, ma tendono a fallire non appena il contesto cambia, aumenta l’incertezza o la fatica da sovraccarico cognitivo.
Un recente studio , sviluppato da ricercatori dell’Università del Maryland, propone un nuovo modello di Theory of Mind (One Model, Two Minds, OM2M), ispirato alla celebre teoria dei “due sistemi”, resa popolare dallo psicologo premio Nobel Daniel Kahneman, secondo la quale gli esseri umani alternano due modi di ragionare: uno rapido, intuitivo ma spesso impreciso, e uno più lento, riflessivo e accurato.
L’idea è semplice ma potente: dotare le macchine di una mente “veloce” e di una mente “lenta”, e insegnare loro a scegliere di volta in volta quale usare. La mente veloce (System 1) è un sistema basato su reti neurali che riconosce schemi abituali sulle relazioni tra agenti, oggetti e luoghi con grande efficienza. La mente lenta (System 2) è un modulo che interviene quando serve riflessione, aggiornando le “convinzioni”del modello (i parametri) alla luce di indizi contestuali. Un “interruttore contestuale” (contextual gating) decide di volta in volta quale dei due sistemi deve prevalere, a seconda della situazione, consentendo al modello di decidere quando “pensare veloce” e quando “pensare lento”, cioè quanto peso attribuire al ragionamento intuitivo e quanto a quello riflessivo.
Messa alla prova in compiti classici di psicologia cognitiva, come il test della falsa credenza (la storia di Sally e Anne):

> Esperimento usato in psicologia per capire se una persona (spesso un bambino) riesce a mettersi nei panni degli altri, cioè se ha sviluppato la cosiddetta “teoria della mente”. Il test è il seguente: Sally ha un cestino e mette dentro una biglia. Poi esce dalla stanza. Mentre Sally è fuori, Anne prende la biglia e la sposta in una scatola. Quando Sally torna, si chiede al bambino: “Dove cercherà la biglia Sally?”. La risposta giusta è: nel cestino, perché Sally non sa che la biglia è stata spostata. Lei ha una credenza “sbagliata” (falsa credenza), ma coerente con quello che ha visto. Se un bambino risponde “nella scatola”, vuol dire che non riesce ancora a distinguere tra ciò che lui sa e ciò che l’altro personaggio crede. In sintesi: è un modo per testare se capiamo che gli altri possono avere pensieri diversi dai nostri, anche se sbagliati

.
L’IA non solo ha mostrato una notevole capacità di sapersi adattare anche a situazioni inedite, ma ha anche riprodotto spontaneamente i bias cognitivi tipici degli esseri umani, come l’ancoraggio, il condizionamento da priming, gli effetti del framing e la fatica da sovraccarico cognitivo (vedi box), senza alcun addestramento specifico per ciascun bias.

> [b]Bias cognitivi[/b]
>
> [b]Ancoraggio (anchoring)[/b]
> Quando dobbiamo prendere una decisione o stimare un valore, tendiamo a fissarci sul primo numero o informazione che riceviamo (l’“ancora”).
> Esempio: in ambito medico è il caso di un paziente con lombalgia che abbia una storia di ernia del disco; si è portati a diagnosticare ancora una volta una lombalgia da ernia discale perché ci si lega a una diagnosi già nota, escludendo importanti diagnosi differenziali quali una frattura osteoporotica, una metastasi vertebrale, ecc..
>
> [b]Priming[/b]
> Il nostro cervello può essere condizionato da stimoli recenti o nascosti che influenzano il comportamento successivo, anche senza che ce ne accorgiamo.
> Esempio: se prima di mostrarti un testo leggi la parola “vecchio”, potresti interpretare più lentamente un’immagine ambigua, perché sei stato inconsciamente orientato verso l’idea di lentezza.
>
> [b]Framing (effetto cornice)[/b]
> Le persone prendono decisioni diverse a seconda di come un’informazione viene presentata, anche se i dati oggettivi sono identici.
> Esempio: un farmaco che “ha il 90% di probabilità di successo” viene percepito meglio dello stesso farmaco descritto come “ha il 10% di probabilità di fallire”.
>
> [b]Sovraccarico cognitivo (cognitive load fatigue)[/b]
> Quando la nostra mente è troppo impegnata da compiti complessi, stress o fatica, cediamo al pensiero veloce e superficiale invece di riflettere a fondo.
> Esempio: se sei stanco dopo una giornata pesante, è più probabile che tu scelga cibi poco salutari o prenda

decisioni impulsive, perché il cervello non ha più “energie” per ragionare con calma

.

In pratica per la prima volta un modello computazionale non solo ragiona in maniera adattiva, ma commette anche i nostri stessi errori rispecchiando le dinamiche dei processi cognitivi umani, nei quali il pensiero intuitivo è efficiente ma fallibile e viene corretto dal pensiero deliberativo in situazioni complesse. In altre parole, OM2M non imita soltanto cosa pensiamo, ma anche come ci lasciamo influenzare dal contesto, aprendo la strada a macchine in grado di pensare in maniera più simile a noi, con comportamenti sociali più realistici e decisioni più flessibili. In questo modo, si dischiudono scenari interessanti per l’integrazione tra neuroscienze cognitive e intelligenza artificiale.

Giampaolo Collecchia e Riccardo De Gobbi

Bibliografia

Shalima Binta Manir, Tim Oates. One Model, Two Minds: A Context-Gated Graph Learner that Recreates Human Biases https://arxiv.org/abs/2509.08705