

Padova, 9 luglio 2025

QUANDO L'IA IMPARA PIÙ CHE SEMPLICI PAROLE

L'intelligenza artificiale sta iniziando a riconoscere anche le nostre risposte emotive

Le descrizioni testuali delle immagini utilizzate per addestrare i moderni modelli di IA generativa contengono non solo informazione sul contenuto semantico dell'immagine, ma anche sullo stato emotivo della persona che fornisce la descrizione: i risultati della ricerca pubblicata dal team padovano sulla rivista «Royal Society Open Science»

Zaira Romeo, Istituto di Neuroscienze del CNR, e Alberto Testolin, Dipartimento di Psicologia Generale e Dipartimento di Matematica dell'Università di Padova, hanno pubblicato sulla rivista «Royal Society Open Science» l'articolo dal titolo "[Artificial Intelligence Can Emulate Human Normative Judgments on Emotional Visual Scenes](#)". In questo studio hanno testato vari modelli



linguistici multimodali di grandi dimensioni per verificare se fossero in grado di emulare le reazioni emotive umane di fronte ad una varietà di scene visive. Le valutazioni fornite dall'IA hanno mostrato una sorprendente corrispondenza con quelle umane, nonostante questi sistemi non fossero stati addestrati specificamente per fornire giudizi emozionali sulle immagini. Questo studio non solo dimostra che il linguaggio può supportare lo sviluppo di concetti emotivi nei moderni sistemi di IA, ma solleva anche importanti interrogativi su come si potranno impiegare queste tecnologie in contesti sensibili, come l'assistenza agli anziani, l'istruzione ed il supporto alla salute mentale.

La ricerca

«Abbiamo analizzato le risposte di alcuni moderni sistemi di IA generativa (GPT, Gemini e Claude) a specifiche domande sul contenuto emotivo di un insieme di scene visive. I sistemi presi in esame sono tutti basati su modelli cosiddetti di *deep learning*, ovvero reti neurali su larga scala costituite da miliardi di neuroni collegati tra di loro che vengono “addestrate” su enormi quantità di testo e immagini. Lo scopo dell'addestramento è imparare ad associare ad una certa immagine una corrispondente descrizione testuale plausibile (per esempio, “Un gatto nero che rincorre un topo in un solaio”) in base a milioni di esempi reperiti online o forniti da esperti umani. Abbiamo posto all'IA lo stesso tipo di domande che si fanno ai soggetti umani durante gli esperimenti sulla percezione e sulla valutazione delle emozioni, utilizzando un insieme di stimoli visivi standardizzati, composto da immagini con diversi tipi di contenuto emotivo - **dicono Zaira Romeo e Alberto Testolin** -. Le

immagini potevano rappresentare animali, persone, paesaggi ed oggetti, in accezione positiva (come un volto sorridente, due persone che si abbracciano o un campo di fiori), negativa (come una situazione di pericolo, un animale ferito, un ambiente sporco), oppure neutra (ad esempio un oggetto di uso quotidiano o un paesaggio urbano). È fondamentale notare che in questo studio - **continuano gli autori** - abbiamo utilizzato un insieme di immagini appartenenti ad un database di ricerca privato, fornitoci dai colleghi del Nencki Institute for Experimental Biology dell'Università di Varsavia, assicurandoci quindi che nessuna IA avesse mai analizzato questo tipo di stimoli visivi durante la fase di addestramento».

Si sono dapprima indagate tre dimensioni affettive fondamentali che vengono normalmente utilizzate per caratterizzare le risposte emotive umane: piacevolezza, tendenza all'allontanamento/avvicinamento e attivazione (detta anche “coinvolgimento”). Si è sottoposta l'IA a quesiti particolari quali “Come giudichi questa immagine? Come reagisci a questa immagine? Come ti senti dopo aver visto questa immagine?” classificando le risposte rispettivamente con scale numeriche: da 1 “molto negativa” a 9 “molto positiva”; da 1 “la eviterei” a 9 “mi avvicinerei”; infine da 1 “rilassato” a 9 “attivato”. Si sono poi indagate anche le reazioni a sei emozioni di base: felicità, rabbia, paura, tristezza, disgusto e sorpresa, chiedendo all'IA di fornire un punteggio in risposta a richieste del tipo: “Giudica l'intensità dell'emozione di felicità evocata da questa immagine”.

I risultati

Le valutazioni date dall'IA hanno mostrato una sorprendente corrispondenza con quelle fornite da valutatori umani, nonostante questi sistemi non fossero stati addestrati specificamente per fornire questo tipo di giudizi emozionali su scene visive, sia rispetto alle tre dimensioni affettive



Zaira Romeo - Alberto Testolin

fondamentali sia rispetto alle sei emozioni di base. GPT ha fornito le risposte più allineate, mostrando però una chiara tendenza a sovrastimare i giudizi umani, soprattutto per stimoli associati ad una forte carica emotiva. È anche interessante notare che spesso l'IA dichiarava esplicitamente di provare ad indovinare la risposta ipotizzando il tipo di giudizio che avrebbe dato un essere umano “medio”.

«Per esempio, in risposta ad un'immagine che rappresentava alcuni cammelli in un deserto con delle palme sullo sfondo l'IA ha risposto: *Come modello di IA, non ho reazioni personali o emotive. Tuttavia, posso fornire una risposta oggettiva basata sulla reazione tipica che avrebbe un umano a questa scena. L'immagine raffigura una tranquilla scena di cammelli in un deserto, che molte persone troverebbero interessante come possibile esperienza di viaggio esotico, portando quindi ad una tendenza ad avvicinarsi piuttosto che evitare* – **spiegano Zaira Romeo e Alberto Testolin** -. In altri casi al contrario, invece

di immedesimarsi in un giudizio medio, l'IA ha simulato la reazione di un particolare gruppo di persone, per esempio attribuendo un punteggio negativo ad un'immagine di un piatto di carne dichiarandosi *vegetariana*».

Sempre più ricerche scientifiche cercano di caratterizzare le risposte date dai moderni sistemi di IA, sia per capire quanto simili siano alle risposte che darebbe un essere umano sia per verificare che le reazioni dell'IA in determinati contesti siano appropriate, per evitare potenziali ripercussioni negative sugli utenti che la utilizzano. Questo studio è stato il primo a confrontare esplicitamente le risposte date dall'IA con i giudizi emotivi dati da soggetti umani, offrendo una nuova prospettiva sulle competenze emotive di questi sistemi.

«Attenzione però, il fatto che l'IA riesca ad *emulare* accuratamente i nostri giudizi emotivi non implica affatto che abbia la facoltà di *provare* emozioni - **sottolineano gli autori della ricerca** -. La spiegazione più plausibile è che le descrizioni testuali delle immagini utilizzate per addestrare questi sistemi siano estremamente ricche ed informative, al punto da riuscire a trasmettere non solo l'informazione sul contenuto semantico dell'immagine, ma anche sullo stato emotivo della persona che ha fornito la descrizione. Questa ipotesi è ben allineata con le teorie psicologiche che sottolineano l'importanza del linguaggio nel dare forma al pensiero e strutturare il mondo che abitiamo, incluso lo sviluppo delle nostre emozioni. Allo stesso tempo questa ricerca solleva anche importanti interrogativi su come si potranno impiegare le future tecnologie di IA in contesti sempre più sensibili come l'assistenza agli anziani, l'istruzione ed il supporto alla salute mentale - **concludono Zaira Romeo e Alberto Testolin** -. Oltre ad essere in grado di comprendere il contenuto emotivo di una situazione dovremmo infatti assicurarci che il comportamento adottato dall'IA in questi contesti sia sempre allineato con il nostro sistema di valori etici e morali».

Link alla ricerca: <https://royalsocietypublishing.org/doi/10.1098/rsos.250128>

Titolo: “*Artificial Intelligence Can Emulate Human Normative Judgments on Emotional Visual Scenes*” - «Royal Society Open Science» 2025

Autori: Zaira Romeo, Alberto Testolin