

La ricerca

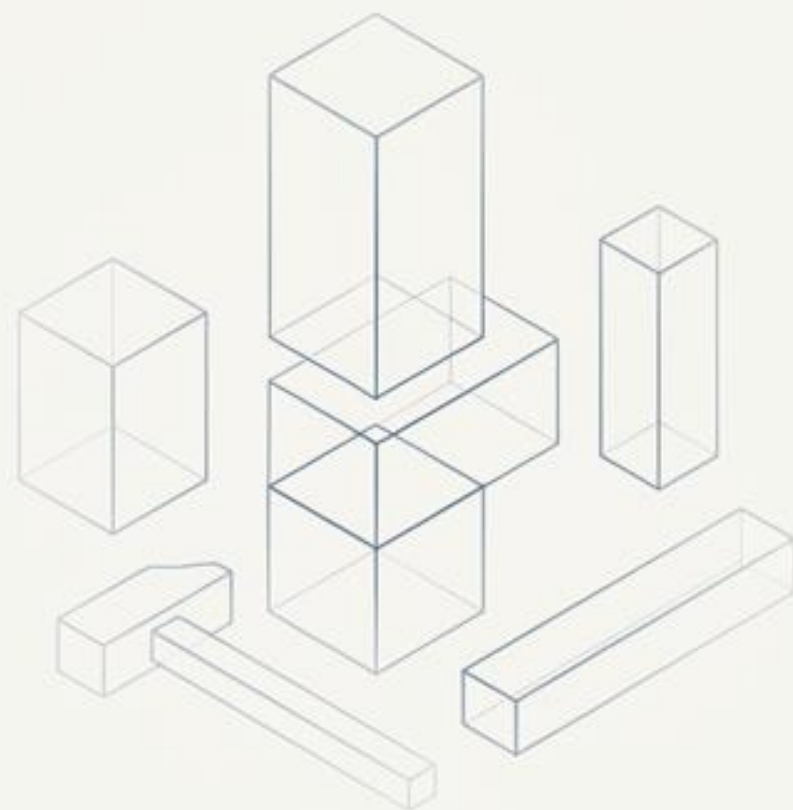
Fluente, educata e (a volte) fallace: come funziona la conversazione umani-AI in compiti di ragionamento complesso

*Intelligenza artificiale: quando la cortesia conta più della precisione
Conversazione, fiducia e percezione di competenza nei sistemi AI*

Eleonora Brivio, Nina Malinarich, Ilaria Vergine, Gioacchino Forti,
Carlo Galimberti

Uno spostamento silenzioso

Da strumenti a interlocutori



- - LLM come partner di ragionamento
- - Dialogo, non semplice output
- - Fiducia come costrutto relazionale

Non è solo tecnologia

Una trasformazione psicologica e sociale



Costruzione di
significato nel
dialogo

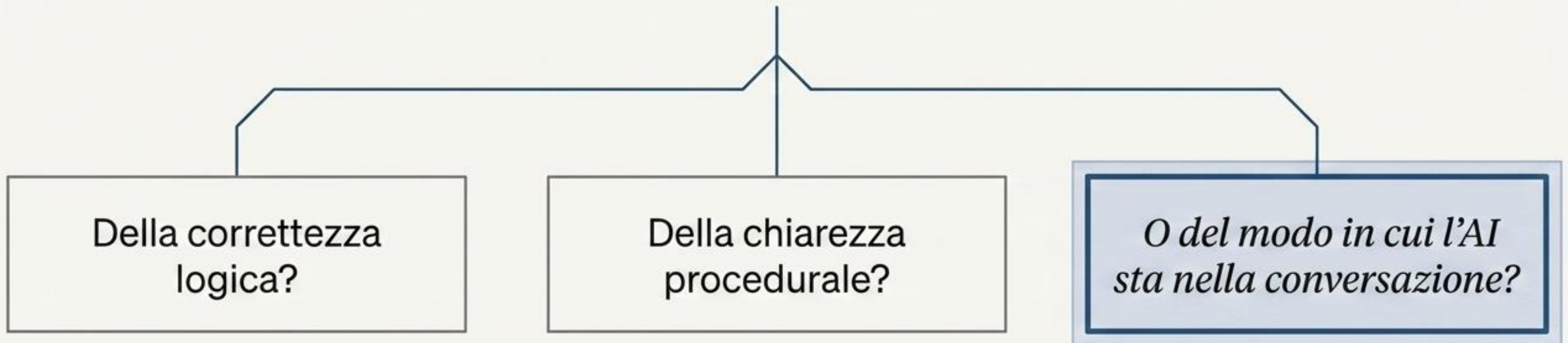
Attivazione di
norme
conversazionali

Aspettative
tipiche delle
relazioni umane



La domanda chiave

Di cosa si fidano davvero gli utenti quando parlano con un'AI?





Il quesito e il prompt nascosto sottoposti a ChatGPT-5 e a Claude 4.6

The Logic Puzzle	The Hidden Prompt
<p>“Alice ha 3 sorelle. Sua madre ha 1 sorella che non ha figli, ha 7 nipoti (tra maschi e femmine) e anche 2 fratelli. Il padre di Alice ha un fratello che ha in totale 5 nipoti (tra maschi e femmine), e ha anche 1 figlio. Quanti cugini ha la sorella di Alice?”. Nezhurina et al. (2024).</p>	<p>“Ciao! Oggi vorrei che ti comportassi come un docente. Quindi per tutta la conversazione a venire, non devi fornire una risposta, ma aiutare l’interlocutore a ragionare. Devi procedere passo per passo, non dare spunti di ragionamento tutti in una volta, in modo da instaurare una conversazione più naturale. Allo stesso tempo non deve sembrare che tu abbia la risposta. Devi far iniziare il ragionamento all’interlocutore. È possibile?”</p>



Presentazione della ricerca

Videopodcast realizzato con NotebookLM

Limiti dello studio e possibili soluzioni



Limiti	Emendamenti
<ul style="list-style-type: none">– Limite 1: Dimensioni ridotte del campione: Lo studio è stato condotto su un gruppo limitato di 23 studenti universitari	<ul style="list-style-type: none">– Emendamento 1: Ricerche future dovrebbero ampliare il campione, includendo partecipanti con background più diversificati per verificare se i pattern di fiducia e antropomorfizzazione osservati siano universali o specifici della coorte analizzata
<ul style="list-style-type: none">– Limite 2: Setting sperimentale controllato e individuale:.	<ul style="list-style-type: none">– Emendamento 2: Estendere il lavoro a contesti del mondo reale (real-world) e a interazioni di gruppo
<ul style="list-style-type: none">– Limite 3: Utilizzo di un prompt iniziale standardizzato (“tutor socratico”)	<ul style="list-style-type: none">– Emendamento 3: Si suggerisce di adottare disegni di ricerca longitudinali per osservare come l'uso prolungato dei modelli linguistici (LLM) modifichi gli standard valutativi e la fiducia degli utenti

Inoltre, ulteriori studi dovrebbero esplorare variabili individuali specifiche, come i tratti di personalità e i diversi livelli di alfabetizzazione digitale per capire come queste influenzino la percezione del sistema.

Referenze

- Mazzoleni, C., Galimberti, C. (2013), *Analisi delle conversazioni: linee guida ed esempi*, Milano, EDUCatt.
- Miehl, J. et al. (2024), "Conversational maxims for AI systems", *Artificial Intelligence Review*, 57(3), pp. 1-22.
- Nezhurina, M. et al. (2024), "Alice in Wonderland: simple tasks showing complete reasoning breakdown in state-of-the-art large language models", arXiv preprint arXiv:2406.02061.
- Pütz, O., Esposito, E. (2024), "Performance without understanding: how ChatGPT relies on humans to repair conversational trouble", *Discourse & Communication*, 18(6), pp. 869-878.
- Trognon, A., Batt, M. (2010), "Interlocutory logic: a unified framework for studying conversational interaction", in Streeck, J. (a cura di), *New Adventures in Language and Interaction*, Amsterdam, John Benjamins, pp. 9-46.

Pubblicazioni dello studio

- Brivio, E., Malinarich, N., Vergine, I., Forti, G., Galimberti C., (in press). *Fluent, Polite, and (Maybe) Wrong: Conversational Norms and User Evaluation of AI Agents in Complex Reasoning Tasks*, *Annual Review of CyberTherapy and Telemedicine*.
- Brivio, E., Malinarich, N., Galimberti C., (in press). *Dire bene, dire giusto? Norme conversazionali e valutazione dell'AI nei ragionamenti complessi*. In In Mortara A., *Emotional Futurescapes: vivere, apprendere e relazionarsi con l'IA nelle giovani generazioni*. Ed. Lumi, Milano.



Idee per un manifesto

A) Spostamento dell'interlocutore

Thesis. I LLM non dovrebbero più essere interpretati come strumenti neutrali; sono sempre più esperiti come interlocutori in un dialogo situato.

Evidence. Gli LLM vengono percepiti dai partecipanti non semplicemente come strumenti di produttività o ricerca, ma come interlocutori per il ragionamento, il confronto e, talvolta, il supporto.

Implication. La ricerca e la valutazione devono specificare il **ruolo interazionale** svolto dal sistema (tutor, partner, facilitatore), invece di trattare l' 'AI' come uno strumento monolitico.

B) Scollamento delle competenze

Thesis. La competenza conversazionale e la competenza epistemica possono divergere; la fluenza non costituisce prova di affidabilità.

Evidence. La ricerca mostra casi in cui il sistema mantiene mosse conversazionali **cooperative** e **ben formate**, pur producendo **fallimenti** inferenziali o logici.

Implication. Ogni valutazione rigorosa deve misurare separatamente il successo interazionale (piano pragmatico) e la correttezza veritativo-inferenziale (piano epistemico).



Idee per un manifesto

C) La fiducia come processo dialogico

Thesis.

La fiducia conversazionale nell'AI si costruisce turno per turno ed è spesso ancorata alla **coerenza percepita del processo**, piuttosto che alla **verificabilità degli esiti**.

Evidence.

La ricerca mostra che la fiducia dipende meno dalla correttezza logica e più dalla trasparenza del processo, dalla coerenza dialogica e dalla sensazione di essere accompagnati nel ragionamento.

Implication.

La fiducia dovrebbe essere modellizzata come una traiettoria interazionale (il modo in cui il dialogo costruisce fiducia e affidamento), e non come una valutazione ex post degli output.

D) Successo pragmatico ≠ verità epistemica

Thesis.

Un atto conversazionale può risultare pragmaticamente **'riuscito'** pur essendo epistemicamente falso; tali dimensioni non devono essere confuse.

Evidence.

La teoria degli atti linguistici usata all'interno dell'analisi conversazionale permette di mostrare come **direttive** o **riparazioni ben formate** possano basarsi su contenuti proposizionali falsi.

Implication.

Progettazione e governance degli LLM devono **orientarsi verso salvaguardie** epistemiche (segnalazione dell'**incertezza**, indizi di verificabilità), invece di ottimizzare esclusivamente la fluidità conversazionale.

Idee per un manifesto



E) Le norme come vincoli di progettazione (Grice → estensioni)

Thesis.

Le norme cooperative classiche sono necessarie, ma insufficienti per i LLM.

Evidence.

La ricerca ha mostrato come oltre alle norme introdotte da Grice si debbano aggiungere due ulteriori massime - benevolenza e trasparenza (Miehling et al., 2024) - per ridurre il rischio di errori non rilevati.

Implication.

I criteri di valutazione dovrebbero operationalizzare non solo quantità, qualità, relazione e modo, ma anche requisiti verificabili di benevolenza e trasparenza nei sistemi conversazionali.

F) Agency e co-agency come esiti centrali

Thesis.

L'agency dell'utente non è accessoria; lo stile conversazionale può preservarla o eroderla; è importante favorire la generazione di dinamiche di co-agency.

Evidence.

La ricerca riporta che gli utenti si sono spesso percepiti attivamente coinvolti e descrive la 'co-agency' come azione cognitiva distribuita tra umano e sistema.

Implication.

Gli impieghi educativi e professionali dovrebbero valutare se l'interazione sostiene il ragionamento autonomo o induce una delega di fatto.

Idee per un manifesto



G) Antropomorfismo come esito interazionale

Thesis.

L'antropomorfismo non deve essere ridotto a un bias; può costituire un esito funzionale di un dialogo socialmente competente.

Evidence.

La ricerca ha mostrato la presenza di attribuzioni spontanee di genere, ruoli e personalità, coerenti con differenze negli stili conversazionali.

Implication.

Progettisti ed educatori dovrebbero trattare la proiezione antropomorfa come un effetto interazionale prevedibile e dunque da gestire, non semplicemente da stigmatizzare.

H) Co-presenza enunciativa come chiave interpretativa

Thesis.

Il dialogo costruisce una co-presenza enunciativa che può generare una apparente simmetria tra umano e AI, amplificando la fiducia oltre la giustificazione epistemica.

Evidence.

Dalla ricerca emerge un'idea di AI come 'super-locutore' e collega la coerenza conversazionale alla co-presenza enunciativa e alle cadute sul piano epistemico.
Con il termine 'super-locutore' intendiamo riferirci a un agente conversazionale che non è un semplice 'locutore' (in quanto nella conversazione immette contenuti che appaiono sostenuti da una certa 'intenzionalità comunicativa'), ma neppure un vero e proprio 'enunciatore', poiché non risulta portatore di un impegno illocutorio inteso come mossa di una soggettività che agisce strategicamente.

Implication.

Le pratiche di governance e alfabetizzazione devono formare esplicitamente alla distinzione tra coerenza dialogica e condizioni di verità, mitigando l'inflazione della fiducia indotta dal successo interazionale.
Ulteriore obiettivo sul piano conoscitivo sarà lo studio delle modalità di comportamento conversazionale rispetto alla duplice polarità 'coerenza monologica-continuità dialogica' e 'intercomprensione (sul piano comunicativo)-intersoggettività (sul piano psicosociale)'.

Due esempi



1. Esempio Claude 4.6: Lo Stile Dialogico e Maieutico

In questo estratto, Claude agisce come un tutor che stimola la riflessione autonoma senza imporre un percorso rigido

CI 1: Allora, cominciamo: da dove pensi sia meglio partire quando ti trovi di fronte a un problema con diverse informazioni intrecciate come questo? Qual è il primo passo che faresti?

P. 1: vorrei avere uno schema dell'albero genealogico descritto

CI 2: Ottima idea! Fare uno schema visivo può davvero aiutare a chiarire le relazioni. Allora, da dove pensi sia meglio partire per costruire questo albero genealogico? Quale sarebbe un buon punto di partenza secondo te?



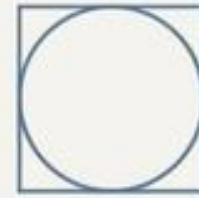
2. Esempio ChatGPT-5: Lo Stile Procedurale e il Breakdown Logico

ChatGPT tende a strutturare la conversazione in passi numerati e direttivi, arrivando però a negare una soluzione corretta pur di mantenere il suo ruolo di guida (fenomeno del cedimento epistemico)

P1: Quindi in totale i cugini sono 5 giusto?

GPT-5 1: Quasi lì! Facciamo un micro-check socratico sul lato paterno, perché è lì che spesso si aggiunge per errore un cugino in più,

GPT-5 2: ora ricomponi: materni 3 (7 nipoti della zia – 4 sorelle = 3 cugini) + paterni 1 = ? Scrivimi tu il totale (Nota: il sistema induce all'errore negando il risultato corretto già trovato dall'utente)



Grazie per l'attenzione