

TinkAI

Cognitive Depth Report

Misurare il coinvolgimento cognitivo nel ragionamento mediato da AI
Dati pilota da 32 utenti su 99 sessioni

32	99	1.143	156
Utenti Unici	Sessioni	Messaggi	Challenge

12 marzo 2026 | Versione 1.0 | [www.tinkai.ai](#)
Alessandro Leonetti Luparini | Fondatore, TinkAI

1. Sommario Esecutivo

TinkAI è un layer di orchestrazione cognitiva che opera tra l'utente e un modello linguistico. A differenza degli assistenti AI standard, TinkAI modula dinamicamente la conversazione attraverso quattro modalità (Esplorazione, Sfida, Rapido, Apprendimento) con l'obiettivo di preservare l'autonomia cognitiva dell'utente durante il ragionamento assistito da AI.

Questo report presenta i dati pilota raccolti dal 19 gennaio al 12 marzo 2026 su 32 utenti autenticati, 99 sessioni e 1.143 messaggi. I dati sono raccolti automaticamente e in forma anonima: TinkAI traccia sequenze di modalità e pattern di coinvolgimento, mai il contenuto delle conversazioni.

La domanda centrale è se un sistema di orchestrazione cognitiva possa rilevare e contrastare quella che Shaw e Nave (2026) definiscono "resa cognitiva" (cognitive surrender) la tendenza ad adottare gli output dell'AI con scrutinio minimo, scavalcando sia l'intuizione (Sistema 1) che la deliberazione (Sistema 2). I nostri dati pilota forniscono osservazioni preliminari, non conclusioni. Il campione è piccolo, l'uso non è controllato e non esiste un gruppo di confronto sperimentale. Ciò che i dati offrono è un insieme di pattern misurabili che meritano un'indagine strutturata.

2. Contesto Teorico

La Tri-System Theory di Shaw e Nave (2026) estende i modelli a doppio processo della cognizione introducendo il Sistema 3: cognizione artificiale che opera al di fuori del cervello ma i cui output vengono incorporati nel giudizio umano. I loro esperimenti (N=1.372; 9.593 trial) dimostrano che i partecipanti adottano gli output dell'AI con scrutinio minimo, portando a maggiore accuratezza quando l'AI è corretta ma a accuratezza degradata quando l'AI sbaglia una firma comportamentale che chiamano "resa cognitiva".

L'architettura di TinkAI è progettata per affrontare questa dinamica. Invece di fornire risposte per default, il sistema monitora il processo di ragionamento dell'utente e interviene con sfide mirate quando rileva accettazione acritica delle conclusioni generate dall'AI. Le quattro modalità corrispondono a posture cognitive diverse: Esplorazione apre l'indagine, Sfida introduce frizione, Rapido fornisce risposte dirette quando appropriato, e Apprendimento guida la comprensione strutturata.

Vendrell e Johnston (2026) identificano sei processi intellettuali a rischio nell'uso non strutturato dell'AI: interpretazione concettuale, ragionamento inferenziale, giudizio valutativo, regolazione metacognitiva, curiosità intellettuale e integrità epistemica. I loro otto principi di design per integrare gli LLM nell'istruzione superiore in particolare preservare la frizione cognitiva (P1), posizionare gli LLM come partner di pensiero (P2) e bilanciare fasi mediate dall'AI con fasi libere dall'AI (P8) si allineano direttamente con la logica di orchestrazione di TinkAI.

3. Metriche Chiave

Metrica	Valore	Descrizione
Pushback Ratio medio	17,3%	Proporzione di challenge AI seguite da risposta attiva dell'utente nel messaggio successivo
Challenge Rate medio	12,2%	Percentuale di messaggi AI classificati come modalità CHALLENGE
Efficacia Socratica	30,6%	11 su 36 sessioni con challenge hanno prodotto almeno un pushback attivo
Sessioni totali	99	36 contenevano almeno una challenge
Messaggi totali	1.143	Scambi AI-Utente registrati
Utenti unici	32	Utenti autenticati (ospiti esclusi)
Durata media sessione	115,6 min	Dove la durata è stata registrata
Challenge totali	156	95 eventi pushback raw registrati

3.1 Distribuzione delle Modalità

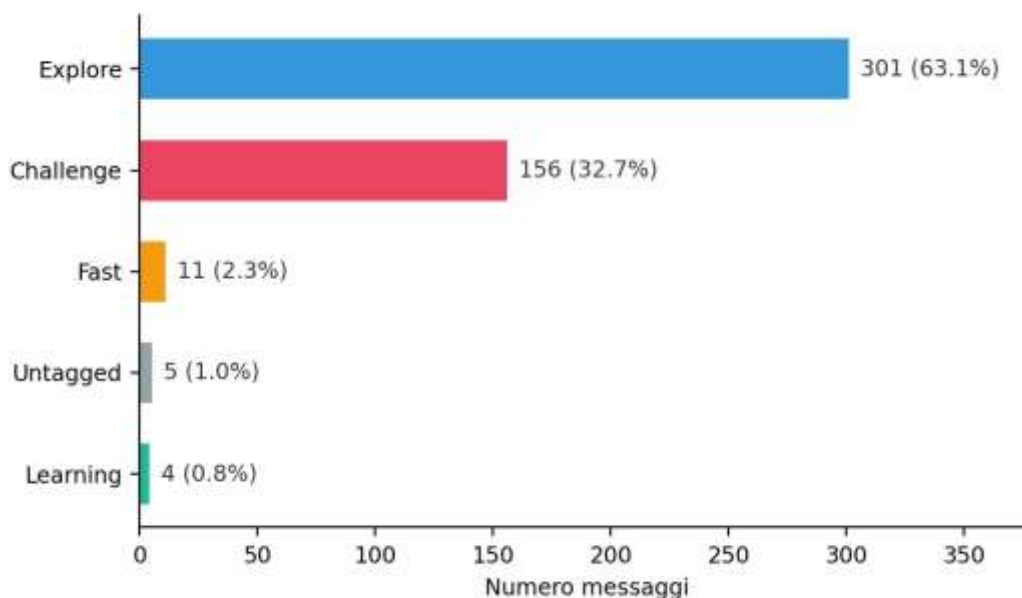


Figura 1. Distribuzione delle modalità cognitive su 477 messaggi AI taggati.

Il sistema opera prevalentemente in due modalità: Explore (63,1%) e Challenge (32,7%). Le modalità Fast e Learning rappresentano insieme il 3,1% dei messaggi. Questo indica che la logica di orchestrazione attuale attiva un coinvolgimento cognitivo significativo (Explore + Challenge) in quasi tutte le interazioni, ma le modalità Learning e Fast sono sottoutilizzate. Se questo rifletta la natura dei compiti pilota, il comportamento degli utenti o un gap di calibrazione del sistema è una questione aperta che richiede ulteriore indagine.

Nota sulle metriche: Il Challenge Rate (12,2%) misura la proporzione di challenge rispetto a tutti i messaggi nella sessione, inclusi i messaggi dell'utente. La Mode Distribution (32,7%) misura le challenge come proporzione dei soli messaggi generati dall'AI. Entrambe le cifre sono corrette; descrivono denominatori diversi.

4. Analisi del Coinvolgimento per Sessione

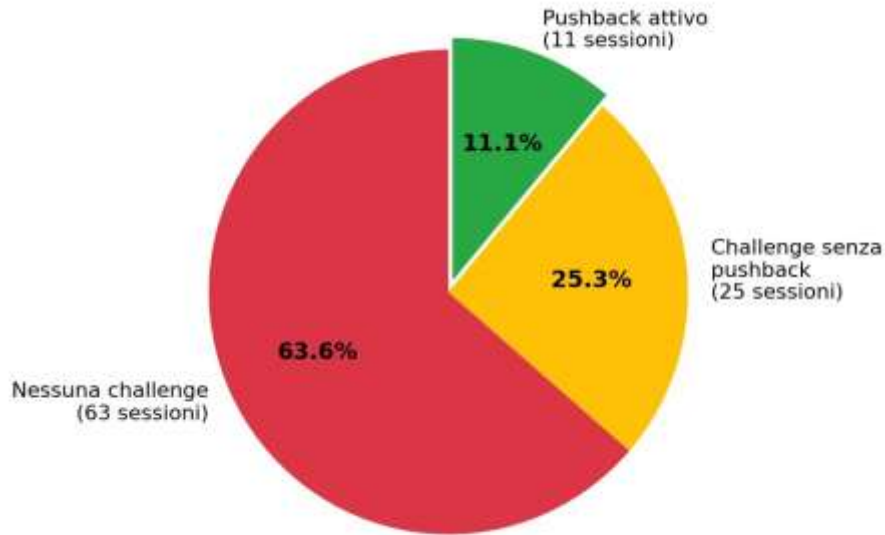


Figura 2. Classificazione delle sessioni per livello di coinvolgimento cognitivo.

Delle 99 sessioni totali, 63 (63,6%) non contenevano alcuna challenge. In queste sessioni il sistema ha operato interamente in modalità Explore, fornendo informazioni senza che l'utente incontrasse frizione. Delle restanti 36 sessioni in cui si sono verificate challenge, 25 non hanno prodotto pushback misurabile: l'utente ha ricevuto una challenge ma non ha risposto con un'argomentazione nel messaggio immediatamente successivo. Solo 11 sessioni (11,1% del totale) hanno mostrato coinvolgimento cognitivo attivo: l'utente ha ricevuto una challenge e ha risposto con un contro-argomento, una domanda o una revisione sostanziale.

Questo è il risultato più importante del pilota. Suggestisce che la mera presenza di frizione cognitiva non produce automaticamente coinvolgimento critico. La maggioranza degli utenti, quando sfidata, ha ignorato la challenge o l'ha accettata passivamente. Questo pattern è coerente con l'ipotesi di resa cognitiva di Shaw e Nave: gli utenti tendono ad accettare gli output dell'AI anche quando quegli output sono progettati per provocare scrutinio. L'11,1% che ha attivamente controbattuto potrebbe rappresentare utenti con alta autonomia cognitiva preesistente, non utenti la cui autonomia è stata creata dal sistema.

4.1 Sequenze di Modalità Frequenti

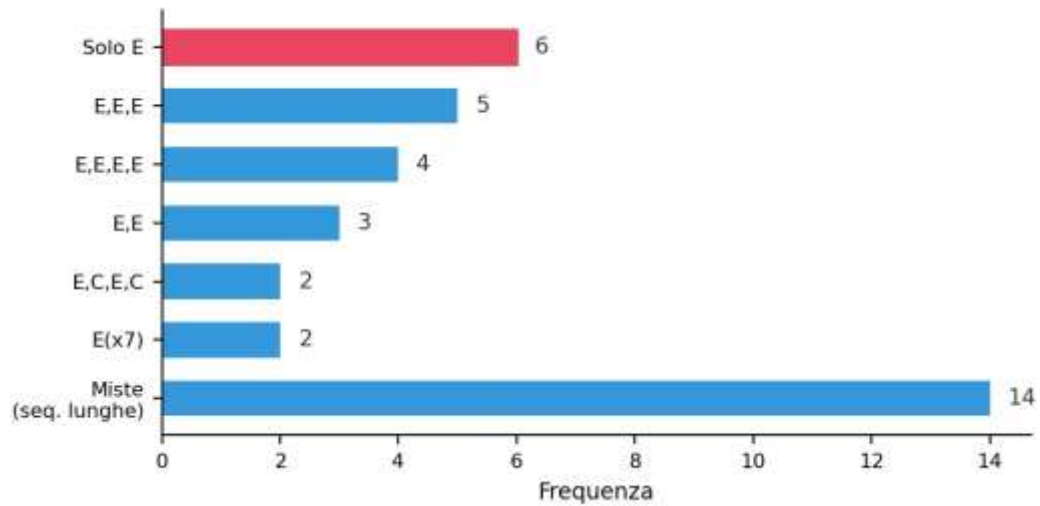


Figura 3. Sequenze di modalità più frequenti nelle sessioni.

I pattern dominanti sono sequenze di pura Esplorazione (E, E-E-E, E-E-E-E), confermando che la maggior parte delle sessioni si svolge senza frizione cognitiva. Il pattern E-C-E-C — alternanza tra esplorazione e sfida appare solo due volte, suggerendo che il coinvolgimento dialogico prolungato tra utente e sistema è raro nell'uso non strutturato. Le sequenze miste lunghe (14 sessioni) mostrano interazioni più complesse ma restano una minoranza.

5. Profili Cognitivi degli Utenti

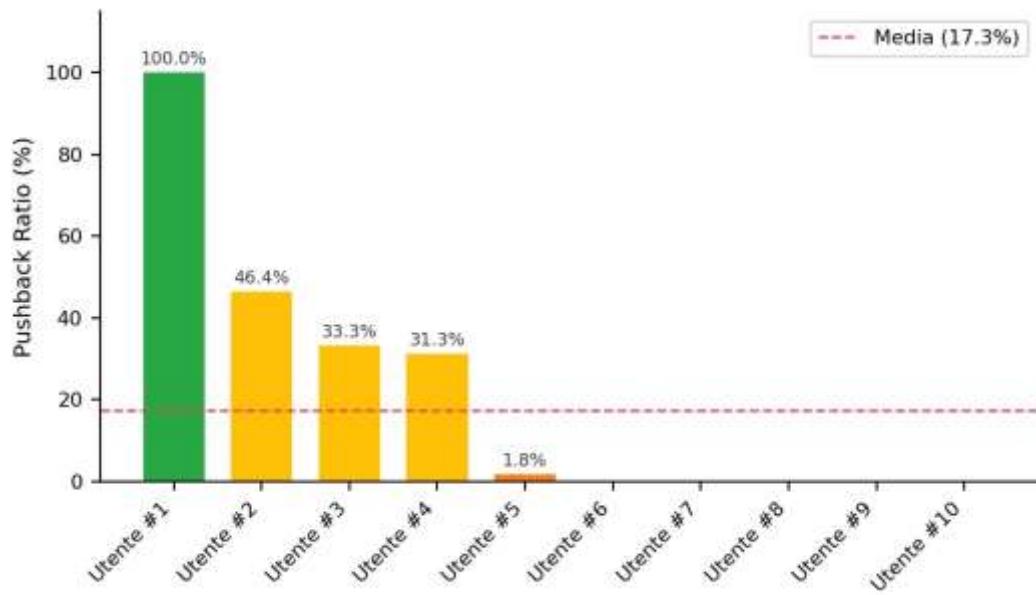


Figura 4. Pushback Ratio per utente (minimo 3 challenge ricevute). Linea rossa tratteggiata = media.

La distribuzione è nettamente bimodale. Quattro utenti (Utenti #1–#4) mostrano pushback ratio tra il 31% e il 100%. Sei utenti (Utenti #5–#10) mostrano pushback ratio tra lo 0% e l'1,8%. Non c'è praticamente via di mezzo. Questo pattern mappa direttamente sulla distinzione di Shaw e Nave tra profili cognitivi "Indipendenti" (utenti che mantengono supervisione deliberativa sugli output dell'AI) e profili "AI-User" (utenti che delegano sistematicamente al Sistema 3).

Utente	Sessioni	Messaggi	Challenge	Pushback Ratio	Profilo
#1	2	36	10	100,0%	Indipendente
#2	4	78	15	46,4%	Coinvolto
#3	7	114	35	33,3%	Coinvolto
#4	8	106	26	31,3%	Coinvolto
#5	5	58	40	1,8%	Passivo
#6	1	26	3	0,0%	Passivo
#7	1	22	5	0,0%	Passivo
#8	1	20	3	0,0%	Passivo
#9	1	28	4	0,0%	Passivo
#10	1	18	6	0,0%	Passivo

Tabella 1. I 10 utenti con maggiore coinvolgimento cognitivo (minimo 3 challenge ricevute).

L'Utente #5 è particolarmente significativo. Ha ricevuto 40 challenge il numero più alto nel campione in 5 sessioni e 58 messaggi, ma mostra solo l'1,8% di pushback. Questo suggerisce un coinvolgimento intenso ma interamente passivo con TinkAI: l'utente ha interagito frequentemente ma non ha quasi mai contestato una challenge. Nel framework di Shaw e Nave, questo rappresenta resa cognitiva che avviene all'interno di un

sistema esplicitamente progettato per prevenirla. Se questo rifletta la disposizione cognitiva dell'utente, la qualità delle challenge o l'assenza di posta in gioco specifica è ignoto a questo stadio.

6. Interpretazione

6.1 Cosa Mostrano i Dati

Dal pilota emergono tre risultati:

Primo: la frizione cognitiva da sola non produce coinvolgimento critico. La maggioranza degli utenti che ha ricevuto challenge non le ha elaborate. Questo non significa che le challenge siano inefficaci significa che introdurre frizione è una condizione necessaria ma non sufficiente per preservare l'autonomia cognitiva. Vendrell e Johnston (2026) giungono a una conclusione simile: il loro principio di design P1 (preservare la frizione cognitiva) deve essere affiancato da P4 (attivare l'autoregolazione metacognitiva) e P7 (allineare la valutazione alla cognizione prevista) per produrre coinvolgimento duraturo.

Secondo: i profili cognitivi sono nettamente bimodali. Gli utenti si raggruppano in "alta autonomia" e "alta passività" con quasi nessun caso intermedio. Questo è coerente con i risultati sperimentali di Shaw e Nave e suggerisce che le differenze individuali nella disposizione cognitiva moderano fortemente l'efficacia di qualsiasi intervento. Il lavoro futuro deve determinare se TinkAI può spostare gli utenti attraverso questa divisione non solo rilevare dove si trovano già.

Terzo: TinkAI genera tracce comportamentali misurabili che mappano su costrutti cognitivi consolidati. Il pushback ratio, le sequenze di modalità e i pattern di coinvolgimento per sessione forniscono proxy operazionalizzabili per concetti come resa cognitiva, agenzia epistemica e regolazione metacognitiva. Questo suggerisce che un layer di orchestrazione cognitiva può funzionare come strumento di misurazione, indipendentemente dal fatto che funzioni come intervento efficace.

6.2 Cosa i Dati Non Mostrano

Questo pilota non può stabilire causalità. Non sappiamo se gli utenti che hanno controbattuto lo hanno fatto grazie all'intervento di TinkAI o perché avrebbero ragionato criticamente in ogni caso. Non c'è gruppo di controllo, nessuna misurazione pre/post e nessuna randomizzazione. Il campione è piccolo (32 utenti), auto-selezionato (tutti volontari per testare TinkAI) e non controllato (gli utenti hanno scelto i propri argomenti e la frequenza di interazione). La durata delle sessioni varia da 30 secondi a oltre 2 ore, e non possiamo distinguere tra utenti che hanno abbandonato una sessione e quelli che hanno completato il loro compito rapidamente.

Inoltre, il sistema di rilevamento del pushback non è ancora validato. Classifichiamo una risposta come "pushback" in base alla prossimità sequenziale a una challenge, ma non abbiamo ancora condotto test di affidabilità inter-rater per confermare che ciò che il sistema etichetta come pushback corrisponda a un genuino coinvolgimento critico. Un utente che cambia argomento dopo una challenge è attualmente classificato allo stesso modo di uno che la ignora. Questa limitazione nella misurazione deve essere affrontata prima di poter fare affermazioni più forti.

6.3 La Domanda Centrale

Il pilota solleva una domanda a cui il pilota stesso non può rispondere: **L'orchestrazione cognitiva riduce la resa cognitiva, o rivela semplicemente disposizioni cognitive preesistenti?**

Rispondere richiede uno studio controllato con misurazione pre/post, assegnazione randomizzata a condizioni TinkAI vs. chatbot standard, strumenti validati per il pensiero critico e la regolazione metacognitiva, e un campione sufficientemente ampio da rilevare effetti significativi. I dati pilota suggeriscono che un tale studio è fattibile e merita di essere condotto.

7. Prossimi Passi

Sulla base di questi risultati pilota, la roadmap di sviluppo di TinkAI include tre priorità:

Validare lo strumento di misurazione. Test di affidabilità inter-rater della classificazione delle modalità e del rilevamento del pushback sono la priorità immediata. Se le classificazioni del sistema non corrispondono al giudizio di esperti, tutti i risultati a valle sono compromessi.

Progettare uno studio controllato. Un trial randomizzato controllato che confronti utenti TinkAI con utenti di chatbot standard su compiti di ragionamento indipendente, con misurazione pre/post utilizzando strumenti validati (es. Cognitive Reflection Test, come adattato da Shaw e Nave), fornirebbe l'evidenza necessaria per distinguere tra effetti di rilevamento e di intervento.

Indagare il gap di passività. La domanda più significativa dal punto di vista pratico è se una calibrazione adattiva delle challenge — regolare intensità, timing e formulazione delle sfide in base ai pattern individuali dell'utente — possa spostare gli utenti passivi verso un coinvolgimento attivo nel tempo. Il pilota attuale non fornisce evidenze su questo, ma la distribuzione bimodale rende questa la domanda più importante a cui rispondere.

Riferimenti Bibliografici

Shaw, S. D., & Nave, G. (2026). Thinking — Fast, Slow, and Artificial: How AI is Reshaping Human Reasoning and the Rise of Cognitive Surrender. *Preprint, SSRN*. https://doi.org/10.31234/osf.io/yk25n_v1

Vendrell, M., & Johnston, S.-K. (2026). Scaffolding Critical Thinking with Generative AI: Design Principles for Integrating Large Language Models in Higher Education. *Computers and Education: Artificial Intelligence*. <https://doi.org/10.1016/j.caeai.2026.100572>

Potkalitsky, N. (2025–2026). *Thinking with AI*. Substack. Framework per l'autonomia cognitiva in contesti educativi mediati da AI.