

10. settembre

I gemelli digitali come specchi deformanti

Gli specchi dovrebbero pensare più a lungo prima di riflettere.

Jean Cocteau

I gemelli digitali, modelli basati su grandi modelli linguistici arricchiti con informazioni personali dettagliate per simulare il comportamento di un individuo specifico, rappresentano una strategia proposta a scienziati sociali, responsabili politici e organizzazioni per ottenere più rapidamente informazioni sulle popolazioni a costi inferiori rispetto alla ricerca tradizionale su soggetti umani.



I modelli linguistici su larga scala (LLM) hanno la capacità di trasformare radicalmente la misurazione psicologica superando le barriere cognitive, etiche e pratiche di lunga data inerenti alla ricerca con partecipanti umani

Attraverso la simulazione scalabile delle risposte umane, gli **LLM** offrono non solo una raccolta dati più rapida ed economica, ma anche dati potenzialmente "migliori" per ricercatori e manager: gli **LLM** non si stancano, possono essere intervistati ripetutamente e possono simulare scenari ad alto rischio o potenzialmente dannosi senza rischi etici.

Sebbene gli **LLM** possano mostrare un comportamento simile a quello umano e replicare in modo direzionale gli effetti del trattamento sperimentale molti ricercatori hanno espresso scetticismo, sia su basi teoriche che empiriche. Gli studi hanno dimostrato che gli LLM spesso faticano a riprodurre le opinioni umane, tendono a mancare di eterogeneità e sovrastimano le dimensioni dell'effetto sperimentale.

Per affrontare queste preoccupazioni, esperti e professionisti hanno proposto l'utilizzo di gemelli digitali: modelli LLM di base (ad esempio, GPT e Llama) arricchiti con informazioni personali dettagliate per simulare specifici comportamenti individuali

A prima vista, i gemelli digitali rappresentano una soluzione straordinariamente elegante, offrendo quella che sembra essere una panacea per le limitazioni fondamentali documentate nelle precedenti ricerche sui dati sintetici.

In primo luogo, modelli distinti e personalizzati offrono il potenziale per infondere la varianza umana nelle risposte, superando l'omogeneizzazione che spesso affligge i modelli generali

A. Wang, J et al Large language models that replace human participants can harmfully misportray and flatten identity groups. Nat. Mach. Intell. 7, 400–411 (2025).

In secondo luogo, ancorare il modello a dati umani ricchi e reali può fornire i dettagli necessari per prevenire allucinazioni o confabulazioni e affrontare la sicofanzia e la normatività predefinite che minano l'accuratezza del modello di base. Insieme, ciò consentirebbe "simulazioni che riflettano meglio la miriade di fattori, spesso idiosincratici, che influenzano gli atteggiamenti, le credenze e i comportamenti degli individui"

J. S. Park et al Generative agent simulations of 1,000 people. arXiv:2411.10109 [cs.AI] (2024)

Il fascino dei gemelli digitali è ulteriormente amplificato dalle applicazioni uniche che rendono possibili, dall'intelligenza artificiale (IA) agentiva, ad esempio negoziare, lavorare o frequentare qualcuno per conto di un'altra persona, all'autoriflessione, in cui gli individui interagiscono con copie digitali di se stessi.

Date queste promesse, le applicazioni di digital twin vengono freneticamente create, implementate e vendute in ambito accademico e industriale. *Ad esempio, la startup simile.ai, basata su, ha recentemente raccolto 100 milioni di dollari, unendosi a un elenco crescente di startup di dati sintetici con valutazioni dell'ordine di 1 miliardo di dollari (ad esempio, Aaru).*

Nonostante il fervido interesse, le basi empiriche dei gemelli digitali rimangono sorprendentemente esigue.

Ad oggi, solo due studi iniziali hanno esaminato direttamente se i gemelli possano fungere da surrogati umani.

Uno studio ha creato gemelli digitali di oltre 1000 individui sulla base delle trascrizioni di interviste di 2 ore e ha riportato risultati incoraggianti (accuratezza relativa dell'85% sul General Social Survey, basata sul rapporto tra l'accuratezza del gemello digitale e l'accuratezza del test-retest). Tuttavia, i dati non sono disponibili pubblicamente, il che rappresenta un ostacolo sostanziale alla comprensione scientifica dei gemelli digitali.

Il secondo studio ha sviluppato Twin-2K-500, un dataset pubblico che consente la creazione di gemelli digitali di oltre 2000 individui sulla base delle loro risposte a più di 500 domande e ha riportato un'accuratezza relativamente elevata dei gemelli digitali sui dati di test (accuratezza media del 72%, accuratezza relativa dell'88% basata sul rapporto tra l'accuratezza del gemello digitale e l'accuratezza del test-retest).

Come primo test, tuttavia, la sua portata era necessariamente limitata: i gemelli non sono stati confrontati con parametri di riferimento come i profili basati esclusivamente su dati demografici.

Inoltre, i gemelli hanno replicato solo circa la metà degli effetti sperimentali osservati negli esseri umani. In aggiunta, in entrambi gli studi, la validazione si è basata su un insieme relativamente ristretto di esperimenti e sondaggi ben noti, che potrebbero sovrapporsi ai dati di addestramento del modello di base, sollevando preoccupazioni in merito alla diffusione di informazioni fuorvianti

J. Ludwig, S. Mullainathan, A. Rambachan, Large language models: An Applied econometric framework. Annu. Rev. Econ. 18, 283–316 (2026).

Di conseguenza, il settore sta rapidamente adottando una tecnologia priva di un'essenziale validazione empirica, lasciando aperto il rischio di prendere decisioni importanti basate su un'illusione di personalizzazione piuttosto che su una reale comprensione a livello individuale.

I gemelli digitali, modelli basati su grandi modelli linguistici arricchiti con informazioni personali dettagliate per simulare il comportamento di un individuo specifico, rappresentano una strategia proposta a scienziati sociali, responsabili politici e organizzazioni per ottenere più rapidamente informazioni sulle popolazioni a costi inferiori rispetto alla ricerca tradizionale su soggetti umani.

Un megastudio che ha coinvolto ricercatori della **Columbia** e della **Yale University**



Peng T et al

Digital twins are funhouse mirrors:

Five systematic distortions.

Sci Adv. 2026 Sep 4;12(36):eae8260.

Ha analizzato **19 esperimenti preregistrati**, comprendente 164 risultati distinti, ha confrontato un campione rappresentativo di 1784 adulti statunitensi con gemelli digitali addestrati su oltre 500 risposte precedenti per partecipante.

Lo studio ha appurato che i gemelli digitali si sono dimostrati **promettenti** nella classificazione degli individui, **ma sono risultati di poco superiori ai modelli linguistici di base nel produrre previsioni accurate delle risposte individuali.**

Tuttavia, i gemelli digitali hanno previsto le risposte assolute in modo simile a profili demografici molto più semplici. I ricercatori hanno documentato cinque distorsioni, fornendo una tabella di marcia diagnostica per lavori futuri: insufficiente individuazione, stereotipizzazione, bias di rappresentazione, pregiudizi ideologici e iperrazionalità.

