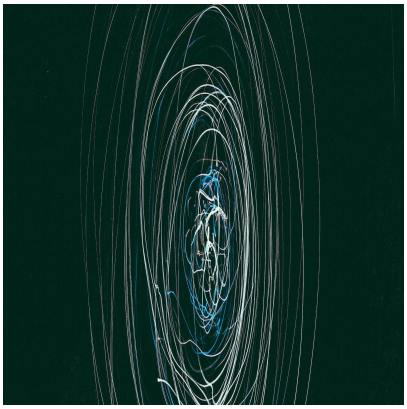

â??capacitÃ emergentiâ?• e â??opacitÃ cognitiveâ?• dellâ??IA: una sfda per la democrazia?

Descrizione

I sistemi di intelligenza artificiale generativa (IA) hanno compiuto un salto qualitativo, per certi versi imprevedibile. Oltre a generare testi e immagini, i modelli piÃ¹ recenti mostrano â??capacitÃ emergentiâ?•. AbilitÃ non programmate esplicitamente che compaiono spontaneamente al crescere della scala computazionale. Ragionamento analogico, pianificazione multi-step, comprensione del contesto. Competenze considerate esclusivamente umane. Questo fenomeno pone interrogativi profondi e la questione della *governance* diventa urgente. Lâ??Unione europea con lâ??*AI Act* ha tentato un primo inquadramento normativo ma la velocitÃ dellâ??innovazione supera quella della legislazione.

Una prima considerazione verte sulla distinzione tra riproduzione intelligente e produzione dellâ??intelligenza. Ovvero tra IA ingegneristiche o riproduttive e IA cognitive o produttive. Secondo questa distinzione lâ??evidenza data da una potenza di calcolo di gran lunga superiore a quella umana significa unâ??alternativa certamente molto utile ma non si puÃ² negare che potrebbe rappresentare anche una prospettiva evolutiva, per certi versi inquietante. Cosa resterebbe della nostra identitÃ incarnata? Stiamo assistendo alla progressiva migrazione del pensiero dallâ??umano allâ??astrazione del calcolo? Un pensiero incorporeo che, abbandonando il limite, si fonda nel computazionale? Siamo di fronte a una nuova forma di soggettivitÃ cognitiva o ci riflettiamo in uno specchio algoritmico scambiando lâ??eco per una voce e il riflesso per un volto?

CapacitÃ emergenti

Analizziamo il fenomeno delle â??capacitÃ emergentiâ?•. In pratica, a mano a mano che aumentano i parametri e i dati di addestramento, questi sistemi raggiungono una complessitÃ tale da far emergere funzionalitÃ inaspettate, non programmate esplicitamente dagli sviluppatori. Possiamo dire che assistiamo a un salto di stato qualitativo, in cui il sistema architetta in modo indipendente strategie risolutive mai tracciate prima, perfezionando spontaneamente la propria abilitÃ nel generare risposte che sorprendono per rigore logico e complessitÃ strutturale.

Giunti a questo punto, potrebbe essere lecito ritenere che lâ??emergere di tali capacitÃ cognitive puÃ² essere considerato un punto di convergenza tra lâ??architettura dei modelli linguistici artificiali e il pensiero umano? Questa inaspettata evoluzione dellâ??IA rappresenterebbe il segnale che si stanno ricalcando i processi cognitivi profondi della mente umana?

Sul tema la posizione dei ricercatori Ã¨ contrastante. Per alcuni stiamo assistendo a una svolta decisiva mentre, per altri, permangono dubbi in merito allâ??interpretazione del fenomeno. Concorde Ã¨ la considerazione che la capacitÃ emergente si rileva nei modelli piÃ¹ grandi mentre Ã¨ assente nei modelli piÃ¹ piccoli. Vale a dire che non si tratta di un semplice affinamento di *skills* preesistenti, bensÃ¬ di qualcosa qualitativamente nuovo. Il modello â??imparaâ?• a farlo solo superando una soglia critica di scala: numero di parametri, qualitÃ /quantitÃ dei dati e degli algoritmi di *training*.

OpacitÃ cognitive

Ebbene, cosÃ¬ descritte, a queste â??capacitÃ emergentiâ?• corrispondono anche â??opacitÃ cognitiveâ?•. Significa che per quanto si possa perfino avere conoscenza di tutti i dati interni al sistema, non siamo in grado di ricostruire con chiarezza perchÃ© una certa decisione Ã¨ stata presa. La motivazione Ã¨ riconducibile ad una varietÃ di fattori. In particolare, lâ??ostacolo non Ã¨ solo il vuoto informativo ma lâ??eccesso di complessitÃ . Ci troviamo davanti a sistemi di IA dove con innumerevoli variabili e relazioni non lineari si producono effetti dâ??insieme non riconducibili a una semplice somma delle parti. Una classica esemplificazione, a tal proposito, Ã¨ riconducibile al volo di uno stormo di uccelli. Ogni singolo uccello segue regole semplici di volo ma il movimento collettivo dello stormo non puÃ² essere spiegato sommando i comportamenti dei singoli uccelli.

Significa che le tradizionali spiegazioni non sono piÃ¹ sufficienti. Se non si Ã¨ in grado di dare una risposta alla classica domanda sul perchÃ© un sistema di IA ha assunto una determinata decisione (â??opacitÃ cognitiveâ?•), la questione si sposta sul piano funzionale. La decisione assunta dal sistema di IA Ã¨ utile? Funziona? Insomma, si passa dal piano della esplicabilitÃ a quello del mero funzionalismo. Proprio perchÃ© il sistema di IA ha operato allâ??interno di spazi decisionali inconoscibili, si sancisce il passaggio dalla cognizione effettiva alla pura osservazione passiva del risultato e dei suoi effetti.

Il cuore del problema, pertanto, consiste nel combinato disposto tra â??capacitÃ emergentiâ?• e â??opacitÃ cognitiveâ?•, da cui la â??validazione funzionaleâ?• con il seguente assunto: lâ??efficacia di un risultato conta piÃ¹ della comprensione del percorso che lo ha generato. Non Ã¨ piÃ¹ necessario chiedersi il â??*perchÃ©*â??, in quanto ci Ã¨ tecnicamente impossibile saperlo, ma solo che lâ??*output* â??*funzioni*â?•.

Accettare risultati senza comprenderne il processo decisionale?

Sebbene questo approccio funzionalista consenta di gestire complessitÃ altrimenti inaffrontabili, porta con sÃ© rischi strategici ed etici profondi. Non piÃ¹ in grado di decifrare i criteri sottostanti, il ruolo dellâ??umano muta da decisore consapevole a mero fruitore che approva *output* nati in spazi che gli sono preclusi.

Lo slittamento, dalla comprensione alla passiva accettazione, Ã¨ la frattura epistemica piÃ¹ insidiosa dellâ??IA in quanto le â??capacitÃ emergentiâ?• e le â??opacitÃ cognitiveâ?• non viaggiano separati. Si alimentano a vicenda. E mettere in discussione un sistema, in grado di evidenziare capacitÃ non

previste ma che â??funzionanoâ?•, si rischierebbe di apparire ostili al progresso, nostalgici di unâ??inefficienza ormai superata.

A fronte del funzionalismo della tecnica si pone il diritto a comprendere una decisione che ci riguarda. Diritto fondativo in qualsiasi sistema che si voglia chiamare democratico ma che le â??opacitÃ cognitiveâ?• vanificano.

La risposta piÃ¹ diffusa a questo problema sarebbe lâ??*Explainable Artificial Intelligence* (XAI) che consiste nello sviluppo di tecniche che possono rendere interpretabili le decisioni assunte dallâ??IA. Ã? un orientamento certamente necessario ma non Ã¨ sufficiente. Le spiegazioni prodotte dai sistemi di XAI sono spesso approssimazioni *post-hoc*, non descrivono il processo reale di elaborazione ma costruiscono una narrazione plausibile a partire dal risultato. In un certo senso possiamo dire che rappresenta una opacitÃ di secondo livello. Da un lato non sappiamo come il sistema ha sviluppato un determinato *output* ma abbiamo un risultato e una storia che sembrano ragionevoli. Il rischio Ã¨ che quella storia diventi il surrogato della comprensione e che la *governance* si accontenti di esso.

La questione richiede un qualcosa di piÃ¹ profondo. Possiamo dire una nuova cultura della responsabilitÃ digitale. Capace di tenere insieme ciÃ² che lâ??IA sa fare e ciÃ² che la societÃ deve poter comprendere e governare. Nel neoumanesimo digitale questa Ã¨ una sfida per la democrazia.

Crediti foto: Gertruda Valaseviciute su Unsplash

Data di creazione
22 Maggio 2026
Autore
lucio_romano