

Lorenzo Perilli

**Coscienza artificiale. Come le macchine
pensano e trasformano l'esperienza umana**

Il Saggiatore, Milano 2025, pp. 308

€ 21.00, ISBN 9788842833932

Giulio Amore

Università degli Studi di Padova

Le tematiche relative alle capacità cognitive, esperienziali, percettive dei sistemi basati sull'intelligenza artificiale (IA) sono oggi discusse all'interno di molteplici discipline.

All'interno di questo dibattito, il taglio che Lorenzo Perilli, professore ordinario all'Università di Roma Tor Vergata, dà nell'opera *Coscienza Artificiale. Come le macchine pensano e trasformano l'esperienza umana* è senza dubbio degno di nota. Soprattutto, lo è se consideriamo come centrale e fondamentale l'idea di uno scambio reciproco e fruttuoso tra le discipline cosiddette scientifiche e umanistiche, proprio come avverte Perilli nel Prologo.

Il libro infatti affronta una questione spesso tralasciata: in che modo le categorie, i concetti, il lessico, che noi utilizzavamo (e utilizziamo ancora) per descrivere le capacità cognitive ed esperienziali e che, a nostra volta, abbiamo traslato in campo informatico per descrivere i "comportamenti" delle IA, si stiano modificando e adattando proprio a questi stessi sistemi.

In altre parole, invece di adattare gli schemi concettuali relativi alla descrizione degli algoritmi e delle loro performance a quelli tradizionalmente riservati all'essere umano e alle sue attività, cognizioni o esperienze, l'autore rintraccia attraverso una precisa analisi storico-concettuale la tendenza opposta: un adattamento delle nostre categorie al lessico e alle performance delle macchine. Citando direttamente Perilli: "Il punto è che noi tendiamo a adattare il nostro concetto di intelligenza, il nostro concetto di coscienza, a quel che le macchine algoritmiche fanno o non fanno fare. Stiamo progressivamente ma rapidamente modificando le nostre categorie." (p. 19).

All'interno della prima parte, l'autore comincia l'analisi di questa tendenza, partendo da come la tecnologia contemporanea stia sollecitando più di ogni altro tipo di riflessione (teologica, metafi-

sica, antropologica, scientifica) la domanda “[c]he cos’è l’uomo?” (p. 26) e di come essa stia producendo un rovesciamento tra soggetto e oggetto, soprattutto in relazione alle idee di attività e passività connesse ai rispettivi concetti. “Da *homo faber*, autore ciascuno del proprio destino, parte attiva nel mondo, creatore, artigiano nel senso più nobile della parola [...] saremo dunque osservatori, fruitori, consumatori. *Homo spectator*.” (p. 28); da soggetti attivi, a osservatori passivi, da agenti che dispongono di mezzi in loro possesso a semplici fruitori di sistemi complessi la cui struttura ed il funzionamento sfugge nei suoi dettagli. È qui, infatti, che si mostra per Perilli lo scarto tra le tecnologie che potremmo definire tradizionali e quelle forme di IA, come i Generative Pre-trained Transformers (GPT): la possibilità di mantenerne sotto controllo i processi, il funzionamento. Se è pur vero, infatti, che i risultati di questi GPT son stati oltre ogni aspettativa, è anche vero che questa opacità, non trasparenza del calcolo, può trasformarsi in un problema.

La seconda parte, la più corposa del volume, continua il lavoro cominciato nella prima, andando a sollecitare la riflessione sugli slittamenti concettuali operati dalla nostra cultura per adattarsi agli algoritmi, a partire dalle riflessioni e dai testi propri degli autori chiave della storia di questi sistemi, come Butler, Wiener, Turing, McCarthy, von Neumann, Rosenblatt e molti altri.

Lo scopo è infatti mostrare come il rovesciamento categoriale contemporaneo, derivante dall’introduzione massiccia delle IA, affondi le sue radici all’interno della storia di questi stessi sistemi e di come molto spesso si siano appiattiti concetti complessi per poterli usare in ambito informatico.

Questo appiattimento, tuttavia, ha in sé un effetto collaterale, che è appunto la ri-significazione dei concetti stessi: “[p]ercezione, sensazione, coscienza, intelligenza, tutto viene normalizzato e reso univoco, anche laddove non lo è.” (p. 74). Così, la stessa formula intelligenza-artificiale, sostiene l’autore, si è nutrita di un’ambiguità di fondo proprio della definizione di intelligenza, in particolare, nella forma data da uno dei padri dell’IA, John McCarthy. La sua definizione, infatti, risulterebbe incompleta, a tratti superficiale, tutta schiacciata sulla capacità di raggiungere obiettivi o semplicemente di risolvere problemi. Questo superficialità ha portato effettivamente, secondo Perilli, ad una trasformazione del concetto di intelligenza: “il risultato è che oggi il concetto di intelligenza si è ridotto ad una capacità operativa. Si è trasformato

per adattarsi all'ambiente delle macchine. [...] oggi è l'intelligenza umana che si ridefinisce [...] prendendo a modello l'intelligenza algoritmica del calcolatore." (pp. 82-83).

In maniera analoga, anche i concetti di memoria e apprendimento rischiano di perdere la loro stratificazione, la loro complessità, in nome di una riduzione concettuale che li renda più adatti alla spiegazione di determinate componenti delle macchine e degli algoritmi, oppure quello di scelta, basato sul giudizio, ben diverso da quello di decisione, processo invece computazionale.

Uno spazio più ampio viene dedicato al concetto di coscienza e all'ipotesi di poterla realizzare artificialmente. Questa ipotesi si fonda su un principio, quello di invarianza organizzativa, "e prevede che due sistemi con una organizzazione funzionale identica in ogni dettaglio avranno esperienze qualitativamente identiche." (p. 136). Tutto questo si realizza nell'idea di poter sostituire le varie componenti dell'esperienza umana (cognitiva, percettiva, di coscienza) con algoritmi e chip funzionalmente analoghi e dunque ottenere esperienze medesime. Il rischio, ancora una volta, non è necessariamente di natura distopica, quanto molto più concretamente parlando, di una reale riduzione di portata dei nostri concetti e categorie fondamentali: "intelligenza, coscienza, mente, cervello, percezioni, *vengono a modificarsi* sotto l'influenza dei loro equivalenti digitali. Siamo cioè portati a interpretare e rappresentare noi stessi per analogia con la rappresentazione digitale di noi stessi e del nostro mondo. Il pensiero algoritmico prevede che non sia più l'algoritmo a imitare i nostri processi mentali, ma i nostri processi mentali a essere considerati come algoritmici." (p. 175).

Tuttavia, il mondo degli algoritmi è disomogeneo a quello che potremmo chiamare il mondo della vita. L'algoritmo infatti si fonda su un calcolo, su un processo, in larga parte determinato da univocità, non-ambiguità, correttezza formale, effettività, strutturazione, riproducibilità, "[m]a piuttosto che prendere coscienza di questo e considerare l'approccio algoritmico come uno strumento che aiuta in modo straordinariamente efficace alcune forme di conoscenza, [...] si è pensato bene di considerare il mondo della vita nella sua interezza come inadatto e disfunzionale." (p. 179).

Un mondo interamente calcolabile, algoritmizzabile, controllabile e manipolabile da software, rappresenta nelle parti III e IV, connesse dal punto di vista concettuale, il fondamento di una società basata sul controllo e su un'organizzazione che ruoti attorno all'efficienza dei sistemi, all'interno dei quali i cittadini rappresen-

tano utenti, consumatori. Purtroppo, però, nel mondo dei software proprietari, l'utente non ha potere di nessuna sorta. Il controllo rimane nelle mani di chi "produce, controlla e guadagna." (p. 230).

Lo slittamento concettuale viene così sostenuto e incoraggiato, nell'ottica di far sembrare sempre più simili a noi questi sistemi, con l'effetto di renderli sempre più indispensabili, non solo dal punto di vista pratico, tecnico, ma anche dal punto di vista della definizione e autocoscienza di noi stessi.

Degno di nota è anche l'Appendice (pp. 255-267), in cui Perilli elenca le principali applicazioni in ambito medico dei sistemi basati sull'IA, un campo, questo, in cui l'introduzione e lo sviluppo di questi sistemi è divenuto ormai insostituibile e che ha già registrato grandi successi.

L'opera di Perilli risulta essere un volume molto scorrevole, diviso in brevi capitoli, brevi excursus, che attraverso l'analisi del lessico specifico della disciplina ed il confronto con uno più genuinamente filosofico, guidano il lettore attraverso un percorso storico-concettuale che ha come scopo quello di mettere in guardia, di avvertire e suonare una sorta di campanello di allarme. L'analisi storica delle fonti permette di dare profondità alle tesi di Perilli, che non vanno derubricate in nessun modo a mero allarmismo o tecno-pessimismo. Lo scopo primario del libro rimane quello di portare alla luce una discussione intorno ad un'inevitabile *messaggio alla prova* dei nostri concetti di fronte ai successi dell'IA. Farlo non è semplicemente una prassi utile ma è anche periodicamente necessario. Tuttavia, l'autore ci mette in guardia da un fattore all'interno di questa discussione: appiattire e ridurre concetti come quelli di intelligenza, coscienza, esperienza, percezione, in nome di una sorta di adattabilità concettuale porta con sé il rischio di perdere una complessità fondamentale che non va però assolutamente sacrificata.