

## L'INTELLIGENZA ARTIFICIALE

## Il potere degli algoritmi non abbiamo capito nulla

LUCIANO FLORIDI

Nell'Intelligenza artificiale si parla di allucinazioni per descrivere errori o deviazioni nell'output di un modello rispetto a rappresentazioni fondate e accurate dei dati di input. C'è grande differenza con le nostre allucinazioni, esperienze percettive disturbanti che si sviluppano in assenza di stimoli esterni. - PAGINA 22



## L'INTERVENTO

# Luciano Floridi Anche l'AI è stata fraintesa

*Pubblichiamo un estratto di un recente articolo del filosofo Luciano Floridi apparso sull'ultimo numero della rivista Minds and Machines, tradotto da Bruno Ruffilli*

L'intelligenza artificiale (AI) può generare fraintendimenti in molti modi. Se gli sviluppi vertiginosi del software e dell'hardware sono al di là della portata della maggior parte di noi, forse la fonte di confusione più profonda deriva dal vocabolario tecnico dell'AI. Così affollata di termini derivati dalle scienze cognitive e dalle neuroscienze (BCS, che comprendono le scienze cognitive e le neuroscienze), l'AI acquisisce proprietà biologiche e cognitive ingiustificate che ne inficia-

no la comprensione. A loro volta, le discipline scientifiche che studiano le funzioni cerebrali alla base dell'apprendimento e del comportamento hanno preso sempre più in prestito dalle scienze informatiche e computazionali su cui si basa l'AI, trasformando l'entità biologica più complessa e multiforme che conosciamo in una semplice macchina calcolatrice.

Gli studiosi di intelligenza artificiale parlano ad esempio di "apprendimento automatico", espressione coniata (o resa popolare, il dibattito è aperto) da Arthur Samuel nel 1959 per indicare «lo sviluppo e lo studio di algoritmi statistici in grado di apprendere dai dati e di generalizzare a nuovi dati, e quindi di eseguire compiti senza istruzioni esplicite». Ma questo apprendimento non significa ciò che neuroscienziati e psicologi cognitivi intendono quando si riferiscono al modo in cui esse-

ri umani o animali acquisiscono nuovi comportamenti o contenuti mentali, o modificano quelli esistenti, come risultato di esperienze nell'ambiente. Allo stesso modo, nell'AI si parla di allucinazioni per descrivere errori o deviazioni nell'output di un modello rispetto a rappresentazioni fondate e accurate dei dati di input. C'è grande differenza con le nostre allucinazioni, esperienze percettive disturbanti che si sviluppano in assenza di stimoli esterni.

Per spiegare questa confusione dobbiamo fare un passo indietro e partire da un'idea di Carl Schmitt, che osserva come «tutti i concetti significativi della moderna teoria dello Stato sono concetti teologici secolarizzati». Ad esempio, nozioni politiche come sovranità, stato di eccezione, volontà sovrana, onnipotenza della legge e legittimità, possono essere ricondotte a concetti teologici: questo prestito concettuale non ha elimi-

nato la struttura o l'influenza dei concetti teologici, ma li ha ricontestualizzati in un quadro secolare.

Non è solo un'osservazione storica, ma anche una critica severa. Il prestito concettuale limita la valutazione critica dei concetti politici proprio a causa delle loro radici teologiche, dalle quali non si sono del tutto emancipati, mentre le dinamiche di potere e i processi decisionali in politica riflettono ancora le strutture stabilite dal pensiero religioso.

Queste considerazioni possono estendersi ad altre discipline. Quando emergono nuove scienze, mancano di un vocabolario tecnico per descrivere e comunicare i loro fenomeni, problemi, ipotesi, osservazioni, formulazioni, teorie, ecc. È urgente la necessità di essere precisi, chiari, coerenti e sintetici; di concordare le definizioni e promuovere la standardizzazione. Le lacune ven-

gono colmate inventando nuovi termini, utilizzando traduzioni dal greco o dal latino, o adottando e adattando espressioni tecniche di altre discipline. L'intelligenza artificiale si è sviluppata molto rapidamente e ha avuto bisogno di prendere in prestito il suo lessico da campi affini: la cibernetica, la logica, l'informatica e la teoria dell'informazione; e soprattutto le scienze che studiano il modo di agire e il comportamento umano e animale e le loro basi biologiche. Il fenomeno si è sviluppato a partire da Alan Turing, che ha influito in modo determinante sul parallelismo con l'intelligenza e il comportamento umani per spiegare come le macchine avrebbero potuto imitare alcuni aspetti della cognizione biologica. Ma probabilmente

il prestito più problematico è avvenuto con l'etichetta che definisce l'intero ambito: "Intelligenza Artificiale", creata dallo scienziato americano John McCarthy a metà degli anni Cinquanta.

Oltre ad apprendimento, utilizzato per *machine learning*, i termini biologici e psicologici nell'AI sono numerosi; ricordiamo, ad esempio, "adattamento", "computer vision", "memoria". Ma sono molti anche i termini con significati tecnici poco o per nulla correlati al senso che hanno nel loro contesto scientifico originale. Prendiamo il caso di "attenzione", un termine estremamente popolare introdotto di recente nell'apprendimento automatico. In BCS si riferisce in generale ai processi di prioritizzazione dei segnali neurali o

psicologici rilevanti per guidare il comportamento adattivo nel contesto corrente, e il sostantivo è spesso accompagnato da altri elementi qualificatori (ad esempio, attenzione selettiva, spaziale, basata sugli oggetti, sulle caratteristiche). Il significato nell'apprendimento automatico è molto diverso, come testimonia anche Wikipedia: «L'attenzione è un meccanismo, all'interno delle reti neurali, in particolare dei modelli basati sui transformer, che "calcola pesi morbidi per ogni parola, più precisamente per il suo incorporamento, nel contesto». È un caso di polisemia, se non di omonimia: le differenze scientifiche tra i due concetti sono significative e profonde, le somiglianze superficiali e trascurabili, eppure il ba-

gaglio psicologico e biologico esercita un potere semantico che spinge verso un maggiore antropomorfismo.

Dall'altra parte, le scienze cognitive e neuroscienze hanno preso in prestito i costrutti tecnici e quantificabili dalla teoria dell'informazione e dalle scienze informatiche, inquadrando il cervello e la mente come sistemi computazionali e di elaborazione delle informazioni. Per esempio, Ulric Neisser, nel testo che segna la nascita della Psicologia Cognitiva, sostiene che «il compito di uno psicologo che cerca di capire la cognizione umana è analogo a quello di un uomo che cerca di scoprire come è stato programmato un computer. In particolare, se il programma sembra immagazzinare e riutilizzare informazioni, vorrebbe sapere con quali routine o procedure ciò avviene». Anche qui l'e-

### Arthur Samuel



Pioniere nel campo dell'AI e dei videogiochi, conìò il termine "machine learning" nel 1959. Scrisse che «Lo sviluppo di algoritmi statistici in grado di apprendere dai dati e di generalizzare a nuovi dati, e quindi di eseguire compiti senza istruzioni esplicite»

**È fuorviante l'idea che le macchine possano imitare la nostra condizione biologica**



Luciano Floridi (1964), ordinario di filosofia ed etica dell'informazione all'Oxford Internet Institute dell'Università di Oxford e professore di Sociologia della comunicazione presso l'Università di Bologna