



## Culture

**TEMPI PRESENTI** Da Melanie Mitchell a Erik Larson, un percorso di letture recenti sull'intelligenza artificiale

Teresa Numerico pagina 10

## TEMPI PRESENTI

\* Un'analisi equilibrata e matura delle conseguenze e delle inadeguatezze intrinseche a questi sistemi

# Nell'artificio dell'intelligenza

Riflessioni intorno ai volumi di Melanie Mitchell e Erik Larson, pubblicati da Einaudi e FrancoAngeli

TERESA NUMERICO

■ ■ ■ L'intelligenza artificiale (IA) continua a mietere successi sempre più sorprendenti mentre cresce l'interesse presso il grande pubblico. Nell'ultimo decennio ci sono stati molti progressi non solo relativi a giochi come scacchi, dama, Go e alcuni videogiochi, ma anche in campi inaspettati come visione artificiale, comprensione e produzione di testi e immagini. Tra le tante letture sull'argomento ci sono due libri recentemente tradotti in italiano: *L'intelligenza artificiale. Una guida per esseri umani pensanti* di Melanie Mitchell (Einaudi, pp. 324, euro 26) e *Il mito dell'intelligenza artificiale*, di Erik Larson, (FrancoAngeli, pp. 273, euro 27). I volumi si distinguono per la capacità di affrontare in modo comprensibile, sebbene non semplicistico, temi complessi e per un atteggiamento equilibrato, maturo e illustrativo delle conseguenze e delle inadeguatezze intrinseche di questi sistemi.

**MITCHELL, UNA SCIENZIATA** informatica, ha una prospettiva possibilista, anche se critica alcune derive entusiaste rispetto ai presunti successi dei sistemi intelligenti. Ritiene che ci sia una grande differenza tra cimentarsi con gli scacchi e Go e comprendere un testo in profondità, differenza che consiste nel fatto che i giochi sono tutti rappresentabili in forma sintattica, e possiamo fornire con chiarezza la funzione obiettivo: ovvero sancire cosa costituisca un successo per il sistema e ci fa vincere la partita. Addestrare le

macchine a giocare significa avere tempo e forza computazionale, oltre a costruire una funzione di rinforzo per premiare i comportamenti che conducono alla vittoria. Diverso è usare gli stessi sistemi nel mondo reale, nel quale c'è bisogno di senso comune e di una capacità di intervento immediato di fronte a situazioni incerte, imprevedibili e non formalizzabili.

In queste circostanze è necessario usare metodi supervisionati, con dati di addestramento che includano tutte le variabili possibili, quindi, in casi come la guida autonoma non solo semafori e strisce pedonali frequenti, ma anche inondazioni, segnaletica orizzontale coperta dalla neve, condizioni di scarsa visibilità e via dicendo. Sono eventi della coda lunga, cioè scarsamente probabili, ma non impossibili. Eppure, non si può insegnare alla macchina a fronteggiare tutti gli imprevisti con una etichettatura esaustiva del mondo. La soluzione suggerita dall'informatico e imprenditore Andrew Ng è che i pedoni, per esempio, debbano essere addestrati a comportarsi in maniera prevedibile quando ci saranno le macchine a guida autonoma in circolazione. Sarà, cioè, l'ambiente circostante a adeguarsi alla capacità di visibilità e rappresentazione di cui abbiamo dotato le nostre macchine. Il mondo dovrà divenire prevedibile per consentire alle macchine di far valere le proprie capacità di anticipazione del futuro. Orizzonte non proprio auspicabile.

La critica principale di Mitchell consiste nella difficoltà di esercitare meccanicamente processi di astrazione complessi e individuare analogie rilevanti nella serie di eventi da classificare. L'elasticità umana si manifesta nell'esercizio dell'astrazione concettuale che interviene con pochi elementi di riconoscimento e collegamento e si attiva per individuare analogie - caratteristiche difficilmente replicabili nella macchina che apprende. A suo parere, quindi, sarebbe ineludibile inserire forme di ragionamento accanto alla struttura dell'apprendimento, per ottenere risultati più solidi nell'astrazione.

**SECONDO LARSON**, un imprenditore informatico, che ha all'attivo diverse start-up finanziate dall'agenzia americana della difesa Darpa, il problema principale degli strumenti intelligenti basati sull'apprendimento automatico è un po' diverso e attiene alle conseguenze negative delle inferenze induttive. Il ragionamento umano è invece concentrato sulle abduzioni, le inferenze degli investigatori: cercare le cause di una certa conclusione, anche se si procede costruendo collegamenti incerti e talvolta indimostrabili, basati su ipotesi e tentativi di verificarle.

L'inferenza induttiva si basa, invece, sull'assunzione implicita che il futuro sarà una proiezione diretta del passato rispetto alla regolarità dei comportamenti della serie in oggetto. Quando addestriamo un algoritmo all'inferenza induttiva dobbiamo comunque inserire un

bias, un pregiudizio di apprendimento, per indicare quali sono le regolarità da considerare, quanto complesse devono essere, quanto scarto dalla regola tollerare e così via; tali aspetti sono implicati direttamente nella costruzione delle regolarità scoperte.

Nei sistemi automatici di apprendimento dobbiamo anche fronteggiare la cosiddetta saturazione del modello. Quando aumentano i dati di addestramento, infatti, esiste una curva di miglioramento delle prestazioni che tende a degradare. Il sistema non continua a incrementare la performatività proporzionalmente alle informazioni, l'aggiunta di nuovi dati comporta vantaggi sempre meno significativi. È semplicistico e sbagliato ritenere che stiamo passando da una prima fase di successi parziali a una soluzione definitiva del problema dell'intelligenza.

Larson è, perciò, scettico nei confronti della Singolarità - cioè l'idea di costruire un'intelligenza generale come quella umana - perché non sappiamo come programmare l'inferenza abduitiva, studiata da Charles Sanders Peirce a cavallo della fine dell'Ottocento, alla quale si deve la nostra inventiva nello scoprire e riconoscere nuovi schemi concettuali. Il pensiero umano è interessante proprio a causa delle sue incertezze e della capacità di fare congetture con pochi elementi connessi in maniera labile. Tali processi cognitivi non si possono meccanicizzare perché richiedono grandi quantità di sapere implicito che noi stessi non siamo consa-



pevoli di usare, quando prendiamo decisioni che hanno effetto sulle nostre inferenze.

**LE CRITICHE DI LARSON** riguardano anche l'impegno nella lotta ai bias e alle discriminazioni dell'intelligenza artificiale induttiva, che non coglierebbe il punto circa l'inadeguatezza di questi strumenti a prendere decisioni e si limiterebbe a cercare di correggerne le storture più vistose. Per lui gli effetti discriminatori sono una conseguenza del mito che attribuisce alle macchine poteri che non posso-

no avere. La questione dell'allineamento tra i nostri scopi e quelli dell'intelligenza artificiale non nasce per i successi dei sistemi tecnici, ma per averli erroneamente elevati a prendere decisioni in contesti delicati, prima di averne accertato le credenziali.

I due volumi ci fanno riflettere sulle domande aperte poste dalle applicazioni di intelligenza artificiale e sulla politica epistemica per definire un nuovo concetto di intelligenza, secondo cui ci possiamo affidare a tali

dispositivi retti da sistemi altamente opachi per prendere decisioni in situazioni incerte. Il mito principale dell'IA riguarda il desiderio di delegare alle macchine la presa di decisione anche in contesti imprevedibili per non sostenerne la responsabilità, sebbene siamo consci di non poter meccanizzare quelle scelte. Purtroppo, i sistemi tecnici nei quali esternalizziamo le nostre capacità di intelligenza e memorizzazione sono come i paradigmi scientifici, di cui parla Kuhn, incommensura-

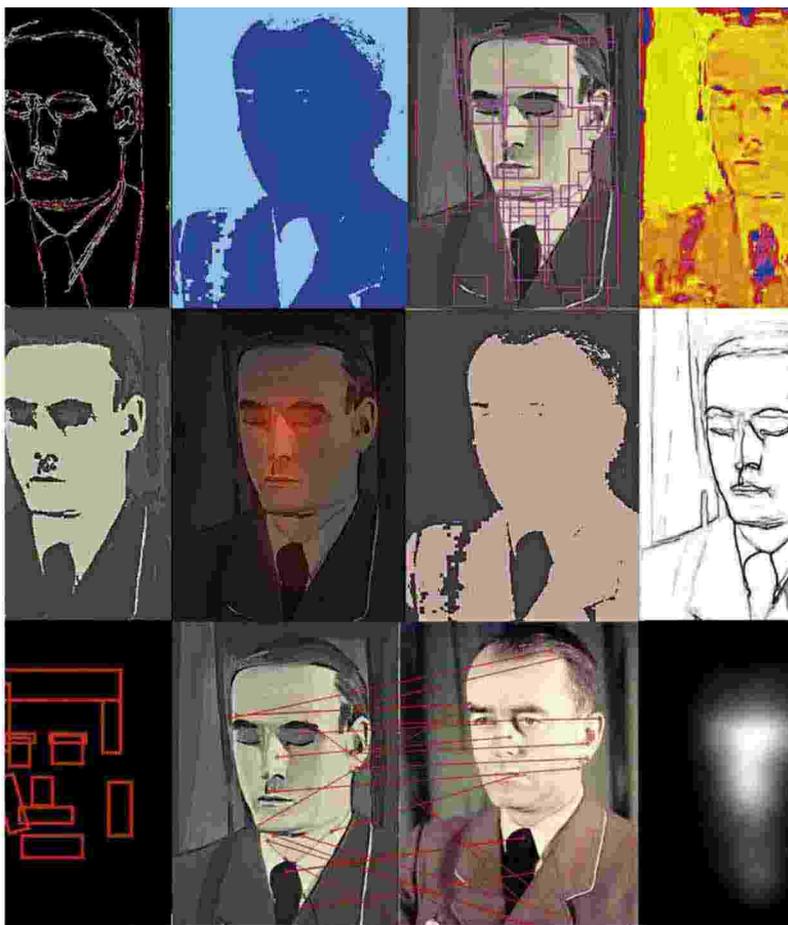
bili. Non siamo certi che sostituiremo un sistema a un altro a causa di criteri epistemici più efficaci nelle scoperte, rispetto all'affidabilità della conoscenza. Stiamo di fatto esercitando una validazione dell'esperienza differente da quella alla quale ci siamo attenuti finora. Stiamo cambiando le convinzioni che abbiamo costruito tra la Rivoluzione Scientifica e l'Illuminismo circa cosa legittima un sapere come credenza vera giustificata, affidandoci a quello che, per i non esperti, è poco più di un oracolo da interpretare.



*La scienziata informatica è possibilista, ma critica l'eccessivo entusiasmo. Il secondo, imprenditore del settore, riflette sulle conseguenze negative delle inferenze induttive*



*Le forme tecniche nelle quali esternalizziamo le nostre capacità sono come i paradigmi scientifici, di cui parlava il fisico e filosofo Thomas Kuhn: incommensurabili*



Particolare da «Secrets» di Luc Tuymans, foto di Paul Hester, dal progetto Muhal della Venice International University



Ritaglio stampa ad uso esclusivo del destinatario, non riproducibile.

003600