

**Margaret A. Boden, *L'Intelligenza Artificiale*, D. Marconi (a cura di), il Mulino, Bologna 2019, pp. 188, € 16.00, ISBN 9788815280954**

*Alberto Giustiniano, Università degli Studi di Padova*

Il saggio di Margaret Boden si presenta come una breve introduzione a quel vasto e altrettanto caotico ambito disciplinare che prospera sotto l'etichetta di Intelligenza Artificiale. Il lettore tuttavia non dovrà aspettarsi un testo divulgativo o manualistico e tantomeno un prodotto di quella pubblicitaria sensazional-scandalistica che purtroppo negli ultimi anni è divenuta particolarmente diffusa anche presso i cataloghi di editori autorevoli. Diversamente si troverà di fronte a una riflessione seria e teoreticamente profonda, scevra da banalizzazioni e semplificazioni, capace di mostrare la complessità di un fenomeno al contempo sociale, scientifico, tecnologico e filosofico. Il testo è costruito su un doppio registro che oscilla tra la ricostruzione storico concettuale e il *pamphlet* polemico. L'autrice riesce così nell'ardua impresa di fornire un quadro complessivo degli sviluppi scientifici e tecnologici dell'IA dalle sue origini ad oggi e al contempo di articolare il complesso dibattito epistemologico che vi è da sempre connesso. Il risultato è un'analisi in grado di mettere in evidenza la pervasività del fenomeno IA innanzitutto sul piano disciplinare, con la sua capacità di ridefinire i rapporti di cooperazione tra differenti ambiti scientifici, ma soprattutto per il suo inedito livello di integrazione circolare tra teoresi e implementazione sul piano della validazione della conoscenza.

Il volume è suddiviso in sette parti autonome ma strettamente collegate l'una con l'altra da rimandi interni resi coerenti dalla questione di fondo che attraversa l'intera riflessione e che viene fissata nel primo capitolo intitolato: "Che cos'è l'intelligenza artificiale?". In queste prime pagine l'autrice sgombra il campo da qualsiasi visione unificante o riduttiva della questione: se l'IA si pone come obiettivo di realizzare macchine in grado di svolgere operazioni simili a quelle delle menti allora il suo ambito di indagine è pluridimensionale e in parte oscuro. Non tutto ciò che riguarda la mente ha a che fare con ciò che definiamo "intelligenza" o "ragionamento" (per esempio la percezione e il controllo motorio) e,

a sua volta, nemmeno l'intelligenza può essere considerata una facoltà univoca semmai "uno spazio riccamente strutturato di diverse capacità di elaborazione dell'informazione" (p.7) il cui funzionamento nell'uomo non ci è affatto chiaro. Per questo, secondo Boden, l'IA non può essere considerata una scienza ma un conglomerato di tecniche e discipline rivolte alla chiarificazione (*obiettivo scientifico*) e, quando possibile, riproduzione (*obiettivo tecnologico*) di quelle capacità fisiche e psicologiche che permettono ai viventi di conseguire i loro obiettivi. Da questi presupposti è più semplice comprendere le ragioni della sua impressionante pervasività ed estensione multidisciplinare, che si riflette tanto sul suo sviluppo storico quanto sulla nozione chiave che sembra guidarla: il concetto di *macchina virtuale*. Per quanto riguarda il piano storico, l'autrice riesce a condensare in poche pagine (pp.12-24) un itinerario complesso, ricco di svolte e ripensamenti, mostrando il proliferare di influenze (matematiche, fisiche, biologiche, filosofiche e ingegneristiche) e il succedersi di paradigmi differenti che negli ultimi anni, maturata la consapevolezza della distanza forse incolmabile che li separa dalla definizione di un modello unitario, stanno trovando nel reciproco coordinamento una nuova vitalità. Tuttavia, è attraverso la definizione aperta di macchina virtuale che Boden permette al lettore di comprendere la complessità del fenomeno. Oggetto dell'IA non è la macchina fisica in sé (cioè il computer) ma ciò che essa *fa* o dovrebbero riuscire a fare: "una macchina virtuale [...] è piuttosto il *sistema di elaborazione dell'informazione* che il programmatore ha in mente quando scrive un programma, e che le persone hanno in mente quando lo usano" (p.9). Dunque la nozione di fondo sembra configurarsi in termini prettamente funzionali, lasciando aperte non solo le modalità e i contesti nei quali i suoi effetti dovrebbero manifestarsi ma anche l'accordo (si legga comprensione) sulla loro *pertinenza*. A questo va aggiunto che il rapporto tra macchina virtuale e macchina fisica è tutt'altro che cartesiano. Uno dei pregi del saggio consiste proprio nel mostrare il rapporto di implicazione reciproca tra la possibilità di sviluppo di macchine virtuali e la concezione di macchine fisiche capaci di funzionare. Diviene sempre più chiaro, anche alla luce delle ricerche nel campo della complessità computazionale, che l'intero

ambito dell'IA sembra focalizzarsi sullo spazio di interpenetrazione tra hardware e software piuttosto che su uno solo dei due poli. Questa prospettiva emerge in maniera chiara dal secondo capitolo intitolato "Il Santo Graal dell'intelligenza artificiale" (pp.25-58). Qui l'autrice mette in evidenza il legame tra la proliferazione di differenti modalità di elaborazione di informazioni e la consapevolezza crescente della grande distanza che ci separa dalla possibilità effettiva di realizzare una macchina in possesso di intelligenza artificiale generale (AGI). La capacità computazionale non è sufficiente di per sé: per realizzare le prestazioni di alto livello che ci si auspica di ottenere sono necessarie strategie capaci di rendere i problemi in oggetto trattabili, ovvero computabili attraverso una serie di passi ben definiti (algoritmi). Questo passaggio rappresenta il limite intrinseco dell'impresa ma paradossalmente è proprio ad esso che dobbiamo la moltiplicazione di strategie di semplificazione e trattamento dei problemi alle quali si devono gli sviluppi impressionanti degli ultimi anni. Questi approcci vengono analizzati da Boden e ricondotti a problemi di euristica, pianificazione, semplificazione matematica e rappresentazione della conoscenza. L'aspetto interessante di queste pagine consiste nell'attenzione rivolta, per ognuno dei casi presi in esame (*problema del frame*, *machine learning*, *computer vision*, cognizione distribuita ecc.), sia ai limiti intrinseci che impediscono il raggiungimento di una prestazione intelligente completa che ai risultati prodotti dallo sviluppo di queste singole strategie. Tali ricadute non si limitano, mostra l'autrice, all'ambito tecnologico ma stanno contribuendo al progresso di altre discipline come la biologia, la linguistica, la psicologia, l'epistemologia, la teoria dell'evoluzione e la matematica. L'orizzonte è quello della reciproca influenza e non dell'assimilazione, non fosse altro per il fatto che queste strategie sono state assunte in origine dall'IA in quanto prodotti di quelle stesse discipline che ora hanno la possibilità di riflettere sui risultati e i fallimenti dei loro assunti grazie a questo differente contesto di indagine. Esempi illuminanti di progressi scientifici e teorici generati dai tentativi falliti nel conseguimento dell'AGI sono forniti nel terzo capitolo, dal titolo "Linguaggio, creatività, emozione" (pp.59-78). In questi tre ambiti funzionali la possibilità di progettare dispositivi in grado di fornire responsività adeguata all'utente umano è stata in

parte raggiunta, in special modo per il trattamento del linguaggio e la realizzazione dell'apprendimento non supervisionato, mentre per quanto riguarda le emozioni le difficoltà appaiono insormontabili. Tuttavia tali prestazioni sul piano scientifico possono fornire soprattutto indicazioni in merito alla complessità del funzionamento dell'intelligenza umana piuttosto che replicarla. In queste ricerche viene messa in evidenza la possibilità di realizzare sottoprestazioni specifiche e precisamente circostanziate (per esempio la traduzione di un testo o l'apprendimento situato) senza che ciò possa significare un reale avvicinamento al funzionamento complessivo della mente. Al contrario, le prestazioni più convincenti indicano di riflesso con la loro limitatezza contestuale e operativa un grado di integrazione tra i differenti piani che vanno a comporre la cognizione umana ben più profondo di quanto potessimo immaginare.

Da qui l'autrice passa in rassegna i cinque tipi di macchine virtuali intorno ai quali ruotano le principali linee di ricerca e finanziamento dell'IA contemporanea: IA classica o simbolica, le reti neurali artificiali (connessionismo), gli algoritmi evolutivi, gli automi cellulari e i sistemi dinamici (*Artificial Life*). Tra queste lo spazio maggiore è dedicato alle reti neurali (cap.4, "Le reti neurali artificiali", pp.79-100) e al campo dell'*A-Life* (cap.5, "I robot e la vita artificiale", pp.101-118) che rappresentano l'avanguardia della ricerca e dai quali sembrano giungere i risultati più interessanti. Lasciamo scoprire al lettore le caratteristiche di queste nuove tecnologie che Boden riesce a illustrare con grande efficacia. Ci limitiamo a segnalare due aspetti significativi. Per quanto riguarda le reti neurali artificiali i successi nella realizzazione di un'elaborazione parallela distribuita (PDP) e una maggiore conoscenza delle sue caratteristiche funzionali comportano profonde implicazioni sia sul piano della teoria neurofisiologica che in filosofia della mente. In particolare in ragione del venir meno in queste tecnologie della necessità di una struttura simbolica della rappresentazione regolata da un rigido programma di istruzioni fissato a monte. Il PDP comporta invece che la rappresentazione sia distribuita ed emerga come stato complessivo della rete in un dato momento. In questo consisterebbe il segreto della sua flessibilità e capacità di riconoscimento di pattern, anche incompleti, in assenza di una programmazione esplicita ma soltanto attraverso l'analisi di esempi.

Suggerendo una sorta di andamento a ritroso rispetto alle convinzioni nel campo della filosofia della mente che dall'innatismo di Chomsky sembra, secondo Boden, ricondurci alla "percezione come inferenza inconscia" di von Helmholtz. Quanto all'ambito dell'*A-Life*, lungi dal richiamare alla mente i sogni di Mary Shelley, i risultati più significativi segnalati da Boden riguardano gli algoritmi evolutivi. Questi particolari programmi sono progettati per generare la variazione causale e la selezione non causale tipica delle strutture genetiche degli organismi. In questo modo è possibile realizzare macchine virtuali in grado di migliorare nello svolgimento di un compito definito muovendo dal semplice succedersi di cicli (generazioni) di attivazione. Il software inizialmente non è in grado di svolgere il compito assegnato ma dopo un certo numero di generazioni sarà stato in grado di conservare quelle mutazioni causali che si sono rivelate più promettenti migliorando così la sua efficienza. Particolarmente sorprendente il caso illustrato dall'autrice di un sistema hardware progettato con algoritmi genetici dal cui funzionamento è emerso un primitivo sensore di onde radio grazie all'inclusione da parte del sistema di una serie di fattori ambientali presenti ma non presi in considerazione dagli ingegneri. Il dispositivo attraverso il funzionamento degli algoritmi genetici in esso contenuti ha *creato* una modalità di orientamento nell'ambiente che non era stata determinata a monte dai progettisti ma è emersa casualmente dalle possibilità offerte dal contesto circostante (pp.110-111).

Alla questione centrale del volume si giunge con il sesto capitolo dal titolo "Ma è vera intelligenza?" (pp.119-142). Qui Boden sottolinea come la possibilità di produrre in tutto o in parte alcune funzioni centrali della cognizione umana non garantisce affatto che ciò che *sembra* intelligente lo sia *realmente*. Entra dunque in gioco nell'itinerario proposto il problema della coscienza. L'autrice mantiene una posizione moderata che tuttavia non cela un velato scetticismo che potremmo definire realista. Interrogarsi sulla coscienza in questo contesto, sostiene, significherebbe giungere a una spiegazione di come possa emergere l'esperienza cosciente dalla materia. Il che sembra produrre esiti paradossali dal momento che la coscienza stessa dovrebbe essere in condizione di esplicitare la sua stessa genesi mettendosi in condizione di osservare se stessa contemporaneamente dall'interno e dall'esterno, prima e dopo se

stessa. Questo non esclude tuttavia la possibilità di considerare a riguardo alcune risposte più ragionevoli di altre. Ed è proprio ciò che si propone l'autrice ripercorrendo un dibattito pluridecennale in filosofia della mente che comprende l'analisi delle argomentazioni di autori quali David Chalmers, John Searle, Paul Churchland, Daniel Dennett, Aaron Sloman, Ruth Millikan e altri. L'originalità dell'analisi consiste nel mantenere sullo sfondo di queste riflessioni anche le modellizzazioni di quegli autori che attualmente utilizzano l'IA per indagare la coscienza in termini computazionali (*machine consciousness*) e di utilizzarle come validi termini di confronto e test di implementazione delle teorie. Un esempio particolarmente importante portato da Boden è il progetto LIDA (*Learning Intelligent Distribution Agent*) di Stan Franklin e Bernard Baars. Un modello concettuale ibrido (simbolico e PDP) a implementazione parziale basato su un'architettura che considera il cervello come un sistema distribuito nel quale diversi sottosistemi specializzati competono e si coordinano per l'accesso alla memoria di lavoro (*spazio di lavoro globale*) conscia. L'accesso a tale piano avviene in modo sequenziale, come per il flusso di coscienza, ma gli elementi che vi accedono generano un riverbero su tutte le altre aree della rete, che possono essere considerate "inconscie". Ciò avvia una concatenazione non casuale di agganci successivi tra gli elementi che si susseguono nello spazio di lavoro globale. I nuovi elementi vi accedono in quanto espressione delle risposte sufficientemente forti di una parte della rete inconscia che sono state a loro volta stimolate in risposta ai precedenti contenuti consci. Ne risulta una complessa dinamica ricorsiva che interessa ogni volta l'intero stato della rete. Sul piano dell'implementazione il modello manifesta proprietà di apprendimento e sembra fornire risposte plausibili in merito al problema del *frame* e alle modalità di collegamento tra stimoli provenienti da diverse aree cerebrali attribuite a una stessa entità (pp.122-127). Secondo Boden è importante sottolineare che questo modello non si sostituisce alle teorie elaborate nel contesto delle scienze umane, al contrario è sviluppato con l'obiettivo di includerne gli assunti al fine di arricchire le possibilità di indagine e implementazione.

In ultimo, nel settimo capitolo dal titolo "La Singolarità" (pp.143-166) l'autrice si concentra sulle conseguenze sociali e politiche

dell'impiego dell'IA e sui pericoli ad essa connessi. In via preliminare le sue riflessioni si focalizzano sullo sgomberare il campo dalle pseudo argomentazioni dei promotori della *singularità*, ovvero di coloro che ritengono vicina e potenzialmente distruttiva la realizzazione di una macchina dotata di AGI e addirittura dalle capacità infinitamente superiori a quelle umane. Boden ritiene che concentrare i nostri sforzi “gestionali” su questa remota possibilità risulti dannoso. Sebbene tali posizioni abbiamo aperto, attraverso il loro forte potere suggestivo, alla riflessione in merito a un uso responsabile di tali tecnologie, rischiano di farci perdere il contatto con i rischi già insiti nell'uso dei dispositivi oggi disponibili.