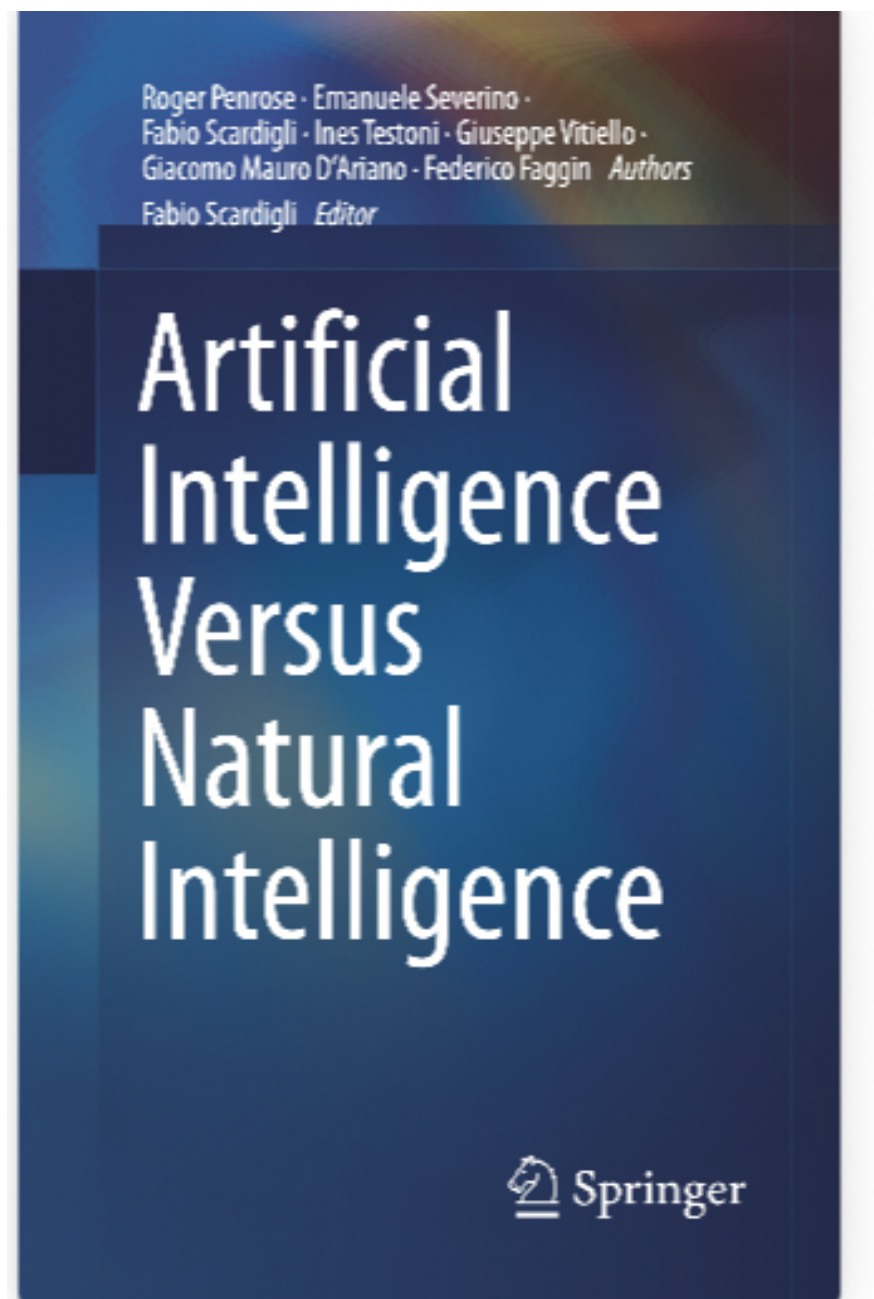


**L'intelligenza è davvero computazione? Coscienza, esperienza e mondo oltre
l'Intelligenza Artificiale**

Recensione di *Artificial Intelligence Versus Natural Intelligence*, Roger Penrose, Emanuele Severino, Fabio Scardigli, Ines Testoni, Giuseppe Vitiello, Giacomo Mauro D'Ariano, Federico Faggin. A cura di Fabio Scardigli. Springer, 2022.



1. Il testo del libro

Artificial Intelligence Versus Natural Intelligence è un libro particolare già nella sua genesi. Il volume nasce infatti dalla conferenza omonima tenutasi a Milano il 12 maggio 2018 e raccoglie il dialogo tra Roger Penrose ed Emanuele Severino, insieme ai contributi di Ines Testoni, Giuseppe Vitiello, Giacomo Mauro D'Ariano e Federico Faggin. Pubblicato da Springer nel 2022, è organizzato attorno a quattro nuclei principali: il confronto Penrose-Severino, il saggio di Testoni, quello di Vitiello e, infine, il contributo di D'Ariano e Faggin sul problema della coscienza e del libero arbitrio.

Il titolo potrebbe far pensare a un volume dedicato alla comparazione tra prestazioni umane e prestazioni delle macchine. Non è questo, tuttavia, il suo oggetto principale. Il vero interrogativo che attraversa le quasi duecento pagine è molto più radicale: *che cosa chiamiamo intelligenza quando parliamo di intelligenza artificiale?*

È possibile identificare l'intelligenza con la capacità di risolvere problemi, manipolare simboli, produrre linguaggio o raggiungere risultati? Oppure l'intelligenza propriamente detta presuppone qualcosa che la computazione, per quanto sofisticata, non possiede: comprensione, esperienza, coscienza, capacità di sentire, apertura a un mondo?

È proprio questo spostamento della domanda a rendere il volume interessante per la psicologia e la psicoterapia. L'AI diviene quasi un esperimento filosofico attraverso il quale interrogare nuovamente l'essere umano.

Il punto di partenza di Penrose è netto. Nel linguaggio ordinario, sostiene, l'intelligenza implica comprensione e la comprensione implica a sua volta una qualche forma di consapevolezza. Una macchina può essere enormemente superiore a un essere umano nella computazione senza che da ciò segua necessariamente che essa comprenda ciò che sta facendo.

Da qui nasce la tesi che accomuna, pur attraverso prospettive diversissime, larga parte degli autori del volume: *una teoria dell'intelligenza non può essere costruita lasciando irrisolto il problema della coscienza.*

2. Sintesi

Dalla prestazione alla comprensione

Penrose apre il confronto opponendosi alla versione forte del computazionalismo. Per chiarire la differenza fra calcolo e comprensione utilizza esempi provenienti dagli scacchi e soprattutto dalla logica matematica.

Il caso della posizione scacchistica proposta al programma Fritz è emblematico: il software possiede una capacità di calcolo immensamente superiore a quella del giocatore medio, ma non coglie immediatamente la configurazione globale della posizione che appare invece evidente a un essere umano. Per Penrose, il punto non è che il programma commetta un errore correggibile con un algoritmo migliore. Il problema è capire se alla prestazione computazionale corrisponda qualcosa che possa essere legittimamente definito *understanding*.

la coscienza come se fosse una cosa localizzabile tra le altre cose. La scienza, osserva, incontra ogni oggetto — cervello compreso — all'interno di quella che definisce la *manifestazione del mondo*. Cercare il luogo della coscienza all'interno del cervello rischia quindi di trasformare la condizione entro cui qualcosa può apparire in un oggetto tra gli oggetti che appaiono.

È una critica di enorme interesse anche al di fuori della filosofia severiniana. Prima di chiedersi se una macchina possa diventare cosciente, occorre infatti chiarire che cosa intendiamo per coscienza.

Ma Severino compie un passo ulteriore. Interroga il verbo che accompagna continuamente il dibattito sull'AI: produrre un'intelligenza, costruire una coscienza, creare una mente.

Dietro il concetto occidentale di *poiesis*, sostiene, si trova l'idea che qualcosa possa passare dal non essere all'essere. L'intero progetto di "produzione" di un'intelligenza artificiale presuppone quindi un'ontologia molto più antica dell'informatica.

Il paradosso cui Severino giunge è particolarmente efficace: l'Occidente ha già concepito l'uomo secondo categorie che lo rendono simile a una macchina, cioè come un ente che organizza mezzi in vista della realizzazione di scopi. La possibilità della macchina intelligente dipenderebbe allora, almeno in parte, dal fatto che *abbiamo precedentemente pensato l'uomo in modo meccanico*. È probabilmente uno dei passaggi filosoficamente più stimolanti del volume.

Testoni: tra coscienza, tecnica ed eternità

Ines Testoni introduce una prospettiva ancora differente, assumendo come punto di partenza il noto episodio dei chatbot sperimentati nel 2017 e utilizzandolo per interrogare il significato dell'autonomia linguistica delle macchine.

Il contributo procede poi verso questioni molto più ampie: il mind-brain problem, il solipsismo, il rapporto tra morte e coscienza, la crisi della metafisica e la concezione severiniana dell'eternità.

Particolarmente interessante, per una psicologia non riduzionistica, è il passaggio in cui Testoni mette in discussione la coincidenza fra mente e cervello richiamando prospettive embodied: la mente non viene descritta semplicemente come qualcosa che accade nella scatola cranica, ma come fenomeno costituito anche nelle relazioni percettive e operative tra corpi e ambiente.

Più speculativa è invece la possibilità, avanzata nel saggio, che sistemi artificiali possano anticipare forme di coscienza quantistica. Qui il testo oltrepassa nettamente ciò che può essere considerato oggi dimostrato. Ma proprio questa radicalità rende visibile una tensione che percorre l'intero volume: il problema non è stabilire soltanto *che cosa una macchina sappia fare*, ma se esista un "esserci" della macchina per il quale qualcosa possa apparire.

Vitiello: il cervello non è una "stella stupida"

Per un lettore proveniente dalla psicologia e dalle scienze della complessità, il saggio di Giuseppe Vitiello è probabilmente uno dei contributi più fecondi del libro.

Il cervello viene concepito come *sistema aperto*, strutturalmente inseparabile dallo scambio continuo con l'ambiente. La descrizione delle singole componenti cellulari e molecolari è necessaria ma non sufficiente perché ciò che deve essere spiegato è la dinamica complessiva del sistema cervello-ambiente.

Ancora più interessante è il modo in cui Vitiello descrive memoria e significato attraverso un *paesaggio di attrattori*.

Una nuova memoria non equivale semplicemente all'aggiunta di un nuovo contenuto in un archivio. L'acquisizione di qualcosa di nuovo modifica l'intero paesaggio: cambia le relazioni fra gli attrattori e quindi modifica retroattivamente anche il significato di ciò che era già presente.

Questo passaggio è fondamentale.

La memoria non appare più come recupero di un'informazione immutata, ma come *riorganizzazione dinamica di significati*. Il sistema può sostare maggiormente presso alcuni attrattori, transitare verso altri, modificare le proprie traiettorie e, in condizioni non patologicamente irrigidite, evitare di restare definitivamente intrappolato in una configurazione.

Il pensiero diviene così *errare*: movimento attraverso un paesaggio continuamente ridisegnato.

Ed è proprio l'errore, in un capovolgimento molto suggestivo, a diventare la caratteristica dell'intelligenza. La macchina ordinaria viene costruita affinché sia prevedibile; il cervello, al contrario, possiede la capacità di generare traiettorie che non appartengono a una catena rigidamente predeterminata. Da qui l'immagine della *mistake-making machine*, una macchina capace di sbagliare non perché guasta, ma perché aperta alla novità.

In questa prospettiva l'errore non rappresenta semplicemente un deficit rispetto alla risposta corretta: può costituire l'emergere di una possibilità che non era contenuta nelle alternative precedentemente disponibili.

D'Ariano e Faggin¹: il punto cieco è l'esperienza

L'ultimo contributo, di Giacomo Mauro D'Ariano e Federico Faggin, affronta direttamente quello che Chalmers ha definito *hard problem of consciousness*.

Il punto di partenza è molto chiaro: descrivere l'elaborazione dell'informazione non equivale ancora a spiegare perché quell'elaborazione sia accompagnata da una esperienza.

Gli autori propongono una teoria non riduzionistica nella quale l'informazione connessa alla coscienza possiede una natura quantistica e arrivano a una forma di panpsichismo basato sull'informazione quantistica. Nella loro costruzione coscienza, qualia e libero arbitrio vengono integrati all'interno di un unico modello teorico.

La proposta è certamente la parte più tecnicamente impegnativa e, insieme, una delle più speculative del volume. È però rilevante il problema da cui prende le mosse: una teoria espressa interamente dal punto di vista di un osservatore esterno sembra lasciare fuori proprio ciò che dovrebbe spiegare, vale a dire **il fatto che per qualcuno qualcosa venga sentito**.

¹ Ricordo che è stato l'inventore dei microchips (n.d.r.)

La distinzione fra stato osservabile dall'esterno e stato vissuto internamente diviene così decisiva. Nel modello degli autori la soggettività non è un sottoprodotto inerte, ma assume addirittura efficacia causale attraverso il rapporto fra coscienza, agency e libero arbitrio.

3. Il valore del volume per il modello cognitivo-neuropsicologico

È proprio qui che il libro acquista un interesse particolare per una prospettiva cognitivo-neuropsicologica di matrice fenomenologico-ermeneutica. Non perché offra una teoria psicologica già pronta. Non la offre. E neppure perché sia necessario assumere le ipotesi quantistiche di Penrose, Vitiello o D'Ariano e Faggin per sostenere una visione non riduzionistica dell'esperienza umana. Sarebbe anzi un errore trasformare la meccanica quantistica in una scorciatoia esplicativa per qualunque fenomeno che il paradigma computazionale non riesce a spiegare. Il contributo più importante è precedente. Il volume costringe a separare concetti che una certa concezione dell'AI tende facilmente a sovrapporre:

Informazione, computazione, comportamento, comprensione, esperienza e coscienza non sono sinonimi.

Da una prospettiva vicina a quella sviluppata da Davide Liccione, la questione assume una forma ancora più radicale. L'essere umano non è innanzitutto un elaboratore di informazioni che riceve un input dal mondo, costruisce internamente una rappresentazione e produce un output comportamentale. È un essere che *si trova già situato in un mondo*, corporalmente ed emotivamente coinvolto in ciò che incontra e storicamente trasformato dagli eventi che vive.

Da questo punto di vista, la difficoltà dell'AI non consiste semplicemente nel fatto che le manchi una quantità sufficiente di informazione o potenza computazionale. Il problema è che la prestazione linguistica e l'esperienza vissuta appartengono a due livelli differenti.

Un sistema artificiale può produrre un enunciato del tipo «ho paura di perderti». Ma la produzione linguisticamente appropriata dell'enunciato non dimostra l'esistenza di quel particolare modo di trovarsi nel mondo nel quale qualcuno teme realmente la perdita di una persona significativa. È precisamente questa distinzione a diventare cruciale anche in psicoterapia. Il paziente non è il testo che produce. Il racconto è una configurazione dell'esperienza, non la sua duplicazione linguistica. E comprendere un paziente non significa semplicemente classificare semanticamente le sue proposizioni, ma cogliere come esse si inscrivano in una storia, in un corpo, in relazioni significative, in possibilità d'azione e in un determinato modo di aprirsi il mondo.

Da questo punto di vista, il contributo di Vitiello entra in una risonanza particolarmente interessante con una concezione dinamica del Sé.

L'idea che l'acquisizione di una nuova memoria produca la *riorganizzazione dell'intero paesaggio degli attrattori*, invece di aggiungere semplicemente un contenuto a quelli precedenti, offre una metafora sistemica potente anche per comprendere il cambiamento psicoterapeutico.

Una trasformazione terapeutica significativa non consiste infatti soltanto nell'aggiungere una nuova cognizione positiva a una cognizione negativa. Modifica le relazioni fra differenti modi di essere.

Quando cambia il paesaggio, può cambiare anche il significato del passato.

Una precedente esperienza di fallimento non viene cancellata, ma può smettere di organizzare nello stesso modo l'esperienza presente. Una nuova possibilità progettuale può modificare retroattivamente il significato di esperienze biografiche precedenti. Un modo di essere precedentemente dominante può perdere forza attrattiva mentre configurazioni prima poco accessibili diventano maggiormente percorribili. Da qui emerge un secondo punto di incontro particolarmente interessante: *la salute non coincide necessariamente con la stabilità*. Nel modello descritto da Vitiello il cervello sano attraversa il paesaggio degli attrattori senza restare permanentemente intrappolato in uno di essi.

Trasposta sul piano psicopatologico — e qui siamo naturalmente oltre ciò che Vitiello sostiene direttamente nel capitolo — questa idea permette di pensare la sofferenza non soltanto come presenza di contenuti patologici, ma come *riduzione della libertà delle traiettorie possibili del sistema*.

Un Sé che ritorna continuamente alla medesima configurazione non è problematico perché possiede quell'attrattore, ma perché non riesce più sufficientemente a uscirne.

Infine, il libro permette di precisare il possibile ruolo dell'intelligenza artificiale nella psicoterapia.

Se la distinzione fra prestazione e comprensione viene mantenuta, non occorre scegliere tra due estremi: considerare l'AI capace di sostituire l'esperienza umana oppure considerarla clinicamente irrilevante. Un sistema artificiale può analizzare testi, riconoscere ricorrenze, quantificare transizioni, individuare cambiamenti longitudinali, confrontare configurazioni linguistiche e rendere visibili regolarità che il terapeuta difficilmente potrebbe osservare su centinaia di migliaia di parole.

Ma l'oggetto ultimo della psicoterapia continua a essere *un'esistenza che vive un mondo*, non il corpus linguistico prodotto da quella esistenza.

L'AI può dunque diventare uno straordinario strumento di osservazione del processo senza che da ciò derivi l'equivalenza fra l'analisi del processo e l'esperienza che quel processo esprime.

4. Conclusioni

Artificial Intelligence Versus Natural Intelligence non è un libro sull'Intelligenza Artificiale nel senso oggi più comune del termine. Chi cercasse una discussione tecnica su reti neurali, modelli linguistici o architetture generative troverebbe poco di tutto questo. Ed è proprio per questo che il libro conserva interesse.

Gli esempi tecnologici appartengono inevitabilmente al momento storico in cui il confronto originario ebbe luogo, nel 2018. Ma l'interrogativo posto dagli autori è diventato, semmai, ancora più urgente con la comparsa di sistemi capaci di produrre linguaggio, immagini e ragionamenti apparentemente sempre più simili a quelli umani.

Più una macchina diventa convincente sul piano della prestazione, più diventa necessario evitare l'errore di dedurre dalla somiglianza della prestazione l'identità del fenomeno che la produce.

Il merito principale del volume consiste allora nel riportare la discussione là dove dovrebbe probabilmente essere iniziata: *prima di stabilire se una macchina possa essere intelligente, dobbiamo decidere che cosa intendiamo quando diciamo che un essere umano lo è*.

Penrose ci obbliga a distinguere computazione e comprensione. Severino domanda quale ontologia sia già implicita nell'idea stessa di “costruire” un'intelligenza. Testoni riapre il problema del rapporto fra coscienza, corpo e tecnica. Vitiello propone un cervello aperto, dinamico, dissipativo, capace di attraversare un paesaggio di attrattori continuamente riconfigurato. D'Ariano e Faggin riportano al centro ciò che ogni descrizione puramente funzionale rischia di lasciare fuori: l'esperienza in prima persona.

Non tutte queste risposte hanno lo stesso grado di solidità empirica e alcune delle ipotesi quantistiche avanzate nel volume devono essere lette per ciò che sono: *proposte teoriche forti e in parte speculative*, non risultati conclusivi. Ma forse il valore del libro non consiste nell'aver risolto il problema della coscienza. Consiste nell'impedire che il successo dell'Intelligenza Artificiale ci induca a considerarlo già risolto.

Per la psicologia e per la psicoterapia questa distinzione è fondamentale. Una macchina può riconoscere configurazioni dell'esperienza, descriverle, confrontarle e forse aiutarci a comprenderne meglio alcune dinamiche. Ma rimane aperta la domanda decisiva posta, in forme differenti, da tutto il volume:

che differenza esiste tra produrre il linguaggio di un'esperienza e avere un'esperienza da raccontare?

È precisamente nello spazio aperto da questa domanda che il confronto tra Intelligenza Artificiale e psicoterapia diventa scientificamente e filosoficamente interessante.