

## INTRODUZIONE

---

Nel settembre 1962, durante la frenesia globale della corsa allo spazio, della crisi dei missili di Cuba e del vaccino antipolio recentemente migliorato, è stato raggiunto un traguardo, meno osannato, ma forse altrettanto decisivo, nella storia umana: è stato nell'autunno del 1962 che abbiamo predetto il futuro.

Sugli schermi dei televisori degli Stati Uniti, da poco a colori, ha debuttato *The Jetsons* (*I pronipoti*, nell'edizione italiana), una serie di animazione di cui è protagonista una famiglia, cento anni nel futuro. Sotto la forma di una sitcom, la serie era, in effetti, una previsione di come avrebbero vissuto in futuro gli esseri umani, di quali tecnologie avrebbero riempito le loro tasche e arredato le loro case.

*I pronipoti* ha previsto correttamente le videochiamate, i televisori a schermo piatto, i telefoni cellulari, la stampa 3D e gli smarwatch; tutte tecnologie che apparivano incredibili nel 1962 eppure onnipresenti entro il 2022. Tuttavia, c'è una tecnologia che non siamo affatto riusciti a creare, un elemento futuristico di cui ancora non possiamo fruire: la robot autonoma di nome Rosey.

Rosey è un'assistente per tutta la famiglia Jetson: si prende cura dei bambini e della casa. Quando Elroy (sei anni) ha difficoltà a scuola, è Rosey che lo aiuta a fare i compiti. Quando Judy, la figlia quindicenne, ha bisogno di una mano per imparare a guidare, è Rosey che le dà lezioni di guida. Rosey prepara i pasti, apparecchia la tavola, lava i piatti. Rosey è leale, sensibile, e ha sempre pronta una battuta. Individua sul nascere i litigi e i fraintendimenti fra i componenti della famiglia e interviene per aiutarli a vedere la prospettiva degli altri. Una volta, si commuove fino alle lacrime per una poesia che Elroy ha scritto per la madre. La stessa Rosey, in un episodio, si innamora.

In altre parole, Rosey aveva l'intelligenza di un essere umano. Non solo le capacità di ragionamento, di buon senso e motorie necessarie per svolgere compiti complessi nel mondo fisico, ma anche quelle di

empatia, di assunzione di prospettiva e di sensibilità sociale necessarie per muoversi nel nostro mondo sociale. Con le parole di Jane Jetson, Rosey era “proprio come una della famiglia”<sup>1</sup>.

Anche se i *pronipoti* hanno previsto correttamente telefoni cellulari e smartwatch, non abbiamo ancora niente di simile a Rosey. Mentre questo libro va in stampa, anche i comportamenti più elementari di Rosey sono ancora fuori della nostra portata. Non è un segreto che la prima azienda che riesca a costruire un robot che sappia semplicemente *caricare una lavastoviglie* avrà immediatamente un prodotto di grande successo: tutti i tentativi fatti in questo senso sono falliti. Non è fondamentalemente un problema *meccanico*, ma *intellettuale* – la capacità di identificare gli oggetti nel lavello, di afferrarli nel modo giusto e di infilarli nella lavastoviglie senza rompere nulla si è dimostrata di gran lunga più difficile da riprodurre di quanto si fosse pensato.

Ovviamente, anche se non abbiamo ancora Rosey, i passi avanti nel campo dell’intelligenza artificiale (AI) dopo il 1962 sono stati notevoli. L’AI ora può battere i migliori al mondo, tra gli esseri umani, in numerosi giochi di abilità, fra cui scacchi e Go. L’AI può riconoscere i tumori nelle immagini radiologiche con la stessa abilità di una radiologa umana. L’AI è sulla soglia delle auto a guida autonoma e, proprio negli ultimissimi anni, nuovi passi avanti nei *Large Language Model* (LLM, modelli linguistici di grandi dimensioni) stanno rendendo possibile a prodotti come ChatGPT, lanciato nell’autunno del 2022, comporre poesie, tradurre tra lingue diverse e addirittura scrivere programmi. Con grande scorno di ogni insegnante di scuola superiore sul pianeta Terra, ChatGPT è in grado di comporre istantaneamente un saggio originale, scritto notevolmente bene, su quasi qualsiasi argomento le possa proporre un intrepido studente. ChatGPT può addirittura superare l’esame per l’abilitazione a esercitare la professione di avvocato, ottenendo un punteggio migliore del 90 per cento degli avvocati.

Parallelamente a questa lunga serie di traguardi raggiunti dall’AI, è sempre stato difficile stabilire quanto ci stessimo avvicinando alla creazione di un’intelligenza di livello umano. Dopo i primi successi degli algoritmi per la risoluzione di problemi negli anni Sessanta, Marvin Minsky, uno dei pionieri dell’AI, ha proclamato che “fra da tre a otto anni avremo una macchina con l’intelligenza di un essere umano medio”. Non è andata così. Dopo i successi dei sistemi esperti negli anni Ottanta, *BusinessWeek* ha proclamato: “L’AI è qui”. Poco dopo c’è stato uno stallo. Ora, con i passi avanti nei modelli LLM, molti ricercatori hanno proclamato di nuovo che “la partita è finita”, perché siamo “sul punto di raggiungere una AI di livello umano”<sup>2</sup>. Che cosa vuol dire: siamo finalmente sul punto di creare un’intelligenza artificiale simile a quella umana come Rosey, oppure i modelli linguistici di

grandi dimensioni come ChatGPT sono solo il traguardo più recente in un lungo viaggio che continuerà ancora per i decenni a venire?

Lungo questo percorso, mentre l'AI continua a diventare più intelligente, diventa sempre più difficile misurare il nostro stato di avanzamento verso questo obiettivo. Se un sistema AI fa meglio degli esseri umani in un compito, questo significa che il sistema ha colto come gli esseri umani risolvono quel compito? Una calcolatrice, che è in grado di macinare numeri più rapidamente di un essere umano, comprende realmente la matematica? ChatGPT, che fa meglio della maggior parte degli avvocati negli esami di abilitazione, comprende realmente la giurisprudenza? Come possiamo stabilire la differenza e in quali circostanze, eventualmente, la differenza conta davvero?

Nel 2021, oltre un anno prima del rilascio di ChatGPT (il chatbot che sta rapidamente diffondendosi in ogni angolo della società), stavo utilizzando il suo precursore, un modello LLM denominato GPT-3. GPT-3 era stato addestrato su grandi quantità di testo (dove grandi significa *tutta Internet*), e poi utilizzava quel corpus per cercare di individuare per corrispondenza di configurazione la risposta più probabile a un prompt. Quando le si chiedeva: "Quali sono due motivi per cui un cane potrebbe essere di cattivo umore?", rispondeva: "Due motivi per cui un cane potrebbe essere di cattivo umore sono se ha fame o se è in calore". Qualcosa nella nuova architettura di questi sistemi consentiva loro di dare risposte con quello che sembrava almeno un grado notevole di intelligenza. Questi modelli erano in grado di generalizzare fatti di cui avevano letto (come le pagine di Wikipedia sui cani e altre pagine sulle cause del cattivo umore) a nuove domande che non avevano mai visto. Nel 2021, esploravo le possibili applicazioni di questi nuovi modelli linguistici: sarebbe stato possibile utilizzarli per fornire nuovi sistemi di sostegno per la salute mentale, o un'assistenza più completa ai clienti, o un accesso più democratizzato alle informazioni mediche.

Quanto più interagivo con GPT-3, tanto più ero affascinato dai suoi successi e dai suoi errori. Per certi versi era brillante, per altri stranamente stupido. Chiedete a GPT-3 di scrivere un saggio sulla coltivazione delle patate nel diciottesimo secolo e il suo rapporto con la globalizzazione, e otterrete un saggio sorprendentemente coerente. Ponetegli una domanda di buon senso su quello che si potrebbe vedere in un seminterrato, e vi risponderà in modo insensato\*. Perché

---

\* Ho chiesto a GPT-3 di completare questa frase "I am in my windowless basement, and I look toward the sky, and I see..." [Sono nel mio seminterrato senza finestre e guardo verso il cielo, e vedo...], GPT-3 ha risposto "a light, and I know it is a star, and I am happy" [una luce, e so che è una stella, e sono felice]. In realtà se in un seminterrato

GPT-3 riusciva a rispondere correttamente ad alcune domande e non ad altre? Quali caratteristiche dell'intelligenza umana cattura, e quali gli sfuggono? E perché, mentre lo sviluppo dell'AI continua ad accelerare, certe domande, a cui era difficile avere risposta un anno, diventano facili negli anni seguenti? In effetti, mentre questo libro va in stampa, la nuova versione aggiornata di GPT-3, denominata GPT-4, rilasciata agli inizi del 2023, può rispondere correttamente a molte domande che mettevano in difficoltà GPT-3. E tuttavia, come vedremo in questo libro, GPT-4 non riesce a catturare caratteristiche essenziali dell'intelligenza umana, qualcosa che succede nel cervello umano.

In effetti, le discrepanze fra intelligenza artificiale e intelligenza umana lasciano profondamente perplessi. Perché l'AI può battere qualsiasi essere umano sulla terra in una partita a scacchi ma non riesce a caricare una lavastoviglie meglio di un bambino di sei anni?

Facciamo fatica a rispondere a queste domande perché non comprendiamo ancora la cosa che stiamo cercando di ricreare. Tutte queste domande sono, essenzialmente, non sull'AI, ma sulla natura dell'intelligenza umana stessa: come funziona, perché funziona in quel modo e, come vedremo tra breve, quella che è la cosa più importante, come è arrivata a essere quello che è.

### *Il suggerimento della natura*

Quando abbiamo voluto comprendere il volo, abbiamo tratto la nostra prima ispirazione dagli uccelli; quando George de Mestral ha inventato il Velcro, ha ricavato l'idea dai frutti della bardana maggiore; quando Benjamin Franklin si è messo a studiare l'elettricità, la prima scintilla di comprensione gli è giunta dal fulmine. La natura, lungo tutta la storia delle innovazioni umane, è stata sempre una guida meravigliosa.

La natura ci offre anche degli indizi su come funziona l'intelligenza: la cui sede più ovvia è, ovviamente, il cervello umano. In questo modo, però, l'AI è diversa da quelle altre innovazioni tecnologiche; il cervello ha dimostrato di essere più impervio da decifrare di quanto non lo fossero le ali o il fulmine. Le scienze hanno studiato per millenni come funziona il cervello: abbiamo fatto passi avanti, ma non abbiamo ancora risposte soddisfacenti.

Il problema è la complessità.

---

guardi verso l'alto non vedrai le stelle, ma solo il soffitto. Nuovi modelli come GPT-4, rilasciato nel 2023, rispondono con maggiore accuratezza a domande di buon senso come questa. Ne parleremo meglio nel Capitolo 22.

Il cervello umano contiene 86 miliardi di neuroni e oltre centomila miliardi di connessioni. Ciascuna di quelle connessioni è così minuscola (è larga meno di trenta nanometri) che risulta a malapena visibile anche con i più potenti fra i microscopi. Queste connessioni sono affastellate insieme in un gran garbuglio: in un solo millimetro cubo (la larghezza di una singola lettera su una moneta da un centesimo) ci sono oltre *un miliardo* di connessioni<sup>3</sup>.

Il numero delle connessioni in sé è solo un aspetto di ciò che rende complesso il cervello; anche se riportassimo su una carta il cablaggio di ogni neurone, saremmo ben lontani dal capire come funziona il cervello. A differenza delle connessioni elettriche nel vostro computer, dove tutti i file comunicano utilizzando lo stesso segnale (elettroni), attraverso ciascuna di queste connessioni neurali transitano centinaia di sostanze chimiche diverse, ciascuna con effetti totalmente diversi dalle altre. Il semplice fatto che due neuroni siano connessi ci dice ben poco di che cosa comunicano; il peggio è che quelle stesse connessioni sono in una condizione di costante cambiamento, con alcuni neuroni che si diramano e formano nuove connessioni, mentre altri si ritirano ed eliminano le vecchie. Nel complesso, questo fa della retroingegnerizzazione del funzionamento del cervello un compito assurdo.

Studiare il cervello è affascinante e al tempo stesso frustrante. Un paio di centimetri dietro i vostri occhi si trova la meraviglia più incredibile dell'universo. Ospita i segreti per comprendere la natura dell'intelligenza, per costruire un'intelligenza artificiale simile a quella umana, per capire perché noi esseri umani pensiamo e ci comportiamo in questo modo. È proprio lì, ricostruito milioni di volte all'anno, a ogni nuovo essere umano che nasce. Possiamo toccarlo, tenerlo in mano, sezionarlo, siamo *letteralmente fatti del cervello*, e tuttavia i suoi segreti rimangono fuori dalla nostra portata, nascosti in bella vista.

Se vogliamo retroingegnerizzare come funziona il cervello, se vogliamo costruire Rosey, se vogliamo scoprire la natura nascosta dell'intelligenza umana, forse il cervello umano non è il migliore degli indizi che può offrire la natura. Il posto in cui è più intuitivo andare a guardare per comprendere il cervello umano è, naturalmente, dentro il cervello umano stesso, ma, controintuitivamente, questo può essere *l'ultimo* posto in cui guardare. Il posto migliore in cui iniziare a guardare può essere nei fossili polverosi sepolti nella crosta terrestre, in geni microscopici all'interno delle cellule di tutto il regno animale, e nel cervello dei molti *altri* animali che popolano il nostro pianeta.

In altre parole, può darsi che la risposta non sia nel presente, ma nei resti nascosti di un lontano passato.

## *Il museo (che non c'è) del cervello*

Sono sempre stato convinto che l'unico modo perché l'intelligenza artificiale funzioni è che effettui la computazione in modo simile al cervello umano<sup>4</sup>.

– Geoffrey Hinton (professore all'Università di Toronto, considerato uno dei “padrini dell'AI”)

Gli esseri umani fanno volare navi spaziali, dividono gli atomi e modificano i geni. Nessun altro animale ha mai inventato la ruota.

Dato il lungo elenco delle invenzioni umane, si potrebbe pensare che ci sia poco da imparare dal cervello di altri animali. Si potrebbe pensare che il cervello umano sia del tutto peculiare e non sia affatto simile al cervello di altri animali, che qualche struttura speciale del cervello sia il segreto della nostra intelligenza. Ma questo non è quello che vediamo.

Ciò che colpisce in modo particolare quando si esaminano i cervelli di altri animali è quanto siano *simili* al nostro. La differenza fra il nostro cervello e quello di uno scimpanzé è praticamente nulla. La differenza fra il nostro cervello e quello di un topo è solo una manciata di variazioni. Il cervello di un pesce ha quasi tutte le stesse strutture del nostro.

Queste somiglianze fra i cervelli in tutto il regno animale significano qualcosa di importante. Sono indizi. Indizi sulla natura dell'intelligenza. Indizi su di noi. Indizi sul nostro passato.

Anche se i cervelli di oggi sono complessi, non lo sono sempre stati. Il cervello è emerso da processi evolutivi caotici senza pensiero; piccole variazioni casuali nei tratti sono state selezionate o scartate a seconda del fatto che supportassero l'ulteriore riproduzione della vita.

Nell'evoluzione, i sistemi all'inizio sono semplici, e la complessità compare solo con il tempo<sup>5</sup>. Il primo cervello, il primo insieme di neuroni nella testa di un animale, è comparso seicento milioni di anni fa in un verme che aveva le dimensioni di un granello di riso. Quel verme è stato l'antenato di tutti gli animali moderni dotati di cervello. Dopo centinaia di milioni di anni di manipolazioni evolutive, attraverso migliaia di miliardi di piccoli cambiamenti nel cablaggio, quel cervello semplice si è trasformato nel grande repertorio dei cervelli moderni. Un ramo dei discendenti di questo antico verme ha portato al cervello nella nostra testa.

\* Anche se i sistemi non diventano *necessariamente* più complessi, la possibilità della complessità aumenta con il tempo.

Se solo potessimo tornare indietro nel tempo ed esaminare quel primo cervello per capire come funzionava e quali trucchi rendeva possibili... Se solo potessimo poi seguire la crescita progressiva della complessità nella linea dei suoi discendenti che ha portato al cervello umano, osservando ogni modifica fisica che si è verificata e le capacità intellettuali che ha reso possibili... Se solo potessimo farlo, potremmo comprendere la complessità che alla fine ne è emersa. In effetti, come ha detto il biologo Theodosius Dobzhansky, “Niente nella biologia ha senso, se non alla luce dell’evoluzione”.

Anche Darwin aveva sognato di ricostruire una storia simile: conclude il suo *L'origine delle specie* sognando un futuro in cui “la psicologia sarà basata su nuove fondamenta, quelle della necessaria acquisizione di ogni facoltà e capacità mentale per gradi”. 150 anni dopo Darwin, forse è finalmente possibile.

Non abbiamo macchine del tempo, ma, almeno in linea di principio, possiamo intraprendere qualche viaggio nel tempo. Solo nell’ultimo decennio, le neuroscienze evolutive hanno compiuto progressi incredibili nella ricostruzione del cervello dei nostri antenati. Lo si può fare attraverso la documentazione fossile: gli scienziati possono utilizzare i crani fossili di creature antiche per retroingegnerizzare la struttura dei loro cervelli. Un altro modo per ricostruire i cervelli dei nostri antenati è attraverso l’esame del cervello di altre specie nel regno animale.

Il motivo per cui i cervelli in tutto il regno animale sono così simili è che derivano tutti da radici comuni, gli antenati condivisi. Ogni cervello nel regno animale è un piccolo indizio in merito a come erano fatti i cervelli dei nostri antenati; ogni cervello non è solo una macchina, ma una capsula del tempo piena di indizi nascosti delle migliaia di miliardi di menti venute prima di noi. Esaminando i traguardi intellettuali che questi altri animali condividono e quelli che non condividono, possiamo iniziare non solo a ricostruire il cervello dei nostri antenati, ma anche a determinare quali capacità intellettuali rendeva loro possibili. Insieme, possiamo iniziare a tracciare l’acquisizione di ogni facoltà mentale per gradi.

È, ovviamente, un lavoro ancora in corso, ma la storia sta diventando straordinariamente chiara.

### *Il mito degli strati*

Non sono certo il primo a proporre una prospettiva evolutiva per comprendere il cervello umano: esiste una lunga tradizione in questo senso. Il modello più famoso fu formulato negli anni Sessanta del se-

colo scorso dal neuroscienziato Paul MacLean, il quale ipotizzò che il cervello umano fosse costituito da tre strati (perciò *trino*), uno sopra l'altro: la *neocorteccia*, evoluta più recentemente, sopra il *sistema limbico*, evolutosi in precedenza e a sua volta posto sopra il *cervello rettiliano*, il primo a evolversi.

MacLean sosteneva che il cervello rettiliano era il centro dei nostri istinti basilari di sopravvivenza, come l'aggressività e la territorialità; il sistema limbico doveva essere il centro delle emozioni, come paura, attaccamento parentale, desiderio sessuale e fame. La neocorteccia poi doveva essere il centro della cognizione, a cui dobbiamo linguaggio, astrazione, capacità di pianificazione e percezione. Il modello di MacLean immaginava che i rettili avessero *solo* un cervello rettiliano, i mammiferi come topi e conigli un cervello rettiliano *e* un sistema limbico, e solo gli esseri umani possedessero tutti i tre sistemi. In effetti, secondo MacLean, queste "tre formazioni evolutive si possono immaginare come tre computer biologici interconnessi, ciascuno dei quali con la propria intelligenza speciale, la propria soggettività, il proprio senso di tempo e spazio, le proprie funzioni di memoria, motorie e di altro tipo"<sup>5</sup>.

Il problema è che l'ipotesi del cervello trino di MacLean è stata ampiamente screditata, non perché sia inesatta (tutti i modelli sono inesatti), ma perché porta alle conclusioni sbagliate sull'evoluzione e il funzionamento del cervello<sup>6</sup>. L'anatomia cerebrale che implica è errata: il cervello dei rettili non è fatto solo delle strutture che MacLean definiva "cervello rettiliano"; i rettili hanno anche la propria versione di un sistema limbico. La suddivisione funzionale si è dimostrata sbagliata: *istinti di sopravvivenza*, *emozioni* e *cognizione* non si distinguono chiaramente, ma emergono da reti di sistemi differenti che si estendono su tutti i tre supposti strati. Anche la storia evolutiva che il modello implicava si è rivelata errata. Non abbiamo un cervello rettiliano nella nostra testa; l'evoluzione non ha semplicemente collocato a strati un sistema sopra l'altro, senza apportare alcuna modifica ai sistemi preesistenti.

Anche se il cervello trino di MacLean si fosse dimostrato più vicino alla verità, il problema maggiore è che le divisioni funzionali non sono particolarmente utili per i nostri scopi. Se il nostro obiettivo è retroingegnerizzare il cervello umano per capire la natura dell'intelligenza, i tre sistemi di MacLean sono troppo ampi, e le funzioni che sono loro attribuite sono troppo vaghe, per darci anche solo un punto da cui partire.



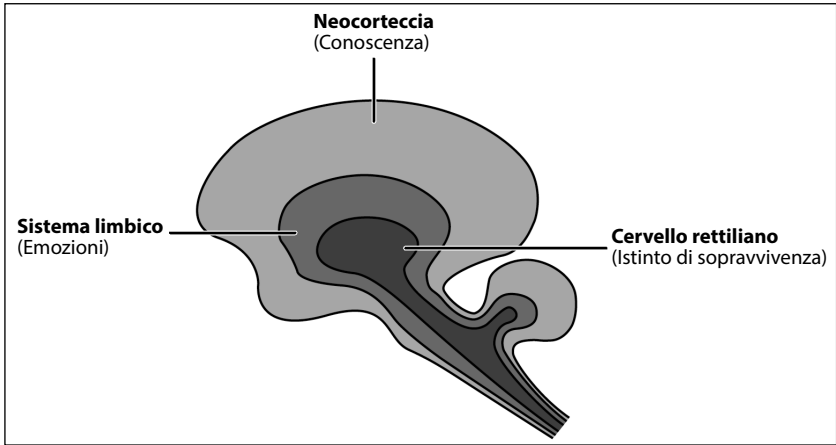


Figura I.1 *Il cervello trino di MacLean.*

Abbiamo bisogno di fondare la nostra comprensione del funzionamento e dell'evoluzione del cervello su ciò che sappiamo del funzionamento dell'intelligenza, e per questo dobbiamo guardare al campo dell'intelligenza artificiale. Il rapporto fra AI e cervello è bidirezionale; certo il cervello può dirci molto su come creare un'intelligenza artificiale di livello umano, ma anche l'AI può insegnarci qualcosa sul cervello. Se pensiamo che una certa parte del cervello usi qualche algoritmo specifico, ma quell'algoritmo non funziona quando lo implementiamo in una macchina, questo ci fornisce una prova che il cervello probabilmente non funziona in quel modo. Viceversa, se troviamo un algoritmo che funziona bene nei sistemi AI e troviamo dei paralleli fra le proprietà di quell'algoritmo e le proprietà dei cervelli animali, questo ci dà qualche indicazione che il cervello potrebbe effettivamente funzionare in quel modo.

Il fisico Richard Feynman, poco prima di morire, ha lasciato scritto su una lavagna: "Quello che non posso creare, non lo capisco". Il cervello è la nostra ispirazione guida per come costruire un'AI, e l'AI è la nostra cartina di tornasole per il livello di comprensione del cervello.

Abbiamo bisogno di una nuova storia evolutiva del cervello, fondata non solo su una comprensione moderna di come l'anatomia del cervello si sia evoluta nel tempo, ma anche su una comprensione moderna dell'intelligenza stessa.

## *Le cinque svolte*

Iniziamo con l'intelligenza artificiale a livello di topo (ARI), poi passiamo all'intelligenza artificiale a livello di gatto (ACI) e così arriveremo all'intelligenza artificiale a livello umano (AHI).<sup>7</sup>

– Yann LeCun, responsabile per l'AI di Meta

Abbiamo una lunga storia evolutiva da percorrere: quattro miliardi di anni. Invece di tentare di fare il resoconto di ogni piccola variazione, ci concentreremo sulle principali svolte dell'evoluzione. In effetti, come approssimazione iniziale (una prima scaletta di questa storia) tutta l'evoluzione del cervello umano si può riassumere ragionevolmente come il punto d'arrivo di solo *cinque* grandi svolte, a partire dai primissimi cervelli per arrivare al cervello umano.

Queste cinque svolte costituiscono la mappa organizzativa del libro, e disegnano l'itinerario per la nostra avventura a ritroso nel tempo. Ogni svolta è emersa da un nuovo insieme di modificazioni del cervello e ha dotato gli animali di un nuovo “portafoglio” di abilità intellettuali. Il libro è diviso quindi in cinque parti, una per ciascuna svolta. In ogni sezione, descriverò perché si siano evolute quelle abilità, come funzionavano e come si manifestino ancora nel cervello umano di oggi.

Ogni svolta si è basata sulle precedenti ed è stata la base per le successive. Le innovazioni del passato hanno reso possibili le innovazioni del futuro. Grazie a questo insieme ordinato di modificazioni, la storia evolutiva del cervello ci aiuta a dare un senso alla complessità che alla fine ne è emersa.

Questa storia non può essere raccontata fedelmente considerando solo la biologia dei cervelli dei nostri antenati. Le grandi svolte sono sempre emerse da periodi in cui i nostri antenati hanno dovuto affrontare situazioni estreme o sono stati coinvolti in potenti anelli di retroazione, pressioni che hanno portato a riconfigurazioni rapide del cervello. Non possiamo comprendere le svolte nell'evoluzione del cervello senza anche comprendere le prove e i trionfi dei nostri antenati: i predatori a cui hanno saputo sfuggire, le calamità ambientali da cui sono passati indenni e le nicchie in cui si sono rifugiati dalla disperazione per poter sopravvivere.

Cosa fondamentale, ci baseremo su quello che abbiamo compreso oggi nel campo dell'AI, perché molte delle svolte nell'intelligenza biologica hanno un parallelo in quello che abbiamo appreso nell'intelligenza artificiale. Alcune di quelle svolte rappresentano “trucchi”

intellettuali che comprendiamo bene nell'AI, mentre altri trucchi sono ancora al di là della nostra capacità di comprensione. In questo modo, forse la storia evolutiva del cervello può gettare luce su quali svolte ci siano sfuggite nello sviluppo di un'intelligenza artificiale di livello umano. Forse ci rivelerà qualcuno degli indizi nascosti della natura.

*Io*

Vorrei potervi dire che ho scritto questo libro perché ho passato tutta la vita a riflettere sull'evoluzione del cervello e a cercare di costruire robot intelligenti, ma non sono né un neuroscienziato, né un esperto di robotica e nemmeno uno scienziato: l'ho scritto perché volevo leggere questo libro.

Sono arrivato a cogliere la sconcertante discrepanza fra intelligenza umana e artificiale cercando di applicare sistemi di AI a problemi concreti. Ho passato la maggior parte della mia carriera in un'azienda, di cui sono stato fra i fondatori, che si chiama Bluecore; abbiamo realizzato software e sistemi AI per aiutare alcune delle aziende più grandi del mondo a personalizzare il proprio marketing. Il nostro software aiutava a prevedere che cosa i consumatori avrebbero acquistato, ancora prima che sapessero che cosa volevano. Eravamo solo una parte minuscola in un mare di un'infinità di aziende che iniziavano a utilizzare i nuovi passi avanti nei sistemi AI, ma tutti questi molti progetti, piccoli e grandi, si scontravano sempre con le stesse domande sconcertanti.

Quando si commercializzano sistemi AI, alla fine si svolge una serie di riunioni fra team di business e team dell'apprendimento automatico. I team di business cercano applicazioni di nuovi sistemi AI che siano *di valore*, mentre solo i team dell'apprendimento automatico sanno quali applicazioni sarebbero *fattibili*. Queste riunioni spesso mettono in luce le nostre intuizioni errate su quello che sappiamo dell'intelligenza. Le persone d'azienda cercano applicazioni dei sistemi AI che a loro sembrano immediate, ma spesso quei compiti sembrano immediati solo perché lo sono per *il nostro cervello*. Le persone esperte di apprendimento automatico a quel punto spiegano pazientemente perché l'idea che sembra semplice sia, in realtà, di difficoltà astronomica. Le discussioni di questo tipo vanno avanti per ogni nuovo processo. La mia curiosità per il cervello nasce da quelle esplorazioni del punto fino a cui possiamo spingere i moderni sistemi AI e dei casi sorprendenti in cui si rivelano inadeguati.

Ovviamente, sono anche un essere umano e, come voi, ho un cervello umano, perciò è stato facile essere affascinato dall'organo che

definisce tanta parte dell'esperienza umana. Il cervello offre risposte non solo sulla natura dell'intelligenza, ma anche sulle ragioni dei nostri comportamenti. Perché spesso facciamo scelte irrazionali e controproducenti? Perché la nostra specie ha una storia tanto lunga di altruismo motivante e allo stesso tempo di crudeltà insondabile?

Il mio progetto personale è iniziato semplicemente con il tentativo di leggere qualche libro per rispondere alle mie domande, ma poi si è esteso a lunghe corrispondenze via email con neuroscienziati che sono stati tanto generosi da tollerare le curiosità di un non addetto ai lavori. Questa ricerca e queste corrispondenze alla fine mi hanno condotto a pubblicare vari articoli, culminati nella decisione di prendermi del tempo libero dal lavoro per trasformare in un libro le idee che mi frullavano per la testa.

Nel corso di questo processo, quanto più entravo in profondità, tanto più mi rendevo conto della mancanza di una sintesi, che potesse offrire un'introduzione accessibile al funzionamento del cervello, al perché funziona proprio in quel modo, e per quali aspetti il suo funzionamento coincide con quello dei moderni sistemi AI (e per quali invece ne differisce); un libro che potesse ricondurre sotto un'unica narrazione le varie idee che sono nate nelle neuroscienze e nell'AI.

Questo libro è una sintesi del lavoro di molte altre persone; fondamentalmente, è solo un tentativo di ricomporre i pezzi che erano già presenti. Ho fatto del mio meglio per dare i giusti riconoscimenti, cercando di onorare le persone che hanno svolto le effettive ricerche scientifiche. Se in qualche caso non l'ho fatto, non è stato intenzionale. Certo, non ho resistito alla tentazione di indulgere a qualche mia speculazione, ma ho sempre cercato di essere chiaro sui momenti in cui mi sono avventurato in quel territorio.

Forse è appropriato che l'origine di questo libro, come l'origine del cervello stesso, non sia venuta da una pianificazione di base ma da un processo caotico di false partenze e strade a fondo cieco, iterazioni e circostanze favorevoli.

### *Un'ultima osservazione (a proposito di scale e sciovinismo)*

Devo fare un'ultima osservazione, prima di iniziare il nostro viaggio nel tempo. Fra le righe di tutta questa storia è pericolosamente in agguato un errore di interpretazione.

Traceremo molti confronti fra le capacità degli esseri umani e quelle di altri animali oggi viventi, ma sempre scegliendo specificamente quegli animali che si pensa siano i più simili ai nostri antenati. Tutto il libro, la struttura stessa basata sulle cinque grandi svolte, è

solo la storia dell'albero genealogico *umano*, la storia di come è arrivato a formarsi il *nostro* cervello; si potrebbe altrettanto facilmente raccontare la storia di come sia arrivato a formarsi il cervello dei polpi o delle api, e avrebbe i suoi colpi di scena e le sue grandi svolte.

Solo perché il nostro cervello gode di più facoltà intellettuali rispetto a quelle dei nostri antenati non vuol dire che il cervello umano moderno sia strettamente superiore intellettualmente a quelli di altri animali moderni.

L'evoluzione converge regolarmente per vie indipendenti su soluzioni comuni. L'innovazione delle ali è evoluta indipendentemente negli insetti, nei pipistrelli e negli uccelli; l'antenato comune di queste creature non aveva ali. Anche gli occhi si pensa siano evoluti in modo indipendente molte volte. Quindi, quando sostengo che una facoltà intellettuale, per esempio la memoria episodica, si è evoluta nei primi mammiferi, questo *non* vuol dire che oggi *solo* i mammiferi possiedano la memoria episodica. Come ali e occhi, altri rami dell'albero dell'evoluzione possono avere evoluto in modo indipendente una memoria episodica. In effetti, molte delle facoltà intellettuali di cui parleremo in questo libro non sono esclusive degli esseri umani, ma sono spuntate in modo indipendente su molti rami dell'albero dell'evoluzione sulla Terra.

Dai tempi di Aristotele, scienziati e filosofi hanno costruito quella che i biologi moderni chiamano una "scala della natura" (o, dato che gli scienziati amano usare i termini latini, una *scala naturae*). Aristotele ha creato una gerarchia di tutte le forme di vita, in cui gli esseri umani erano superiori agli altri mammiferi, a loro volta superiori a rettili e pesci, a loro volta superiori agli insetti, a loro volta superiori alle piante.

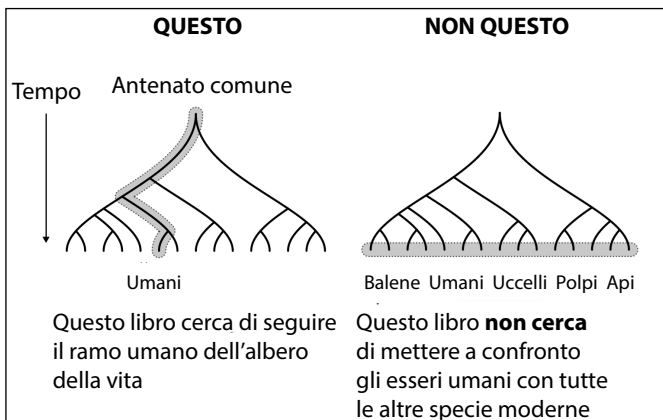


Figura I.2 L'obiettivo di questo libro.

Anche dopo la scoperta dell'evoluzione, l'idea di una *scala naturae* continua a rimanere viva, ma l'idea che esista una gerarchia delle specie è del tutto errata. Tutte le specie vive oggi sono, beh, *vive*; i loro antenati sono sopravvissuti agli ultimi 3,5 miliardi di anni di evoluzione. Perciò, in quel senso (l'unico senso possibile per l'evoluzione), tutte le forme di vita oggi esistenti sono prime a pari merito.

Le varie specie ricadono in nicchie di sopravvivenza diverse, per ciascuna delle quali l'ottimo è rappresentato da cose diverse. Per molte nicchie (in effetti, *la maggior parte* delle nicchie) sono più funzionali cervelli *più piccoli e più semplici* (o l'assenza di cervello). Le scimmie antropomorfe dal cervello molto sviluppato sono il risultato di una strategia di sopravvivenza diversa rispetto a quelle dei vermi, dei batteri o delle farfalle, ma nessuna è "migliore" delle altre. Agli occhi dell'evoluzione, la gerarchia ha solo due livelli: sul primo, le specie che sono sopravvissute, sull'altro, quelle che non ce l'hanno fatta.

Forse, invece, qualcuno vorrà parlare di *migliore* in base a qualche specifica caratteristica dell'intelligenza. Anche in questo caso, però, la classifica dipenderà completamente da quale specifica capacità intellettuale andiamo a misurare. Un polpo ha un cervello indipendente in ciascuno dei suoi tentacoli e può stracciare un essere umano nel multitasking. Colombi, chipmunk, tonni e perfino *iguane* possono elaborare le informazioni visuali più rapidamente di un essere umano<sup>8</sup>. I pesci hanno un'elaborazione in tempo reale incredibilmente accurata: avete mai visto con quale rapidità un pesce guizzi in mezzo a un labirinto di rocce se tentate di afferrarlo? Un essere umano sicuramente andrebbe a sbattere contro qualche ostacolo, se cercasse di spostarsi altrettanto rapidamente lungo un percorso accidentato.

Mi raccomando: nel ripercorrere la nostra storia, dobbiamo evitare di pensare che l'aumento della complessità dal passato al futuro possa significare che gli esseri umani moderni sono strettamente superiori agli animali moderni. Dobbiamo evitare di ricadere di nuovo nell'idea di una *scala naturae*. Tutti gli animali oggi viventi si sono evoluti per lo stesso periodo di tempo.

Esistono, ovviamente, aspetti che rendono unici gli esseri umani e, poiché *siamo umani*, ha senso che nutriamo un interesse particolare per capire noi stessi, e ha senso che cerchiamo di creare intelligenze artificiali *simili a quelle umane*. Spero perciò che affronteremo una storia centrata sugli esseri umani senza cedere allo sciovinismo umano. Si potrebbe raccontare una storia altrettanto valida per qualsiasi altro animale, dalle api ai pappagalli ai polpi, con cui condividiamo questo pianeta. Non racconteremo, però, quelle storie: questo libro racconta la storia di *una sola* di quelle intelligenze, la nostra.