



IL MITO
DELL'INTELLIGENZA
ARTIFICIALE

...

*Perché i computer
non possono pensare come noi*

...

ERIK J. LARSON

Informazioni per il lettore

Questo file PDF è una versione gratuita di sole 20 pagine ed è leggibile con **Adobe Acrobat Reader**



La versione completa dell'e-book (a pagamento) è leggibile **con Adobe Digital Editions**.

Per tutte le informazioni sulle condizioni dei nostri e-book (con quali dispositivi leggerli e quali funzioni sono consentite) consulta [cliccando qui](#) le nostre F.A.Q.

Tracce

I nuovi passaggi della contemporaneità

I lettori che desiderano informarsi sui libri e le riviste da noi pubblicati possono consultare il nostro sito Internet: www.francoangeli.it e iscriversi nella home page al servizio “Informatemi” per ricevere via e.mail le segnalazioni delle novità o scrivere, inviando il loro indirizzo, a “FrancoAngeli, viale Monza 106, 20127 Milano”.

**IL MITO
DELL'INTELLIGENZA
ARTIFICIALE**

•••

*Perché i computer
non possono pensare come noi*

•••

ERIK J. LARSON

Titolo originale: *The Myth of Artificial Intelligence*
Why Computers Can't Think the Way We Do
The Belknap Press of Harvard University Press
Cambridge, Massachusetts, 2021

Copyright © 2021 by Erik J. Larson
Published by arrangement with Harvard University Press

Traduzione dall'inglese di Pierluigi Micalizzi

1a edizione. Copyright © 2022 by FrancoAngeli s.r.l., Milano, Italy

L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore. L'Utente nel momento in cui effettua il download dell'opera accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sito www.francoangeli.it.

A Brooke e Ben

Indice

Introduzione pag. 9

Parte 1 Il mondo semplificato

- 1. L'errore dell'intelligenza » 17
- 2. Turing a Bletchley » 27
- 3. L'errore della superintelligenza » 40
- 4. La "Singolarità", passato e presente » 51
- 5. La comprensione del linguaggio naturale » 57
- 6. L'intelligenza artificiale come kitsch tecnologico » 67
- 7. Misteri e semplificazioni » 75

Parte 2 Il problema dell'inferenza

- 8. Non calcolate, analizzate » 93
- 9. L'enigma di Peirce » 98
- 10. Problemi relativi alla deduzione e all'induzione » 108
- 11. Machine learning e big data » 133
- 12. Inferenza abduttiva » 156
- 13. Inferenza e linguaggio I » 187
- 14. Inferenza e linguaggio II » 200

Parte 3
Il futuro del mito

15. Miti ed eroi	pag. 231
16. La mitologia dell'IA invade le neuroscienze	» 238
17. Teorie neocorticali dell'intelligenza umana	» 255
18. Fine della scienza?	» 261
Ringraziamenti	» 273

Introduzione

Nelle pagine che seguono leggerete del mito dell'intelligenza artificiale (IA). Il mito non consiste nell'impossibilità di realizzare l'IA autentica. Da questo punto di vista, il futuro dell'intelligenza artificiale è un'incognita scientifica. Il mito consiste invece nella sua presunta inevitabilità, nel fatto che sia solo una questione di tempo, perché ormai abbiamo imboccato la strada che condurrà all'IA di livello umano e poi alla superintelligenza. Non è così. Quella strada esiste solo nella nostra immaginazione. Eppure, l'ineludibilità dell'IA è talmente radicata nel discorso comune – alimentato da esperti dei media, *maître à penser* come Elon Musk e persino da molti scienziati del settore (per quanto non tutti) – che schierarsi contro è spesso considerato una forma di luddismo o quanto meno una visione miope del futuro della tecnologia e una pericolosa incapacità di prepararsi a un mondo di macchine intelligenti.

Come illustrerò nel libro, la scienza dell'IA ha svelato un grande mistero relativo all'intelligenza che nessuno ha idea di come risolvere. I fautori dell'intelligenza artificiale hanno tutto l'interesse a minimizzarne i limiti risaputi. In fin dei conti, l'IA è “big business” e sta acquisendo un ruolo sempre più rilevante nella cultura. Ma, che ci piaccia o no, la possibilità di avere sistemi intelligenti in futuro è limitata da ciò che attualmente sappiamo sulla natura dell'intelligenza. E a questo punto lo si deve affermare esplicitamente: tutte le evidenze indicano che l'intelligenza umana e quella delle macchine sono radicalmente diverse.

Il mito dell'IA continua a ribadire che le differenze sono temporanee e che i sistemi più potenti saranno infine in grado di cancellarle. Futurologi come Ray Kurzweil e filosofi come Nick Bostrom, illustri divulgatori del mito, ne parlano come se un'intelligenza artificiale pari a

quella umana sia inevitabile e si spingono a sostenere che, una volta raggiunto questo livello, le macchine superintelligenti ci surclasseranno.

Questo libro intende spiegare due aspetti importanti del mito dell'intelligenza artificiale, uno scientifico e uno culturale. La componente scientifica del mito ipotizza che dobbiamo semplicemente progredire sugli aspetti specifici dell'intelligenza, come i giochi o il riconoscimento delle immagini. Si tratta di un grave errore: il successo relativo alle applicazioni limitate non ci avvicina di un passo all'intelligenza generale. Le inferenze necessarie ai sistemi per l'intelligenza generale – leggere un giornale, sostenere una conversazione di base o diventare una collaboratrice domestica come Rosie, il robot de *I pronipoti* – non si possono programmare, apprendere e ingegnerizzare con la nostra attuale conoscenza dell'IA. Applicando con successo versioni semplici e ristrette dell'intelligenza che beneficiano di computer sempre più veloci e di molti dati, non facciamo reali progressi, ma ci limitiamo semplicemente a raccogliere i frutti dai rami più bassi. Il salto verso il “buon senso” generale è qualcosa di completamente diverso e non conosciamo il percorso per arrivarvi. Non esiste un algoritmo per l'intelligenza generale e vi sono motivi fondati per dubitare che questo emergerà continuando a lavorare sui sistemi di deep learning o attraverso altri approcci oggi comuni. Molto più probabilmente, servirà una scoperta scientifica rivoluzionaria, ma al momento nessuno ha la benché minima idea di come essa possa presentarsi, né, tanto meno, dei dettagli per arrivarvi.

Mitizzare l'intelligenza artificiale è negativo perché maschera un mistero scientifico con la narrazione infinita di un progresso costante. Il mito alimenta la fiducia nel successo inevitabile, mentre un rispetto autentico per l'impresa scientifica dovrebbe riportarci al punto di partenza. Questo ci conduce al secondo argomento del libro: le conseguenze culturali del mito. Rincorrere il mito non è un buon modo per seguire lo “smart money” e nemmeno un buon modo per mantenere una posizione neutrale. Fa male alla scienza e anche a noi. Perché? In primo luogo perché è improbabile che si riesca a innovare se decidiamo di ignorare un mistero fondamentale anziché affrontarlo. Una sana cultura dell'innovazione incoraggia l'esplorazione dei territori sconosciuti, non enfatizza lo sviluppo dei metodi esistenti, soprattutto quando questi si sono dimostrati inadeguati a farci progredire. Il mito dell'inevitabile successo dell'intelligenza artificiale tende a soffocare la cultura dell'invenzione necessaria per il progresso reale, con o senza un'IA di livello umano. Esso incoraggia inoltre la rassegnazione all'or-

rore di una terra delle macchine dove l'invenzione autentica è spinta ai margini da scenari futuristici che fanno affidamento sui metodi esistenti, spesso a causa di interessi ormai consolidati.

A chi si rivolge questo libro? Certamente a chiunque provi entusiasmo per l'IA, ma che si domanda perché essa sia sempre dieci o vent'anni di là da venire. Ciò ha una spiegazione scientifica che illustrerò nel seguito. Dovreste leggerlo anche nel caso in cui riteniate che il progresso dell'IA verso la superintelligenza sia ineluttabile e vi preoccupi il momento in cui si arriverà a quel punto. Benché non sia in mio potere dimostrare che non possa venire il giorno in cui l'IA dominerà il mondo, posso spiegarvi perché potete fare senz'altro a meno di considerare la possibilità di un tale scenario. Più in generale, dovreste leggere questo libro se siete semplicemente curiosi ma confusi dalle mistificazioni dell'IA diffuse nella nostra società. Spiegherò le origini del mito, ciò che sappiamo e non sappiamo relativamente alla possibilità di realizzare davvero un'intelligenza artificiale come quella umana e il motivo per cui dobbiamo apprezzare maggiormente l'unica intelligenza che conosciamo: la nostra.

I contenuti del libro

Nella Parte 1, "Il mondo semplificato", chiarirò in che modo la cultura dell'IA abbia semplificato il discorso sulle persone e abbia invece magnificato quello relativo alla tecnologia. Il tutto è cominciato con Alan Turing, l'iniziatore dell'intelligenza artificiale, e ha comportato semplificazioni comprensibili ma infelici che ho chiamato "errori dell'intelligenza". Gli errori iniziali furono trasformati in un'ideologia da I.J. Good, statistico e amico di Turing, che introdusse l'idea di "ultraintelligenza" come risultato prevedibile una volta raggiunto il livello umano di intelligenza. Fra Turing e Good, vedremo prendere forma il mito moderno dell'IA. Il suo sviluppo ci ha condotti nell'era di quello che ho definito il "kitsch tecnologico", imitazioni scadenti di idee più profonde che impediscono il coinvolgimento intelligente e indeboliscono la nostra cultura. Il kitsch ci dice come pensare e come percepire. Chi diffonde il kitsch ne trae giovamento mentre i suoi fruitori sperimentano una perdita. Questi ultimi – noi – si ritrovano così in un mondo superficiale.

Nella Parte 2, "Il problema dell'inferenza", sostengo che l'unico tipo di inferenza – in altre parole, di pensiero – adeguato all'intelligen-

za artificiale di livello umano (o qualcosa che gli si avvicini) è quello che non abbiamo idea di come programmare o progettare. Il problema dell'inferenza va al cuore del dibattito sull'IA perché riguarda direttamente l'intelligenza delle persone e delle macchine. La nostra conoscenza dei diversi tipi di inferenza risale ad Aristotele e ad altri pensatori della Grecia antica ed è stata sviluppata in campo logico e matematico. L'inferenza viene già descritta attraverso sistemi formali e simbolici come i programmi informatici, quindi attraverso una sua analisi è possibile avere un'idea molto chiara del progetto di costruzione di un'intelligenza artificiale. Esistono tre tipi di inferenza. L'intelligenza artificiale classica ne ha esplorato uno (deduzione), l'IA moderna un altro (induzione). Al terzo tipo di inferenza (abduzione), che produce l'intelligenza generale, nessuno sta lavorando!¹ Infine, poiché ogni tipo di inferenza è diverso – nel senso che non si può ridurre un tipo all'altro – sappiamo che l'impossibilità di costruire sistemi di IA utilizzando il tipo di inferenza che assicura l'intelligenza generale impedirà di fare progressi in direzione dell'intelligenza artificiale generale (AGI – Artificial General Intelligence).

Nella Parte 3, “Il futuro del mito”, sosterrò che il mito, se preso seriamente, ha conseguenze molto negative perché sovverte la scienza. In particolare, esso erode la cultura dell'intelligenza e dell'invenzione umana, che è necessaria proprio per le scoperte fondamentali di cui avremo bisogno per comprendere il nostro futuro. La scienza dei dati (l'applicazione dell'intelligenza artificiale ai “big data”) può tutt'al più rappresentare una protesi per l'ingegno umano, che, se usata correttamente, può venirci in aiuto nell'affrontare l'attuale “diluvio di dati”. Se però la si utilizza per sostituire l'intelligenza individuale, essa tende a bruciare investimenti senza fornire risultati. Mi concentrerò in particolare sulla spiegazione del modo in cui il mito abbia influenzato, tra i diversi ambiti scientifici, le neuroscienze. Il prezzo che paghiamo per il mito è troppo elevato. Poiché non abbiamo un valido motivo scientifico per ritenere che il mito sia vero e abbiamo invece molti motivi per respingerlo, consentendo un futuro di prosperità, dobbiamo ripensare radicalmente il discorso sull'intelligenza artificiale.

Note

¹ Non intendo sostenere che i ricercatori non si siano confrontati con l'abduzione nell'ambito delle ricerche sull'IA, lo hanno fatto. Negli anni Ottanta e Novanta i ricercatori lavorarono sugli approcci logici all'abduzione, la cosiddetta

“programmazione logica abduttiva”. Ma questi metodi erano abduttivi solo di nome, perché si basavano in realtà sulla deduzione, non sulla vera abduzione. Essi non ebbero successo e furono rapidamente abbandonati mentre gli studi sull'intelligenza artificiale si affacciavano sull'era del web. Più di recente, all'incirca dal 2010 a oggi, si sono adottati diversi approcci probabilistici (bayesiani, in particolare) come possibili traiettorie verso l'inferenza abduttiva autentica. Tuttavia, anche questi sistemi non costituiscono una modalità compiutamente abduttiva. Anziché essere approcci deduttivi mascherati come i loro predecessori, essi sono approcci induttivi o probabilistici camuffati. L'abduzione solo nominale non è ciò che intendo con abduzione e i sistemi che ne utilizzano il nome ma non risolvono il problema non ci aiuteranno a fare progressi in questo campo. Nel seguito del libro spiegherò le mie posizioni.

Parte 1

Il mondo semplificato

1

L'errore dell'intelligenza

La storia dell'intelligenza artificiale ha inizio con le idee di qualcuno dotato di un'immensa intelligenza umana: il pioniere dell'informatica Alan Turing.

Nel 1950, Turing pubblicò un articolo provocatorio, "Computer Machinery and Intelligence", che esplorava la possibilità di realizzare macchine intelligenti¹. Si trattava di uno scritto coraggioso, pubblicato quando i computer erano una novità e insignificanti rispetto agli standard odierni. Si trattava infatti di macchine lente e pesanti che velocizzavano operazioni come la decrittazione. Dopo una lunga preparazione, essi potevano per esempio ricevere equazioni fisiche e dati sulle condizioni iniziali, riuscendo così a calcolare il raggio di un'esplosione nucleare. IBM comprese fin da subito le potenzialità che essi avevano di sostituire gli esseri umani nell'esecuzione di calcoli per lavoro, per esempio nell'aggiornamento dei fogli di calcolo. Ma per considerare i computer come macchine "pensanti" serviva molta immaginazione.

L'idea di Turing faceva riferimento a un gioco popolare, conosciuto come "imitation game". Nella versione originale del gioco vi sono un uomo e una donna nascosti. Una terza persona pone loro per iscritto delle domande. Leggendo le risposte cerca quindi di indovinare quali siano da attribuire all'uomo e quali alla donna. La situazione è ingarbugliata dal fatto che l'uomo deve cercare di ingannare l'interrogatore, mentre la donna cerca di rendere ambigue tutte le risposte. Turing sostituì l'uomo e la donna con un computer e un essere umano e quindi diede avvio a quello che oggi chiamiamo "test di Turing". In esso un giudice umano riceve una risposta dattilografata da un computer e da un essere umano. Se non è in grado di identificare con certezza quella del computer, quest'ultimo supera il test. Turing sosteneva che nel caso in cui si fosse verificato questo esito non si poteva ritenere la mac-

china non intelligente, indipendentemente dalla sua eventuale natura umana. In tal modo, la questione dell'intelligenza della macchina si sostituì a quella della sua reale capacità di pensare.

Il test di Turing è davvero molto arduo, nessun computer lo ha mai superato. Ovviamente, Turing non poteva conoscere questo risultato di lungo termine nel 1950; tuttavia, sostituendo sgraditi interrogativi filosofici riguardanti la “coscienza” e il “pensiero” con un test con un esito osservabile, egli incoraggiava una visione dell'IA come una scienza legittima con uno scopo ben definito. Mentre l'IA prendeva forma nel corso degli anni Cinquanta, molti dei suoi pionieri e fautori si trovarono d'accordo con Turing: qualsiasi computer che conversava in modo prolungato e convincente con una persona, stava facendo qualcosa, come la maggior parte di noi ammetterebbe, che richiede il pensiero (di qualunque cosa si tratti).

La distinzione di Turing tra intuizione e ingegno

Turing si era guadagnato la sua reputazione come matematico molto prima di cominciare a occuparsi di intelligenza artificiale. Nel 1936, egli pubblicò un breve articolo matematico dedicato al significato esatto di “calcolatore”, termine che a quel tempo si riferiva a una persona che attraverso una serie di passaggi giungeva a una risultato definito (come l'esecuzione di un calcolo)². In questo articolo, Turing sostituì il calcolatore umano con l'idea di una macchina che svolgeva la stessa operazione. L'articolo si avventurava in calcoli matematici difficili, ma nella discussione sulle macchine non faceva riferimento al pensiero umano o alla mente. Le macchine possono funzionare in modo automatico, sosteneva Turing, e i problemi che risolvono non richiedono alcun ausilio “esterno” o intelligenza. Questa intelligenza esterna – il fattore umano – è ciò che i matematici talvolta chiamano “intuizione”. Il lavoro di Turing del 1936 contribuì a dare all'informatica la consistenza di una disciplina scientifica e rappresentava un contributo rilevante alla logica matematica. Turing non sembrava però soddisfatto di questa sua prima definizione che non coglieva qualcosa di essenziale. In effetti, la stessa idea di mente, o delle facoltà umane che presiedono alla soluzione dei problemi, ricomparve due anni dopo nella sua tesi di dottorato, un tentativo ingegnoso ma di fatto non riuscito di aggirare il risultato ottenuto da Kurt Gödel (che approfondiremo più avanti). La tesi di Turing contiene questo curioso passaggio sull'intuizione, che egli compara con un'altra facoltà mentale che chiama “ingegno”.

In modo alquanto schematico, si può considerare il ragionamento matematico come l'esercizio che combina due facoltà, che possiamo chiamare intuizione e ingegno. L'attività dell'intuizione consiste nel formulare giudizi spontanei che non sono il risultato dei flussi di ragionamento coscienti. Questi giudizi sono spesso ma non invariabilmente corretti (trascurando ciò che si debba intendere con "corretti"). Spesso è possibile trovare qualche altro modo per verificare la correttezza di un giudizio intuitivo. Si potrebbe per esempio ritenere che tutti i numeri interi positivi sono unicamente scomponibili nei loro fattori primi; un ragionamento matematico approfondito conduce al medesimo risultato. Esso comprenderà anche giudizi intuitivi, ma questi saranno quelli meno suscettibili di critiche rispetto al giudizio originale della scomposizione. Non mi addentrerò in ulteriori spiegazioni di questa idea di "intuizione".

Turing procede quindi con la spiegazione dell'ingegno: "L'esercizio dell'ingegno in matematica consiste nel coadiuvare l'intuizione attraverso un'organizzazione appropriata delle proposizioni e magari di figure geometriche o disegni. Quando le proposizioni sono disposte in modo davvero adeguato, i passaggi intuitivi necessari non possono essere messi in dubbio"³.

Benché utilizzi un linguaggio specialistico, Turing sta evidenziando un'ovvietà: i matematici di solito selezionano i problemi o "vedono" un problema interessante di cui occuparsi utilizzando una capacità che *apparentemente* non può essere suddivisa in fasi e quindi è evidentemente inadatta alla programmazione informatica.

La visione di Gödel

Anche Gödel aveva riflettuto sull'intelligenza meccanica. Come Turing, egli era ossessionato dalla differenza tra ingegno (meccanica) e intuizione (mente). La sua distinzione era essenzialmente la stessa evidenziata da Turing, ma formulata con un linguaggio diverso: dimostrazione contro verità (o, nel gergo matematico, "teoria della dimostrazione" e "teoria dei modelli"). I due concetti di dimostrazione e verità, si domandava Gödel, sono in fin dei conti la stessa cosa? In tal caso, la matematica e anche la scienza sarebbero comprensibili in termini puramente meccanici. Non solo. Adottando questa prospettiva, anche il pensiero umano sarebbe meccanico. Il concetto di "intelligenza artificiale", espressione allora non ancora coniata, aleggiava sulla questione. L'intuizione della mente, la sua

capacità di cogliere la verità e il significato, è riducibile a qualcosa di meccanico, alla capacità di calcolo?

Questa era la domanda che Gödel si poneva. E nella sua ricerca di una risposta, egli si imbatté in un imprevisto che di lì a breve gli avrebbe procurato fama mondiale. Nel 1931, Gödel pubblicò due teoremi di logica matematica, noti come “teoremi di incompletezza”. Attraverso di essi, l'autore dimostrò i limiti intrinseci di tutti i sistemi matematici formali. Un colpo di genio. Gödel dimostrò in modo inequivocabile che la matematica – *tutta* la matematica, sotto certe semplici condizioni – non è, a rigor di termini, meccanica o “formalizzabile”. Più nello specifico, egli dimostrò che in qualunque sistema formale (matematico o computazionale) devono esistere enunciati Veri, con la “V” maiuscola, che non sono dimostrabili all'interno del sistema stesso, utilizzandone le regole. L'enunciato Vero è riconoscibile dalla mente umana ma è (in modo dimostrabile) non dimostrabile dal sistema all'interno del quale è formulato.

In che modo Gödel giunse a queste conclusioni? I dettagli sono tecnici e complicati, ma l'idea di base di Gödel è che possiamo considerare un sistema matematico sufficientemente complesso da poter eseguire un'addizione come un sistema di significato, quasi al pari di una lingua naturale come l'inglese o il tedesco, e lo stesso principio è applicabile a tutti i sistemi più complessi. Considerandolo in questo modo, consentiamo al sistema di dare conto di se stesso. Per esempio, esso ci può rivelare che ha alcuni limiti. Questa è stata la scoperta di Gödel.

I sistemi formali come quelli matematici consentono di enunciare con esattezza verità e falsità. Di solito, stabiliamo la verità utilizzando lo strumento della dimostrazione (utilizziamo regole per dimostrare qualcosa e in tal modo sappiamo che è sicuramente vero). Ma esistono enunciati veri che non si possono dimostrare? La mente è in grado di conoscere qualcosa che il sistema non è in grado di conoscere? Nel caso semplice dell'aritmetica, esprimiamo la verità attraverso equazioni come “ $2 + 2 = 4$ ”. Le equazioni comuni sono enunciati veri nel sistema dell'aritmetica e sono dimostrabili impiegando le sue regole. In questo caso, dimostrabile equivale a vero. Prima di Gödel, i matematici pensavano che tutta la matematica fosse dotata di questa proprietà. Ciò comportava che le macchine potessero produrre tutte le verità nei diversi sistemi matematici semplicemente applicando le regole in modo corretto. È un'idea stupenda ma non è vera.