

**Nello Cristianini, “Sovrumano – Oltre i limiti della nostra intelligenza”,
Il Mulino, 2025**

[Riduzione di Carlo Dellabella]

Prologo → La domanda non è più oggi se le macchine possono essere intelligenti, ma se possono esserlo **come e più** di un essere umano. Siamo restii ad ammettere: 1) che una macchina possa essere più intelligente di noi; 2) a considerare cosa c'è oltre i limiti della nostra intelligenza.

Già Alan **Turing** nel lontano 1950 parlava di questa nostra *riluttanza* ed elencava una serie di obiezioni che aveva sentito. Questa memorabile e divertente: «Va beh, ma nessuna macchina intelligente potrà mai apprezzare le fragole con la panna.»

PARTE I → ANI

In questa prima parte, incontriamo agenti intelligenti – come AlphaGo – ResNet – AlphaFold – che possono **superare gli esseri umani**, anche quelli più esperti, sia pure **in settori specializzati**.

[Siamo all'inizio degli anni Venti o un po' prima.]

L'autore comincia col portare l'esempio del **Go** (gioco da tavolo di tipo strategico) – secondo me era migliore quello degli **scacchi** – dove ormai la macchina supera di gran lunga i campioni mondiali.

Nota – L'accelerazione nelle prestazioni dell'IA comincia nel 2012 con la soluzione di un problema che resisteva da anni: il **riconoscimento visivo di un oggetto**. Prima le macchine intelligenti non raggiungevano neanche la capacità di un bambino di un anno quanto a percezione visiva.

Nel 2007 si cominciò a capire che occorreva creare una collezione di immagini con cui insegnare alla macchina a riconoscere **ogni** oggetto esistente. Il risultato fu ImageNet, che al momento contiene 12 milioni di immagini, divise in 20 mila categorie: gli algoritmi di *machine learning* le usano per imparare a riconoscere gli oggetti.

Si aprì così una gara tra scienziati **addestratori** di algoritmi sempre più potenti e scienziati **creatori di test** sempre più sofisticati per valutarne le capacità.

Nell'ottobre 2012 a Firenze il gruppo vincitore ottenne il risultato di **83,6 % di immagini corrette**. La scala era enorme: si trattava della più grande rete neurale mai addestrata, con 60 milioni di connessioni tra i neuroni. Il loro algoritmo di chiamava AlexNet e tra gli autori c'erano Geoff **Hinton**, il padre delle reti neurali (Premio Nobel nel 2024) e Ilya **Sutskever**, co-inventore di GPT.

L'anno dopo un altro algoritmo, ResNet, raggiunse una precisione del 94,6 %. Da quel momento prese il via una corsa con la creazione di modelli sempre più potenti, fino a toccare il 98%. Ormai le macchine ci superano in compiti di riconoscimento visivo (accuratezza+quantità di categorie).

Considerazione generale – La natura dell'IA è questa: raggiunge prestazioni altissime, **a condizione di essere fortemente specializzata.**

Oltre gli esperti → Così nel giro di pochi anni esperti di discipline diverse sono stati eguagliati, superati, talvolta distaccati. In particolare in medicina, nell'esame di radiologie e TAC, ma anche in molti altri campi. È inutile che ci domandiamo che aspetto avrebbe un mondo in cui agenti intelligenti ci superano: **ci siamo già dentro**, almeno nel caso di ANI.

I benefici collettivi sono evidenti, specialmente in medicina. Ma è tutto qui **il sogno delle macchine intelligenti**? Specialisti simili ad AlexNet, ciascuno concentrato su un problema diverso?

PARTE II →AGI

Il successo delle reti neurali “profonde” (come Resnet) ha acceso la speranza di un'Intelligenza Artificiale Generale, cioè “una macchina in grado di svolgere qualsiasi compito intellettuale che un essere umano possa svolgere”: questo l'obiettivo a lungo termine di OpenAI (che ha costruito GPT) e di altri gruppi della Silicon Valley.

È iniziata l'avventura di cercare un'intelligenza **flessibile** comparabile alla nostra. Ma **come definirla e come misurarla?**

Un modo è questo: sottoporre gli agenti intelligenti a test simili ai test psicometrici umani, test che valutano le nostre capacità cognitive.

Così la sintesi unificata che oggi si chiama CHC – creata per misurare le prestazioni degli agenti intelligenti – si basa sullo studio (1993) dello psicologo americano John **Carroll**, che – sull'analisi di 460 test di intelligenza – proponeva un modello di intelligenza umana “a 3 strati” (*narrow, broad, general*).

Domanda → È *razionale* paragonare l'IA all'intelligenza umana? Non può essere quella artificiale un'*altra forma* di intelligenza?

Comunque l'obiettivo del momento sembra essere quello di “eguagliare le prestazioni umane”.

NOTA – Il saggio che stiamo esaminando dedica molte pagine ai “lavori in corso”, cioè a dare un'idea su quello che si sta facendo per “misurare” l'Intelligenza Artificiale Generale.

Qui – per brevità - non possiamo seguirlo.

Diciamo solo questo: addestratori di IA e creatori di test [v. nel glossario *benchmark*] sono impegnati in una lunga partita per esplorare i limiti delle macchine intelligenti. Per i primi l'obiettivo sarebbe una macchina in grado di risolvere qualsiasi problema, per i secondi un test insuperabile per qualunque macchina, ma non per noi.

Alla fine nel 2021 dai ricercatori di Berkeley fu creato MMLU (*Massive Multitask Language Understanding*), un *benchmark* davvero formidabile per mettere alla prova le prestazioni di GPT e simili (v. glossario).

Ciò costrinse le case produttrici ad aumentare sempre più le dimensioni dei modelli, addestrandoli con enormi quantità di dati.

L'anno scorso (2024) è accaduto l'inevitabile: l'ultima versione di Gemini, chiamata *Ultra* (dotata di 1.760 miliardi di parametri), ha raggiunto il 90% di risposte corrette, superando gli esperti umani. Poco dopo anche GPT-4 ha effettuato il sorpasso.

Dove ci porterà questa (apparentemente) inarrestabile competizione? Forse in un futuro ci saranno macchine intelligenti specializzate nel valutare le capacità cognitive di altre macchine intelligenti.

Ed è questo un segno che l'AGI è più facile del previsto oppure quei test non sono abbastanza difficili?

NOTA – Nella mente umana si misurano anche altri aspetti: attitudini, valori, personalità. Sarebbe sbagliato farlo anche in una IA? Già l'*AI Act* dell'UE menziona per l'IA alcuni obblighi legali, quali l'equo trattamento dei cittadini o le questioni di *genere*. Per le nostre macchine intelligenti gli esami potrebbero non finire mai.

AGI all'orizzonte?

Ci sono diverse definizioni di AGI e questo complica il modo in cui misuriamo il progresso verso questa meta.

Gli scienziati ricercatori hanno opinioni diverse su quali caratteristiche debba avere l'AGI. Questo certo non aiuta, anche se fa parte della sfida: cercare qualcosa senza sapere con precisione cos'è.

Tuttavia, sembra abbiano convenuto che ci sono due strade da seguire: una per i problemi che risolviamo per intuito, “a colpo d'occhio” [Sistema 1]; un'altra per quelli che richiedono di ragionare, di “pensare un passo alla volta” [Sistema 2]. (Secondo la definizione del premio Nobel Daniel **Kahneman**)

Quasi tutti i problemi del Sistema 1 sono stati risolti aumentando la scala dei modelli. Ma ciò non sembra funzionare per i problemi del Sistema 2, che richiedono la capacità, per arrivare ad una soluzione, di scomporre il problema in due o più passaggi. [Esempio p.78]. Modelli linguistici come GPT non superavano il 50%, a paragone del 98% di esperti umani. Per cui esiste una debolezza critica dei moderni modelli linguistici riguardo i problemi Sistema 2.

Cosa fare di fronte a questo *impasse*? Secondo alcuni la strada da percorrere è aumentare la dimensione dei modelli di IA, perchè si è visto che certe proprietà cognitive emergono **solo ad una data scala** [v. Glossario *Scaling Hypothesis*]. Per fare un esempio: GPT-2 aveva 1 miliardo di parametri, GPT-4 ne ha 1.000 miliardi.

Altri invece pensano che la via non è questa e saranno necessarie innovazioni prima di raggiungere una forma di AGI.

Nel primo caso si deve investire in **maggiori risorse**, nel secondo in **maggior ricerca**. Al momento l'industria sta scommettendo sulla prima *chance*.

NOTA – L'Ipotesi di scala potrebbe portare l'IA in territori sovrumani (almeno per il Sistema 1), però non solo in termini di prestazioni, anche di risorse. Infatti esistono dei “*muri*”, difficilmente valicabili.

Il muro della computazione (ovvero l'hardware) → Oggi la quantità delle operazioni richieste per addestrare un modello è tale che presto la domanda sarà superiore all'offerta. Il FLOP [*Floating Point Operation*] è il nome che si dà ad una singola operazione elementare. Nel 2022 GPT-3,5 ha usato 2,6 miliardi di petaFLOP; nel 2024 Llama-3 ne ha usato 38 miliardi. Un petaFLOP corrisponde in operazioni al numero 10 elevato alla quindicesima potenza.

Il muro dell'energia → Usare i computer richiede energia, non solo per i calcoli, ma anche per il raffreddamento. Addestrare GPT-3 è costato 4 milioni di dollari, addestrare GPT-4 è costato 40 milioni di dollari. Il costo economico sommato all'impatto ambientale comincia a diventare rilevante.

Il muro dei dati → Una delle leggi del *machine learning* è che per addestrare modelli più grandi ci vogliono più dati. Quanto è ancora disponibile in termini di dati? Nel giugno 2024 uno studio ha certificato che il consumo di dati cresce più velocemente della loro produzione. Le previsioni dicono che all'inizio degli anni Trenta i modelli non potranno più essere addestrati con una massa di dati crescente.

Da qui il dubbio: quanta energia, dati, computazione ha senso spendere per creare – forse – una AGI?

Da qui anche, se non converrebbe imboccare **un'altra strada**. Ma è possibile indurre gli attuali modelli linguistici a ragionare “un passo alla volta” per risolvere problemi di ragionamento?

Nell'aprile 2024 il prof. Subbarao **Kambhampati** scriveva in un articolo: «Nulla per ora mi dà ragione per credere che i modelli di linguaggio svolgano attività di ragionamento.»

Per confermare questo suo giudizio, Kambhampati creò una serie di test, il *PlanBench*, al fine di valutare se un modello come GPT fosse in grado di ragionare. La conclusione fu che perfino GPT-4 risolveva solo una piccola percentuale di domande di ragionamento.

Emily **Bender** definì i modelli linguistici “pappagalli” in grado solo di ripetere le cose che avevano dentro, senza comprenderle.

Il fatto è che OpenAI aveva definito la capacità di ragionamento come “il secondo livello nella scalata verso un'AGI”. Per cui gli investitori, un po' dubbiosi, stavano a guardare.

Ora, c'è un'idea che potrebbe risolvere il problema e deriva dal fatto che noi “pensiamo un passo alla volta”. Pertanto – sulla base di un'osservazione fatta da Jason **Wei** anni prima (2022) – si adottò un metodo chiamato *catena di pensieri* [*chain of thought*], che in pratica consiste nel chiedere a GPT di pensare “ad alta voce” descrivendo i passaggi uno alla volta. E si vide che ciò migliorava la capacità dei grandi modelli linguistici di eseguire ragionamenti.

Così nella gara ricordata prima per superare il *benchmark* MMLU OpenAI – sul finire dell'anno scorso – aggiunse un nuovo candidato, chiamato ufficialmente «o1», ma noto a tutti come “*strawberry*” (fragola), in grado di superare i modelli precedenti e di risolvere problemi di ragionamento matematico. I dettagli tecnici sono segreti, ma l'idea di fondo fu spiegata: «o1» riusciva a controllare “un monologo interiore”, che lo faceva procedere “un passo dopo l'altro”.

Naturalmente ciò ha fugato i dubbi degli investitori, mentre gli scienziati valutatori sono impegnati a creare nuovi test, sulla base della nostra innata riluttanza ad arrenderci: «Sì, va bene, però...» [v. il famoso articolo di Alan Turing del 1950]

Nel 2023 Peter **Norvig** ha scritto un articolo analogo a quello di Turing, in cui sostiene: 1) una prima forma di AGI è già esistente; 2) tuttavia siamo restii ad accettare quest'idea.

Poi ne specifica i motivi (tre mi sembrano importanti): 1) un essere umano è una creatura eccezionale, biologica e mentale, non paragonabile ad una macchina; 2) preoccupazione per le conseguenze economiche e sociali; 3) dubbio sulla validità dei test psicometrici che misurano l'intelligenza. Scrive Nello Cristianini: «Siamo una strana specie: spendiamo miliardi per costruire una macchina in grado di pensare e nello stesso tempo ci rifiutiamo di accettarlo... Cerchiamo e insieme temiamo quell'incontro... Ad ogni singolo test la vediamo avvicinarsi e allora ci affrettiamo a creare un nuovo test, che ci sembri rassicurante.»

PARTE III →ASI

Un'ipotetica forma di intelligenza *sovrumana* – ASI – riguarderebbe →
1) livelli di prestazione che non possiamo raggiungere
2) compiti non accessibili agli esseri umani

Nel caso, *una* questione, forse *la* questione sarebbe: come condividere il mondo con macchine più intelligenti di noi.

Il premio Nobel per la Fisica 2024 Geoff **Hinton** (co-inventore di Alexnet) ha detto sull'IA: «Sarà paragonabile alla rivoluzione industriale: invece di superare le persone in forza fisica, le supererà in capacità intellettuali. Non abbiamo esperienza di cosa significhi avere a che fare con “cose” più intelligenti di noi.»

Questo argomento sembra fantascienza, tuttavia oggi se ne discute seriamente nelle università di tutto il mondo (e le aziende produttrici sono in attesa di risultati).

Entrambe cioè considerano – almeno come possibilità – l'idea di macchine “super intelligenti”.

Le implicazioni sarebbe profonde e gravi per la società, creando “competizione” con una gran parte di lavoratori.

Cristianini elenca i settori in cui l'IA “può batterci al nostro gioco”: scrittura di testi, traduzione simultanea, qualsiasi lavoro che si possa svolgere in *smart working*, risultati scientifici di valore, ecc.

Tuttavia – secondo l'autore – in tutto questo, anche superandoci, l'IA resterebbe “umana”.

Allora si chiede: una futura ASI non potrebbe **comprendere cose che non sono comprensibili per noi**, quanto la meccanica quantistica non lo è per la sua gatta? Non sarebbe possibile che un agente intelligente ottenga una comprensione di aspetti del mondo che a noi non sono comprensibili?

E fa l'esempio: «La mia gatta non comprende la meccanica quantistica e nemmeno sa cosa si perde. Chissà cosa pensa lei delle mie capacità, dato che non sento arrivare il postino e non vedo un topo in fondo al giardino. Chissà cosa mi perdo.»

Secondo la psicologia cognitiva noi – fin dall'infanzia, assieme ad altri primati – abbiamo una conoscenza di base (*core knowledge*), come permanenza dell'oggetto, concetti numerici semplici, ragionamento causale, cognizione sociale.

Ora, una intelligenza diversa – che non fosse dotata di un'uguale “conoscenza di base” – potrebbe descrivere lo stesso mondo in modo diverso.

Ancora. Il nostro cervello non può seguire il metabolismo interno di una cellula, dove migliaia di sostanze chimiche interagiscono in modi diversi. Una macchina può farlo.

Quale allora il futuro in una convivenza con un'eventuale ASI? La nostra mente è limitata, ma è duttile e flessibile, per cui andremo in cerca e forse troveremo le debolezze, le vulnerabilità di una simile intelligenza.

Quest'anno – 2025 – gli scienziati valutatori si stanno ponendo la questione: esistono compiti cognitivi inaccessibili alle macchine? Il progetto è selezionare almeno 1.000 domande (a cui solo individui eccezionali sanno rispondere) da sottoporre all'IA attuale. È stato chiamato “l'ultimo esame dell'umanità” [*Humanity's Last Exam*], in quanto implica il rischio che le macchine eguagliarci e sorpassarci.

Si tratta di una sfida alla nostra riluttanza ad ammettere che le macchine possano superarci in intelligenza.

Epilogo

Un giorno forse i potenti meccanismi che stiamo costruendo saranno “sovrumani”.

Ricordiamoci allora che quella sarà solo intelligenza: non coscienza, non emozione, non libero arbitrio, non volontà. Tutte queste cose ci appartengono, sono la nostra umanità e non appartengono a nessuna macchina.

Il saggio si chiude con queste parole: «Se incontreremo qualcosa di superiore a noi, gli diremo fieri: Sì, però...»