

Intelligenza artificiale: un percorso per termini chiave

Irene Fabbri

Il linguaggio è la lente attraverso cui comprendiamo e interpretiamo il mondo; per secoli, è stato considerato una delle principali caratteristiche distintive dell'intelligenza umana. La nostra capacità di comunicare idee complesse, di costruire narrazioni e di trasmettere conoscenza attraverso le parole è stata spesso vista come un tratto unico, come ciò che ci distingue dalle altre specie.

In questo articolo si propone un percorso a tema intelligenza artificiale, seguendo come filo conduttore la spiegazione di alcuni termini chiave sull'argomento. Sebbene noi tutti ormai abbiamo in mente e conosciamo gran parte delle parole in uso sul tema, esplorare il linguaggio legato all'intelligenza artificiale non è solo un esercizio accademico, è un modo per prendere la rincorsa al fine di spingersi verso una comprensione più profonda di come queste tecnologie influenzeranno e modelleranno il futuro.

Il percorso non può che aprirsi con il termine stesso **Intelligenza artificiale**.

Dal vocabolario **Treccani** <https://www.treccani.it/enciclopedia/intelligenza-artificiale/>:

intelligenza artificiale (IA) *Disciplina che studia se e in che modo si possano riprodurre i processi mentali più complessi mediante l'uso di un computer. Tale ricerca si sviluppa secondo due percorsi complementari: da un lato l'i. artificiale cerca di avvicinare il funzionamento dei computer alle capacità dell'intelligenza umana, dall'altro usa le simulazioni informatiche per fare ipotesi sui meccanismi utilizzati dalla mente umana.*

Si noti che, sebbene si parli spesso di intelligenza artificiale come una nuova proprietà delle macchine o come una nuova particolare tecnologia o un particolare strumento (ad esempio si dice che ChatGPT sia un'intelligenza artificiale), in realtà noi spesso parliamo di qualcos'altro, sottintendendo la parola "sistemi".

Andando a guardare un'altra fonte risulta tutto più chiaro, dal documento ISO/IEC 22989:2022(en) Information technology — Artificial intelligence — Artificial intelligence concepts and terminology:

3.1.3

intelligenza artificiale

IA

<disciplina> *ricerca e sviluppo di meccanismi e applicazioni di sistemi IA (3.1.4)*

Nota 1 alla voce: La ricerca e lo sviluppo possono avere luogo in qualsiasi numero di campi come l'informatica, la scienza dei dati, le scienze umanistiche, la matematica e le scienze naturali.

3.1.4

sistema di intelligenza artificiale

sistema IA

sistema ingegnerizzato che genera output come contenuti, previsioni, raccomandazioni o decisioni per un dato insieme di obiettivi definiti dall'uomo

Nota 1 alla voce: Il sistema ingegnerizzato può usare varie tecniche e approcci legati all'intelligenza artificiale (3.1.3) per sviluppare un modello (3.1.23) per rappresentare dati, conoscenza (3.1.21), processi, ecc., che possono essere usati per svolgere compiti (3.1.35).

Nota 2 alla voce: I sistemi IA sono progettati per operare con vari livelli di automazione (3.1.7).

Tornando all'esempio di prima, ChatGPT è quindi un sistema di intelligenza artificiale, una tecnologia che nasce grazie a una serie di tecniche sviluppate da persone esperte di IA.

La nascita del termine Intelligenza Artificiale

Il termine "Intelligenza artificiale" fu usato per la prima volta in occasione della Dartmouth Conference del 1956, evento riconosciuto come momento fondativo della disciplina dell'IA come campo di studio autonomo. Qualche mese prima della conferenza, il 31 Agosto del 1955 fu pubblicato il documento che proponeva l'organizzazione del convegno: "A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE" di J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon.

Nel testo viene utilizzato per la prima volta il termine "intelligenza artificiale" ma se ne fa un uso spontaneo e non è presente alcuna definizione formale. Nella proposal è presente invece un elenco di punti che evidenzia i problemi legati all'intelligenza artificiale, che saranno oggetto del convegno:

1. Calcolatori automatici
2. Come si può programmare un computer in modo che possa usare il linguaggio
3. Reti neurali
4. Teoria delle dimensioni di un calcolo
5. Auto-miglioramento
6. Astrazione
7. Casualità e creatività

Per contestualizzare un po' questa lista, c'è da dire che gli organizzatori della conferenza sono di formazione informatici e matematici, non linguisti. Eppure, nel loro sogno di voler raggiungere questa "intelligenza artificiale" al secondo posto individuano come fondamentale la necessità che il computer manipoli il linguaggio poiché scrissero: "Si può ipotizzare che una larga parte del pensiero umano consista nell'usare parole in base a regole di ragionamento e a regole di congettura." Quindi nella loro idea l'intelligenza artificiale avrebbe dovuto avere un forte legame con il linguaggio, con la possibilità di interagire con noi usando la nostra lingua.

Oggi si parla di tre categorie di intelligenza artificiale

Oggi quando si parla di intelligenza artificiale, il concetto si suddivide in tre categorie principali, ciascuna rappresentativa di un diverso livello di capacità e complessità. **L'intelligenza artificiale ristretta**, nota anche come IA debole, è quella che abbiamo ora e che punta allo sviluppo di sistemi per eseguire compiti molto specifici, come il riconoscimento delle immagini o la comprensione del linguaggio naturale. Procedendo lungo il continuum dell'evoluzione dell'IA, incontriamo **l'intelligenza artificiale generale (AGI)**, che mira allo sviluppo di sistemi in grado di emulare la versatilità dell'intelligenza umana, immaginando macchine in grado di apprendere e operare in contesti vari e non predefiniti, analogamente a quanto avviene con il pensiero umano. Infine, al vertice di questa gerarchia teorica si colloca la **superintelligenza artificiale**, che punta ad implementare sistemi non solo in grado di replicare ma addirittura di superare alcune capacità umane, inclusi il ragionamento astratto e la creatività. Ad oggi noi abbiamo solo sistemi che

rientrano nell'intelligenza artificiale ristretta. Alcune aziende si pubblicizzano dicendo di stare arrivando a sistemi di AGI ma al momento non ci sono e parlare di superintelligenza è ad oggi una speculazione.

Invece, passando a un tipo di categorizzazione diversa, basata sul *task* (compito) per il quale nasce un sistema di IA, oggi si parla moltissimo delle intelligenze artificiali generative.

Intelligenza artificiale generativa (IA generativa): sistema di Intelligenza Artificiale progettato per creare contenuti nuovi e originali. Esistono varie tipologie di IA generative, alcuni esempi sono:

- Text-to-Text: trasformano testo in testo (come nel caso di ChatGPT o Gemini)
- Text-to-Image: generano immagini a partire da descrizioni testuali.
- Text-to-Audio: generano audio o musica basati su descrizioni testuali.
- Speech-to-Text: convertono l'audio parlato in testo scritto. Sono alla base di assistenti virtuali e di software di trascrizione.

L'interazione con i sistemi di IA generativa avviene attraverso i *prompt*.

Prompt: con il termine prompt si indica un'istruzione o a una serie di istruzioni fornite a un sistema di elaborazione del linguaggio naturale (come ad esempio ChatGPT), al fine di ottenere una risposta specifica o guidare la generazione di un output coerente e pertinente. I prompt sono utilizzati per interagire con i modelli di Intelligenza Artificiale e ottenere risposte desiderate o testi / immagini / suoni conformi a un certo contesto o scopo. Una scelta accurata del prompt può aiutare a ottenere risultati coerenti.

Prompt-tuning: il prompt-tuning è una pratica nell'ambito dell'Intelligenza Artificiale e del Machine Learning, in particolare quando si lavora con modelli di linguaggio o sistemi di generazione di testo. Si riferisce al processo di ottimizzazione o affinamento dei prompt al fine di ottenere risposte più accurate, coerenti o pertinenti da parte del modello.

La base dei sistemi di IA

Come detto nel primo paragrafo, i primi a parlare di intelligenza artificiale furono informatici e matematici. Alla base di tutta la tecnologia e di tutta la conoscenza nell'ambito dell'IA ci sono macchine calcolatrici e teorie matematiche, nonché algoritmi.

Un **algoritmo** è una sequenza di istruzioni ben definite e logicamente strutturate, utilizzate per risolvere un problema o svolgere una specifica attività. Nei sistemi di intelligenza artificiale, gli algoritmi rappresentano il cuore dei processi di calcolo e decisione, svolgono quindi un ruolo fondamentale nell'elaborazione dei dati e nel completamento dei task.

Un algoritmo che prende spunto dal cervello

Nel documento citato in apertura, il punto 3 cita le reti neurali e sembra quindi una buona idea quella di definire anche in questo caso di cosa si tratta.

Rete neurale artificiale: un modello computazionale ispirato al funzionamento dei neuroni nel cervello umano. Si tratta di una struttura matematica composta da unità chiamate "neuroni artificiali" o "nodi", organizzate in strati e connesse da pesi sinaptici. Le reti neurali artificiali vengono utilizzate nell'ambito dell'intelligenza artificiale, in particolare nel Machine Learning, per elaborare dati complessi e svolgere compiti di apprendimento e inferenza. Ogni neurone artificiale è una piccola unità di calcolo che riceve input, applica una trasformazione matematica (una funzione) a tali input e produce un output. I pesi sinaptici tra i neuroni rappresentano l'intensità e la direzione delle connessioni tra di essi. Nel Machine Learning, durante il processo di

apprendimento, questi pesi vengono aggiornati in base agli esempi di addestramento al fine di far sì che la rete neurale apprenda dalle relazioni nei dati.

Dopo questa definizione, mi sento di voler fare una precisazione, citando Korzybski “La mappa non è il territorio”: è vero che il modello matematico delle reti neurali artificiali è ispirato al cervello umano ma questo non rende intelligente la macchina che esegue i calcoli secondo quell’algoritmo. Le macchine non pensano nel modo in cui pensiamo noi ma semplicemente eseguono dei calcoli, calcoli complessi, calcoli ispirati in qualche modo alla struttura del nostro cervello ma pur sempre calcoli.

Macchine e apprendimento

Si è soliti dire che nel Machine Learning le macchine “apprendono”, questa capacità di apprendimento è il frutto dell’applicazione di algoritmi che analizzano, interpretano e inferiscono informazioni a partire da ampi dataset.

Un **dataset** è una collezione strutturata di dati utilizzati per allenare, testare e valutare modelli di machine learning e deep learning. Questi dati possono essere di varia natura, come testo, immagini, audio o video, e sono organizzati in modo che un algoritmo possa apprenderne le caratteristiche, le strutture o i pattern per eseguire compiti specifici come classificazione, riconoscimento, previsione e generazione. Nel caso in cui si lavori con dei modelli di linguaggio (come ChatGPT) poiché il dataset è composto da testi, può essere anche chiamato con il nome **corpus**.

Ancora una volta, l’apprendimento di una macchina è il risultato dell’esecuzione di una serie di calcoli, ma cosa apprendono le macchine? Le macchine “apprendono” nel senso che sono in grado di riconoscere e riproporre dei pattern.

Il termine **pattern** si riferisce a regolarità, strutture o tendenze ricorrenti nei dati, che i modelli di Machine Learning e Deep Learning cercano di identificare, imparare e sfruttare per compiere previsioni o prendere decisioni. I pattern possono manifestarsi in vari modi a seconda della natura dei dati e del contesto specifico, e possono includere sequenze, associazioni, correlazioni tra variabili, distribuzioni di probabilità, configurazioni spaziali o temporali.

Un termine chiave *Bias*

Poiché le macchine apprendono riconoscendo degli schemi, le macchine “conoscono” solo ed esclusivamente ciò che è presente all’interno del dataset. Maggiore sarà la presenza di certi schemi o dati, più facilmente la macchina “apprenderà” tali schemi. Ad esempio, volendo fare un sistema di IA per il riconoscimento di immagini, un sistema che abbia il compito di distinguere i gatti dai cani, in tal caso avrei bisogno di una collezione di immagini (un dataset) che contenga sia foto di gatti che di cani. Tuttavia, se nel dataset ci saranno moltissime foto di gatti e poche foto di cani non riuscirò mai ad avere un buon modello, perché la macchina sarà abituata a riconoscere meglio i gatti. La qualità e la quantità dei dati in un dataset sono cruciali per lo sviluppo di modelli di IA.

I sistemi di IA possono presentare distorsioni se i dataset su cui vengono allenati non sono rappresentativi. Tali distorsioni vengono dette *bias*.

Bias: questo termine si riferisce a pregiudizi o inclinazioni sistematiche nei dati, negli algoritmi, o nel processo decisionale dei modelli di IA. Tali distorsioni possono portare a risultati ingiusti, discriminatori o non accurati. Il *bias* può essere causato da un dataset se i dati non rappresentano

in modo equo tutte le variabili o i gruppi di interesse. Questo può accadere per vari motivi, come la raccolta di dati in contesti non diversificati o l'utilizzo di fonti di dati che riflettono pregiudizi preesistenti nella società.

Essere consapevoli che i sistemi di IA possono essere affetti da *bias* è fondamentale quando si usano questi strumenti. Non è corretto pensare che una macchina sia imparziale o infallibile solo perché è una macchina, giacché tecnologie come quelle IA sono programmate e attingono dalla nostra società, amplificandone le caratteristiche. La presenza di un *bias* nei dati può portare a risultati errati o non accurati. Per esempio, un modello di testo allenato principalmente sull'inglese (come la gran parte di quelli attualmente disponibili) potrebbe avere difficoltà o commettere errori in lingua italiana.

Per gli insegnanti, la consapevolezza del *bias* nei sistemi di intelligenza artificiale offre un'opportunità di educazione e riflessione critica. È fondamentale far capire che i bias possono influenzare non solo la tecnologia che utilizziamo, ma anche il modo in cui percepiamo e interagiamo con il mondo. Gli insegnanti possono guidare gli studenti attraverso le complessità dei sistemi AI, spiegando l'importanza di una programmazione e un uso etico della tecnologia.

Sul tema dei *bias* è possibile trovare molti *case study*, come il caso di ProPublica "Machine Bias" o il caso di uno dei primi sistemi di recruitment di Amazon, che risultò discriminatorio verso le donne.

Per approfondire

Live streaming:

<https://www.youtube.com/watch?v=687OD0FQPaE>

Documenti e case study:

<https://community.omniavis.it/uploads/short-url/mCbuKUvCJErGh6BeuOn1hvSOuPs.pdf>

<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

<https://www.bbc.com/news/technology-45809919>