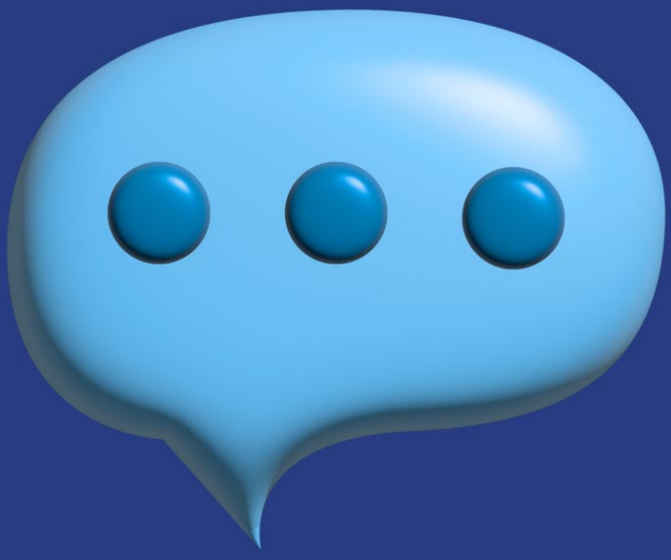
GIUSEPPE LEVI

Strategie di prompting

Una guida pratica all'uso degli LLM per studenti e insegnanti



CLUEB

Strategie di prompting

Una guida pratica all'uso degli LLM per studenti e insegnanti

Giuseppe Levi

CLUEB

Strumenti

ISBN 978-88-491-5820-5

Clueb è un marchio di Casa editrice prof. Riccardo Pàtron editore & C.
Via Marsala, 31 – 40126 Bologna
Info@clueb.it – www.clueb.it
Per informazioni sul copyright e il catalogo consultare www.clueb.it.

Indice

1

Di che cosa stiamo parlando

TL;DR

Questo capitolo esplora il significato del termine "Intelligenza Artificiale", analizzando la differenza tra l'intelligenza umana e quella simulata dalle macchine. Si evidenzia come l'AI imiti le capacità umane senza comprenderle veramente, attraverso algoritmi e set di dati complessi. Viene discusso l'ampio spettro di applicazioni dell'AI, dalla visione artificiale alla robotica, con particolare attenzione ai Large Language Models (LLM). Il capitolo introduce inoltre l'importanza di comprendere e ottimizzare l'uso di questi strumenti, in particolare in ambito universitario. Il capitolo si conclude illustrando il contenuto degli altri capitoli.

Il termine "Intelligenza Artificiale" è una traduzione letterale dall'inglese "Artificial Intelligence" (AI), ma questa semplice trasposizione linguistica nasconde sfumature significative che meritano un'analisi approfondita. In inglese, la parola "intelligence" racchiude in sé una molteplicità di significati fondamentali. Non si riferisce solo all'intelligenza come comunemente intesa, ma abbraccia anche le capacità di apprendere, elaborare e acquisire informazioni, nonché l'informazione stessa. Questa ricchezza semantica si perde parzialmente nella traduzione italiana. D'altra parte, il termine "Artificial" porta con sé tre connotazioni principali:

1. Creato dall'uomo
2. Non naturale
3. Non reale

Combinando questi elementi, l'Intelligenza Artificiale si configura come la creazione di un artefatto capace di acquisire, imparare ed elaborare informazioni, deducendone di nuove. Tuttavia, la definizione stessa sottintende che l'oggetto creato imita le funzioni umane in modo non reale o autentico. L'attribuzione di qualità umane alla macchina è quindi frutto di una **proiezione**. Questa proiezione nasce dalla nostra tendenza a riconoscere pattern familiari e a proiettare la nostra identità su oggetti esterni. Ciò che percepiamo come "intelligenza" in questi sistemi è un comportamento risultante da

complessi algoritmi e vasti set di dati, non di una vera coscienza o comprensione nel senso umano del termine. Possiamo quindi porci alcune domande cruciali:

- Fino a che punto possiamo parlare di vera intelligenza nelle macchine?
- Quali sono i limiti e le potenzialità di questa tecnologia?
- In quali campi di applicazione potrà essere usata?
- E come influenzerà il nostro futuro e la nostra comprensione dell'intelligenza stessa?

Per cominciare a rispondere a queste domande dobbiamo meglio comprendere l'estensione della materia trattata.

1.1. AI: un campo molto vasto

È cruciale comprendere che l'Intelligenza Artificiale è un campo estremamente vasto e diversificato, che va ben oltre la percezione comune attuale. Mentre il pubblico generale tende ad associare l'AI principalmente ai Large Language Models (LLM), come ChatGPT, il campo comprende in realtà una moltitudine di ambiti e applicazioni, ciascuno con le proprie sfide e potenzialità uniche. Per esempio, in modo non esaustivo, possiamo citare:

- Visione artificiale: questo settore si concentra non solo sull'acquisizione di immagini, ma sulla capacità di interpretarle e trarne significato, similmente a come fanno l'occhio umano e la corteccia visiva. La prima applicazione si ebbe anni or sono con la creazione di reti neurali in grado di decodificare i codici di avviamento postale scritti a mano, ora le applicazioni spaziano dal riconoscimento facciale alla guida autonoma.
- Elaborazione del Linguaggio Naturale:oltre ai LLM, questo campo include la traduzione automatica, l'analisi del sentimento, e la comprensione contestuale del linguaggio umano. Questo campo esiste da almeno 50 anni e fonda le sue radici nella linguistica [21, 30, 35].
- Classificazione e riconoscimento audio: sistemi in grado di identificare e classificare suoni, voci, musica, con applicazioni che vanno dall'assistenza vocale, alla diagnostica medica basata su suoni corporei fino all'identificazione delle armi da fuoco [144].
- Generazione di contenuti (AIGC): AI capaci di creare testi, immagini, musica e persino video, aprendo non solo nuove frontiere nella creatività assistita da computer, ma anche permettendo di creare dati sintetici per testare e allenare dei sistemi dedicati all'analisi scientifica [13, 129].
- Robotica: l'integrazione dell'AI nei sistemi fisici, permettendo ai robot di interagire con l'ambiente in modi sempre più sofisticati.
- Apprendimento per rinforzo: sistemi che imparano attraverso tentativi ed errori, ottimizzando le loro prestazioni nel tempo.
- Sistemi esperti: AI specializzate in domini specifici, capaci di emulare il processo decisionale di esperti umani in campi come la medicina o la finanza.

Questa diversità sottolinea come l'AI non sia un monolite, ma un ecosistema di tecnologie in rapida evoluzione, ciascuna con le proprie peculiarità e implicazioni etiche e sociali. La comprensione di questa vastità è essenziale per una discussione informata sull'AI, le sue potenzialità e i suoi limiti. Mentre i recenti progressi nei modelli linguistici hanno catturato l'attenzione del pubblico, è importante ricordare che rappresentano solo una frazione delle capacità e delle applicazioni dell'Intelligenza Artificiale nel suo complesso.

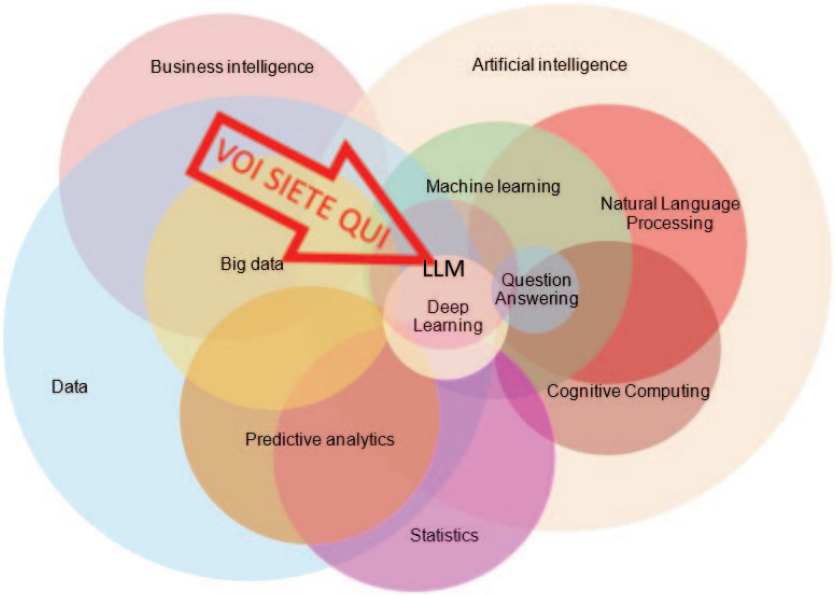


Figura 1.1: Una mappa di Venn dei diversi campi dell'Intelligenza Artificiale. Gli LLM sono collegati in modo interdisciplinare con più campi.

1.2. I Large Language Models

In questo testo ci concentreremo principalmente su un tipo di AI che sta riscontrando un grande successo mediatico ovvero i Large Language Models. I Modelli Linguistici di Grandi Dimensioni (Large Language Models o LLM) rappresentano un'importante evoluzione nell'ambito dell'intelligenza artificiale e del trattamento automatico del linguaggio naturale (NLP, Natural Language Processing). Questi modelli, basati su architetture di reti neurali, in particolare i trasformatori (transformers), hanno dimostrato capacità straordinarie nella comprensione e generazione del linguaggio umano.

Gli LLM sono addestrati su enormi quantità di dati testuali, che permettono loro di apprendere una vasta gamma di conoscenze linguistiche e concettuali. L'applicazione di questi modelli spazia dalla traduzione automatica alla generazione di testi, dalla sintesi di informazioni alla conversazione con utenti in linguaggio naturale. La loro versatilità li rende strumenti potenti in diversi settori, dall'assistenza clienti all'analisi dei dati, dall'educazione alla ricerca scientifica [94]. Tra l'altro, con la capacità di leggere le immagini *tokenizzandole* ovvero dividendole in componenti significative trattabili come fossero testi, gli LLM sono diventati anche multimodali, ovvero in grado di leggere il contenuto di una immagine (per esempio una lettera scritta a mano), dare un significato al contenuto ed elaborare questa informazione [14]. L'apparizione di capacità inaspettate per le quali gli LLM non erano stati espressamente programmati fu una sorpresa per gli stessi ricercatori [35]. Gli LLM che abbiamo a disposizione dimostrano di poter svolgere compiti considerati complessi e di trovare la soluzione a problemi logici, mostrandosi quindi come possibili potenti alleati in molti lavori ma anche per loro vale il cosiddetto *Moravec's*

Paradox, noto fin dagli anni '90 del secolo scorso, per cui funzioni che per gli umani sono semplici possono essere estremamente difficili per una macchina specie se involvono compiti visivi, senso-motori, pianificazione sequenziale o una profonda comprensione del linguaggio [23, 84]. Ne vedremo un esempio col problema visivo presentato in figura 5.3b a pagina 44 e col prompt 3.10 a pagina 27.

Da queste considerazioni nasce la necessità di studiare l'uso degli LLM in ambiente scolastico, universitario e professionale sia per dare delle indicazioni su come ottenere buoni risultati, sia per indicare quali siano i limiti attuali e le differenze di prestazioni dei diversi modelli. Come ogni testo di informatica anche questo nascerà già "obsoleto", ma forse potrà essere un buon punto di partenza e, per permettere al lettore di approfondire gli argomenti trattati, si è voluto dotarlo di una bibliografia, che è una parte fondamentale negli scritti scientifici [69], selezionando prevalentemente articoli da fonti aperte in modo che siano consultabili da tutti.

Dato che nella maggior parte dei casi la struttura interna degli LLM appare all'utente come inaccessibile, almeno nella parte iniziale del testo tratteremo questi come delle *scatole nere* concentrandoci solo sui dati in entrata e sulle risposte prodotte. Questo tipo di approccio non è limitativo ed è stato adottato da diversi autori sia per migliorare la qualità delle risposte prodotte che per provare la resistenza degli LLM a possibili attacchi informatici. [16, 65].

Il modo di stimolare un LLM a generare una risposta è un testo chiamato *prompt*. Può essere una domanda, un'istruzione, un inizio di frase o qualsiasi altro testo che guidi il modello nella generazione del contenuto desiderato. Nel capitolo 2 vedremo delle raccomandazioni generali riguardo a come scrivere prompt che generino risposte significative.

Approfondiremo l'argomento nel capitolo 3 dedicandoci in special modo ai prompt che ci possono aiutare in compiti di programmazione e riprendendo dalla letteratura l'idea dei *prompt patterns*, ovvero particolari strutture di prompt che sono usabili in diversi contesti applicativi.

Nel capitolo 4 affronteremo i temi delle cosiddette *allucinazioni* che sono contenuti completamente inventati o contraddittori generati dall'LLM. Questo tipo di contenuti è particolarmente pericoloso per uno studente perché pur presentandosi bene nella forma è invece errato nella sostanza. Quindi bisogna imparare sia come stimolare l'LLM in modo corretto per minimizzare questo rischio (a meno che non vi vogliate esplicitamente divertire con delle castronerie) e soprattutto essere in grado di controllare la correttezza di quanto prodotto.

La generazione di contenuti falsi o incoerenti può avvenire anche quando chiediamo agli LLM di produrre immagini. Nel capitplo 5 vedremo come poter stimolare un LLM con prompt multimediali formati da figure e testi ed esploreremo i limiti attuali nella accuratezza delle sia delle immagini prodotte, sia della capacità di ragionamento del sistema quando viene posto difronte a problemi visivi.

Il Jailbreaking, ovvero della costruzione di prompt in grado di indurre LLM a generare contenuti dannosi o offensivi è affrontato nel capitolo 6.

Due esempi di progetti completi, sono presentati nel capitolo 7 e per indicare al lettore come costruire progetti più complessi e avanzati, nel capitolo 8, vedremo come ospitare un proprio LLM personale su una propria macchina, ottenendo una maggiore privacy e un completo controllo sui parametri di funzionamento.

Il testo si chiude coi capitoli 9 e 10, mostrando come gli LLM possano essere trattati come sistemi programmabili e, dato che un libro universitario non è un vero libro universitario senza degli esercizi, proponendo degli esperimenti in difficoltà crescente.

La Bibliografia, con oltre 150 lavori scientifici e fonti citate, non vuol essere solo un supporto al testo ma anche uno strumento di studio per chi volesse approfondire l'argomento o fosse semplicemente curioso. Molti degli articoli sono in inglese ma usando strumenti come NotebookLM di Google, di cui diamo l'indirizzo nel prossimo paragrafo, potrete facilmente accedere ai loro contenuti che, nella maggioranza dei casi, sono scaricabili liberamente cercandoli con Google Scholar.

1.3. Strumenti Utili

Per scelta presentiamo e consigliamo solo strumenti che abbiano almeno una versione, anche se con limitazioni, gratuita o siano completamente gratuiti. I principali LLM che potrete usare per fare esperimenti sono:

1. ChatGPT: `https://chatgpt.com/`
2. Copilot: `https://copilot.cloud.microsoft/`
3. Claude: `https://claude.ai/`
4. Gemini: `https://gemini.google.com/`
5. Mistral: `https://chat.mistral.ai/chat` è un modello europeo.
6. Grok2: `https://x.com/i/grok`
7. Deepseek: `https://chat.deepseek.com/` è un modello opensource cinese.
8. QwenLM AI: `https://chat.qwenlm.ai/` permette di usare i modelli della cinese Alibaba.

Alcuni siti vi permettono funzioni particolari:

9. DuckDuckGo AI Chat: `https://duck.ai` In versione beta vi permette di scegliere l'LLM e (dice) di mantenere il vostro anonimato.
10. Groq: `https://groq.com/` è gratuito e vi permette di scegliere fra numerosi modelli Open Source.
11. Hugging Face: `https://huggingface.co/chat/` Hugging Face è un immenso repository di modelli Open Source e di dati. Ne parleremo nel capitolo8. Permette di creare degli agenti personalizzati gratis.
12. Pi AI: `https://pi.ai/discover` Cerca di avere una interazione calda e piena di emozioni verso l'utente.

Ma vi sono molti altri strumenti e siti che vi permettono di sperimentare e mettere a confronto fra loro diversi modelli gratuitamente. Fra i tanti citiamo:

13. Chatbot Arena: `https://lmarena.ai/`

Di particolare importanza per lo studio sono i sistemi che permettono di caricare documenti e ottenere risposte basate su questi. Ne segnaliamo due tralasciandone altre che non forniscono opzioni gratuite:

14. NotebookLM: `https://notebooklm.google.com/` Attualmente il miglior assistente alla ricerca e allo studio. Gratuito. Permette di creare dei blocchi di appunti in cui riassumere fino 50 files (pdf ed altri formati) creando una guida lo studio, facendo domande e generando (per ora solo in inglese) un riassunto audio in forma di podcast.
15. Perplexity AI: `https://www.perplexity.ai/` Permette di caricare e riassumere do-