

# **PRINCIPI DI FUNZIONAMENTO DEI LARGE LANGUAGE MODELS (LLM)**

## SOMMARIO

LLM .....	2
COME FUNZIONANO.....	2
STRUTTURA INTERNA.....	2
IL PROCESSO DI ADDESTRAMENTO.....	2
LLM GENERICO VS LLM AZIENDALE.....	3

## LLM

I LLM (Large Language Models) sono modelli di intelligenza artificiale progettati per comprendere e generare linguaggio umano. Immagina di avere un assistente virtuale che ha letto miliardi di parole: libri, articoli, siti web, manuali tecnici... tutto ciò che serve per "capire" come si costruisce una frase, come si fa una domanda, come si risponde in modo coerente

## COME FUNZIONANO

I LLM ricevono un input (una domanda, una frase, un comando) e generano un output coerente. Non hanno una coscienza o opinioni proprie, ma statisticamente scelgono la risposta più probabile, basandosi su tutto ciò che hanno "letto" durante l'addestramento. Alla radice, un LLM è un sistema predittivo: dato un pezzo di testo, prova a indovinare quale parola (o parte di parola) viene dopo. Questa abilità, estesa su larga scala, diventa molto potente e si basa su pattern linguistici appresi leggendo miliardi di frasi.

## STRUTTURA INTERNA

Il cervello umano ha neuroni. I LLM hanno reti neurali artificiali, ispirate a quelli biologici. Questi modelli sono composti da strati (layer) con tantissimi "nodi" connessi tra loro. Ogni nodo elabora numeri e passa il risultato agli altri. Ogni connessione ha un "peso" che viene addestrato per produrre buone risposte. Nel caso dei LLM moderni, queste reti hanno decine o centinaia di strati e miliardi di parametri.

## IL PROCESSO DI ADDESTRAMENTO

Durante l'addestramento, il modello:

- legge una frase con una parola nascosta

### **Oriens Consulting S.r.l. a socio unico**

Via Zamenhof 200, Vicenza 36100

+39 0444 1834081 - C.F. e P. IVA: 03801360243

info@oriens.consulting - oriens.consulting

- cerca di indovinare la parola giusta
- confronta la sua previsione con la parola reale
- aggiusta i suoi parametri per sbagliare meno la volta dopo

I LLM come GPT si basano su un'architettura chiamata Transformer, introdotta da Google nel 2017. Il cuore del Transformer è un meccanismo chiamato "attenzione" (attention), che permette al modello di focalizzarsi sulle parti più rilevanti del contesto di una frase.

Quando usi un LLM, gli dai un prompt: un testo di input. Il modello lo legge, costruisce una rappresentazione interna del significato, e poi genera la continuazione una parola alla volta. Il prompting è un'arte a sé: con prompt diversi puoi guidare il modello a fare riassunti, generare codice, simulare ruoli, tradurre, ecc. Una volta addestrato, il modello non impara più automaticamente. Ma può essere ottimizzato con tecniche di fine-tuning, per specializzarlo in un ambito (es. medicina, finanza, diritto), oppure istruito con il Reinforcement Learning from Human Feedback (RLHF), in cui persone valutano le risposte per renderlo più utile e meno pericoloso.

## **LLM GENERICO VS LLM AZIENDALE**

### **LLM Generico**

Addestrato su dati pubblici e generalisti (Wikipedia, libri, siti web, ecc.)

Sa un po' di tutto, ma non è specializzato in nulla

Perfetto per compiti ampi: scrivere testi, rispondere a domande generiche, fare brainstorming

Non conosce il tuo contesto aziendale specifico

### **LLM Aziendale**

Basato su un LLM generico, ma raffinato (con fine-tuning o RAG) usando documenti aziendali

Usa il linguaggio e i concetti tipici della tua azienda o settore

Più preciso e rilevante per i compiti interni, usando policy, report, database aziendali

Specializzato in modelli, processi, dati e linguaggio aziendale

#### **Oriens Consulting S.r.l. a socio unico**

Via Zamenhof 200, Vicenza 36100  
 +39 0444 1834081 - C.F. e P. IVA: 03801360243  
 info@oriens.consulting - oriens.consulting

**Oriens Consulting S.r.l. a socio unico**

Via Zamenhof 200, Vicenza 36100

+39 0444 1834081 – C.F. e P. IVA: 03801360243

info@oriens.consulting – oriens.consulting