



LLM (Large Language Models): perché ne parlano tutti?

“Nel panorama dell’Intelligenza Artificiale, il trattamento del linguaggio naturale (NLP) è sempre stato uno dei campi più impegnativi e affascinanti. Negli ultimi anni, un’innovazione senza precedenti ha scosso le fondamenta di questo settore: l’avvento dei Large Language Model (LLM). Questi modelli, alimentati da algoritmi di apprendimento chiamati deep learning e addestrati su enormi quantità di testi, hanno dimostrato una comprensione sorprendente e una capacità di generazione del linguaggio umano.”

Questa risposta ce l’ha fornita proprio ChatGPT!

MA COSA SONO ESATTAMENTE I LARGE LANGUAGE MODEL?

Gli LLM, come detto anche da ChatGPT, sono modelli di AI, per la precisione AI generativa, in quanto creano del contenuto. Infatti, questi modelli comprendono, riassumono, e generano delle risposte. I primi modelli sono stati elaborati al MIT e risalgono agli anni 60, ma, per la mancanza di risorse tra cui quelle computazionali, non si sono ottenuti grandi risultati. Oggi, le risorse sono maggiori e conseguentemente la performance è migliorata e sono tra i modelli in maggiore evoluzione al momento.

MA COME FUNZIONANO ESATTAMENTE?

Il punto focale degli LLM è come questi comprendono le parole. Inizialmente venivano utilizzate tabelle numeriche semplici in cui ogni parola veniva rappresentata nella suddetta tabella, metodo che però si è rivelato obsoleto, in quanto non venivano correttamente riconosciuti i collegamenti e i pattern connessi ai termini, come per esempio contrari o sinonimi. Oggi invece, gli LLM si basano sui cosiddetti modelli transformers, nonché reti neurali che imparano il contesto e il significato partendo da una sequenza di dati che nel nostro caso potrebbe essere una frase. Infatti, vengono usati i word embeddings, ossia le rappresentazioni delle parole o frasi (da cui partiamo) che vengono trasformate in vettori numerici in uno spazio multidimensionale. In sostanza si tratta di una rappresentazione numerica dei token/parole che cerca di cogliere il contesto della frase e la relazione tra le stesse. Per questo, le parole simili tra di loro vengono poste con una distanza minore all'interno dello spazio.

PERCHÉ I LLM SONO COSÌ IMPORTANTI?

Nel 2021 sono stati ribattezzati dal centro di ricerca sull'AI di Stanford "Foundation models" proprio per la loro importanza nella ricerca sull'AI degli ultimi anni. Questi modelli devono la loro fama alla loro capacità di adattamento: un singolo modello riesce a tradurre, riassumere e rispondere a domande complesse, riesce a creare codici e a revisionare interi documenti, il tutto partendo sia da prompt piccoli che corposi, e in più lingue! In alcuni casi, si parla, infatti, di modelli veramente giganti: **Claude** (il modello di Anthropic) per esempio, accetta prompt in input da 100k token (una parola o parte di una parola), dunque parliamo di articoli molto lunghi o addirittura libri. Mentre **GPT-3** si basa su almeno 175 miliardi di parametri. Inoltre, secondo il Consiglio Europeo sulla protezione dei dati (1), gli LLM sono un mezzo estremamente potente per la tutela dei dati sensibili, in quanto, soprattutto in testi lunghi, riescono a identificare e analizzare tutti quelli che potrebbero essere dati soggetti a privacy ed escluderli da determinati contesti, tra cui eventuali set di dati utilizzati per addestrare gli stessi modelli! Questo permette di evitare bias e di pulire anche i dati di partenza.

LE LIMITAZIONI DEI LLM

Ovviamente bisogna anche considerare che, come per ogni cosa, anche i LLM hanno delle limitazioni. Spesso possono "allucinare", ovvero restituire output sbagliati credendo siano corretti, il che diventa problematico nel momento in cui si ha a che fare con dati sensibili o magari informazioni sbagliate su determinati avvenimenti. Per tale motivo, nasce anche la necessità di avere dati di addestramento puliti in quanto risalire ai singoli dati distorti è pressoché impossibile o comunque molto difficile in quanto si tratta di modelli che si basano su milioni di parametri. In ogni caso, gli sviluppi di questa tecnologia sono molteplici e in rapidissima crescita e se sfruttati bene possono essere la svolta nella rivoluzione tecnologica che stiamo vivendo.

FONTI

- https://www.edps.europa.eu/data-protection/technology-monitoring/techsonar/large-language-models-llm_en

Rubrica a cura di **Generazione Stem**

L'AUTRICE

Inès El Gataa, Master's Degree in Data science and artificial intelligence alla SISSA di Trieste e contributor di Generazione Stem.