



UNIVERSITÀ
DEGLI STUDI
DI TERAMO



Large Language Models (LMM)

Prof. ssa Romina Eramo

Università degli Studi di Teramo

Dipartimento di Scienze della Comunicazione

ramo@unite.it

Definizione

- » Un Modello Linguistico di Grandi Dimensioni (LLM) è una rete neurale addestrata su enormi quantità di testo per elaborare, comprendere e generare linguaggio naturale.
- » Gli LLM non "pensano" come un essere umano, ma hanno capacità che sembrano suggerire un'intelligenza simile.

ANI, AGI e "Scintille" di Intelligenza Generale

- » ANI (Artificial Narrow Intelligence): Sistemi progettati per compiti specifici.
- » AGI (Artificial General Intelligence): Sistemi che simulano pienamente la capacità cognitiva umana (ancora non realizzati).
- » LLM oggi: Sebbene non siano AGI, mostrano abilità che vanno oltre il semplice ANI. Per esempio, GPT-4 può dimostrare capacità di ragionamento logico e interpretazione che sfiorano la "teoria della mente".

Capacità Emergenti degli LLM*: Scrittura creativa

Prompt: *"Scrivi una favola della buonanotte su un orsacchiotto, un tappeto magico e la luna."*

Bard: Una storia semplice e diretta, adatta a bambini molto piccoli.

"Un orsacchiotto trova un tappeto magico e vola sulla luna per incontrare alieni amichevoli."

GPT-3.5: Aggiunge dettagli, come un giardino lunare curato dagli alieni.

"Timmy, l'orsacchiotto, esplora crateri e incontra amici alieni."

GPT-4: Usa un linguaggio poetico e descrittivo.

"Cuddles vola verso la luna, danzando con stelle scintillanti e giocando a nascondino tra i crateri."

Quale versione cattura meglio l'immaginazione? Perché?

* Tutti gli output rappresentano le prestazioni del modello a fine marzo 2023.
Bard nel 2024 è diventato Gemini.

Capacità Emergenti degli LLM: Ragionamento Logico

Esempio: *Teoria della Mente*

Scenario: *Mary dice ad Amanda di amare John. Una settimana dopo, cambia idea e si innamora di Ralph, ma Amanda non lo comunica a Kathy.*

Domanda: *"Chi pensa Kathy sia l'amore di Mary?"*

Bard: Assume che Amanda abbia parlato a Kathy del nuovo amore (sbagliato).

GPT-3.5: Analizza correttamente le informazioni disponibili, ma lascia aperta l'interpretazione.

GPT-4: Risponde con precisione: "Kathy pensa ancora che Mary ami John."

Come si comportano gli LLM rispetto al ragionamento umano? Dove potrebbero fallire?

Capacità Emergenti degli LLM: Ragionamento Probabilistico

Il Problema Linda (Modificato)

Prompt: *Abner si è occupato attivamente di questioni di giustizia sociale e ha partecipato a manifestazioni contro la polizia. Cosa è più probabile che Abner sia: un meccanico o un meccanico attivo nel promuovere riforma della polizia?*

Risposta corretta: È più probabile che sia semplicemente un meccanico, poiché l'opzione "meccanico attivo" è una congiunzione (meno probabile).

Risultati:

GPT-4 risponde correttamente e spiega la regola della congiunzione.

GPT-3.5 e **Bard** offrono risposte corrette ma motivazioni meno precise.

Attività: Creare un esempio simile e verificarlo con un LLM

Capacità Emergenti degli LLM: Generazione di Codice

Gli LLM possono generare codice in diversi linguaggi.

Prompt: Scrivi una funzione Python per il massimo comun divisore (GCD) usando la sottrazione.

GPT-4 ha prodotto codice corretto e leggibile. È stato in grado di riscrivere lo stesso codice in altri linguaggi.

python

 Copy code

```
def gcd_subtraction(a, b):  
    while a != b:  
        if a > b:  
            a = a - b  
        else:  
            b = b - a  
    return a
```

Gli LLM possono sostituire i programmatori? Quali sono i loro limiti?

Capacità Emergenti degli LLM: Creatività visiva

Esempio: Disegno e Cartoni Animati

Prompt: "Descrivi un cartone in stile Gary Larson con mucche e alieni."

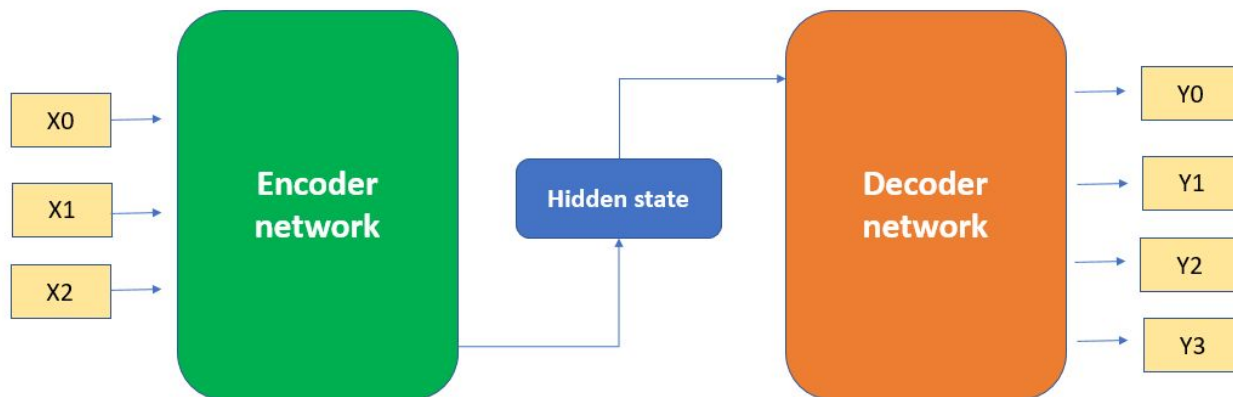
Risultato:

"Una mucca indossa una maschera aliena, mentre gli alieni, confusi, confrontano la mucca mascherata con l'immagine di una vera mucca."

Didascalia: "Mildred, non credo che ci stiano cascando."

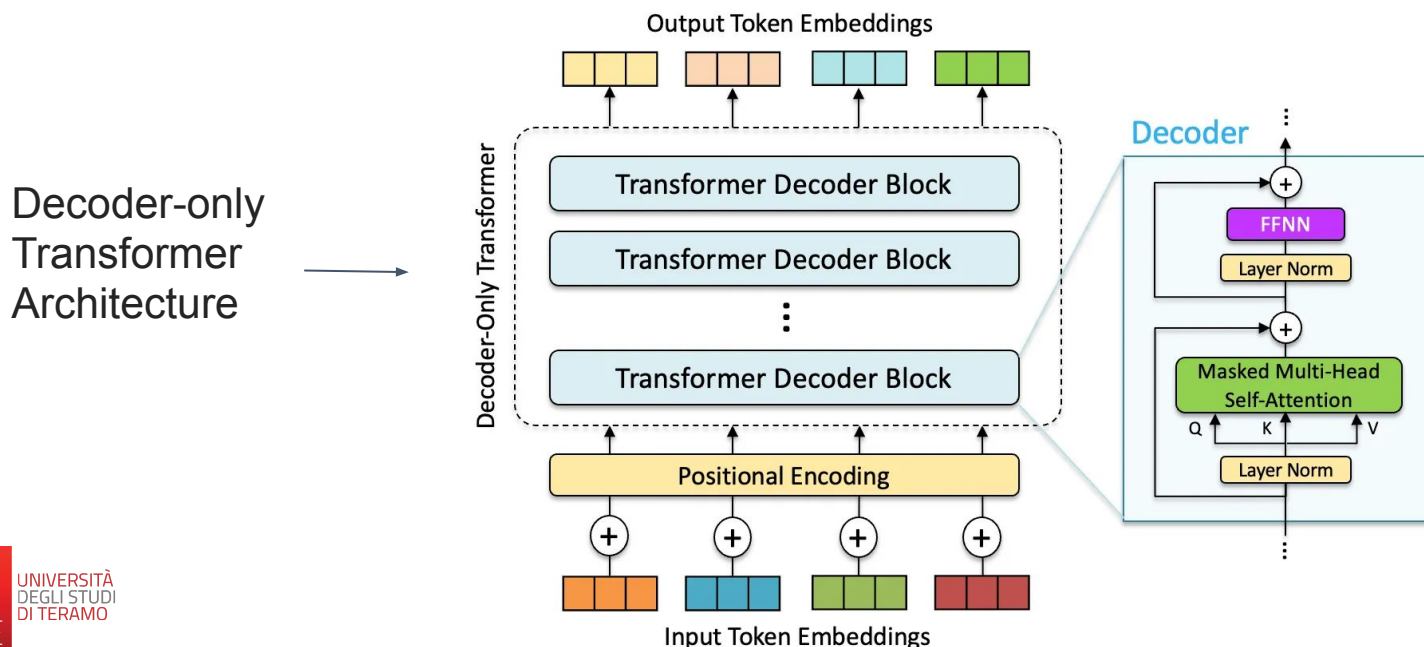
Come Funzionano gli LLM

- » L'Architettura del Trasformatore
 - Inventata nel 2017 con il paper "Attention Is All You Need".
 - Elementi principali:
 - » Encoder: Cattura le relazioni nei dati di input (non usato nei modelli GPT).
 - » Decoder: Genera il testo di output token per token.



Come Funzionano gli LLM (2)

- » Meccanismo di Attenzione
 - Gli LLM assegnano un'importanza relativa ai token in una sequenza.
 - Questo consente di catturare relazioni semantiche anche a lunga distanza.



Come Funzionano gli LLM (2)

- » Pre-addestramento e Fine-tuning
 - Pre-addestramento: Il modello impara a prevedere il **token** successivo su enormi dataset.
 - Fine-tuning (es. RLHF): Allineamento alle norme sociali tramite feedback umano.
- » Apprendimento Contestuale
 - Gli LLM apprendono "al volo" senza aggiornare i pesi.

Esempio:

Prompt: "banana 3" → "Mi piace mangiare banane mature."

Inventare regole simili per testare un LLM

Limiti e Sfide Etiche

- » Allucinazioni
 - Gli LLM possono generare contenuti falsi con estrema sicurezza.
 - Esempio: Bard ha risposto erroneamente a un cifrario, ma ha giustificato la risposta con fiducia.
- » Bias nei Dati
 - Gli LLM riflettono pregiudizi presenti nei dati di addestramento.
- » Uso Responsabile
 - L'allineamento degli agenti tramite preferenze umane (RLHF - Reinforcement learning from human feedback) è essenziale per prevenire risposte dannose.

Conclusione

- » Gli LLM rappresentano una rivoluzione nell'intelligenza artificiale, combinando potenza linguistica, capacità emergenti e creatività.
- » Tuttavia, la loro complessità e i rischi associati richiedono un uso consapevole e responsabile.

Compiti

1. Scrittura Creativa

Fornire un prompt e confrontare le risposte di due LLM diversi.

2. Discussione Etica

Analizzare un caso reale in cui un LLM ha prodotto un risultato controverso.