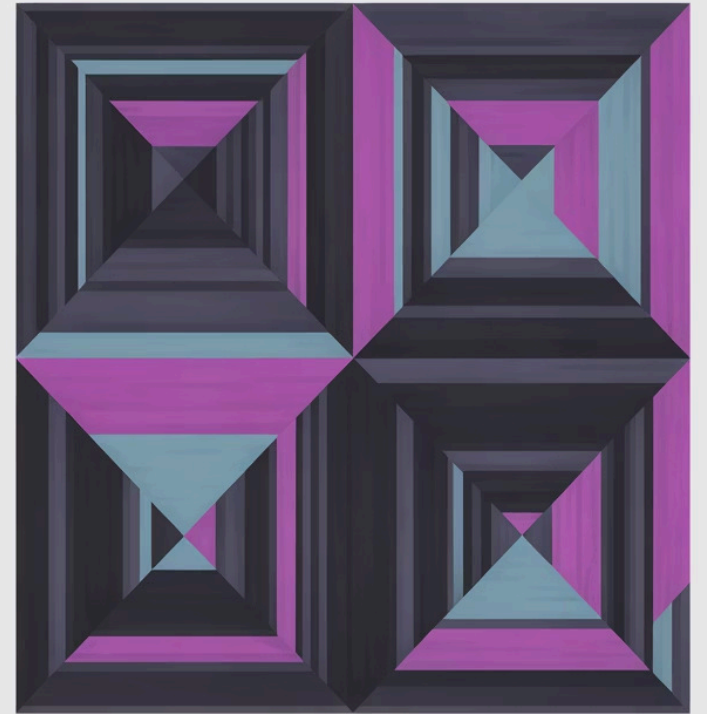
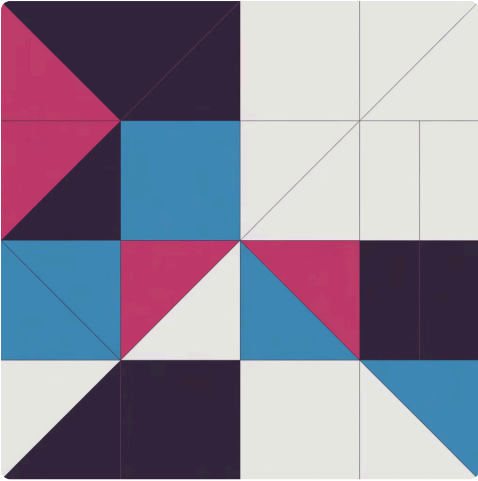


Gli LLM pensano?

Come funziona davvero il "cervello" digitale dell'IA

Il dibattito sull'intelligenza artificiale divide esperti e filosofi. Mentre alcuni negano capacità cognitive alle macchine, altri sostengono che i meccanismi degli LLM riproducano forme genuine di elaborazione intelligente del linguaggio e della conoscenza.





Utilizzando [Gamma.app](#) ho elaborato una sintesi grafica di un articolo del filosofo Riccardo Manzotti che si presta ad essere letto in parallelo al libro di Paolucci.

Aluisi Tosolini

Riccardo Manzotti

Ordinario di Filosofia Teoretica, IULM, Milano



 Agenda Digitale



Gli LLM pensano? Come funziona davvero il "cervello" ...

Analisi del dibattito sull'intelligenza artificiale: gli LLM pensano davvero? Confronto tra meccanismi cerebrali e algoritmi...

Pubblicato il 3 ottobre 2025

Il contesto e il grande dibattito

Intelligenza Artificiale

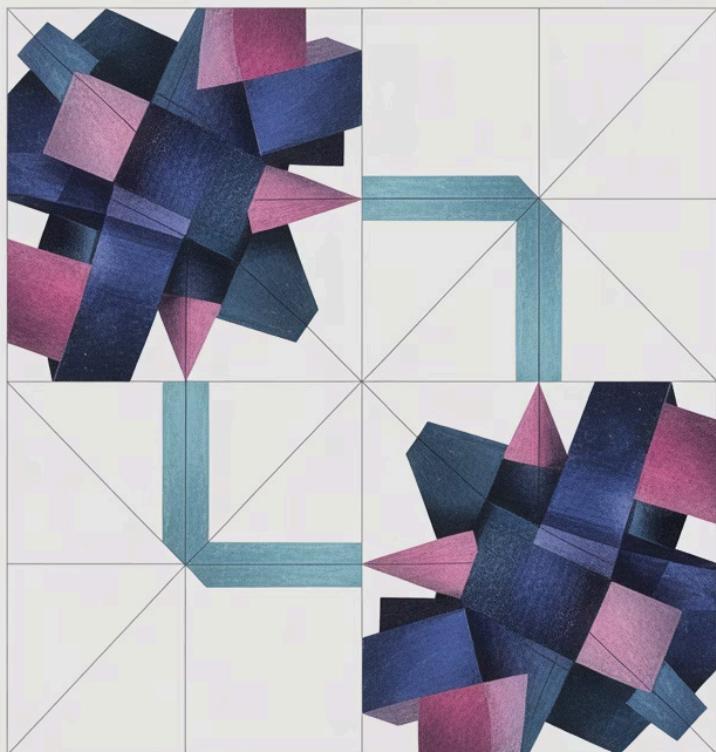
I modelli linguistici di grandi dimensioni hanno conquistato il dominio pubblico

Il Grande Dibattito

Gli LLM sono una forma di intelligenza o solo imitazione matematica?

Alleanza Improbabile

Filosofi e informatici uniti nel negare l'intelligenza alle macchine



La Questione fondamentale

Da quando, nel 2022, gli **LLM (modelli di linguaggio di grandi dimensioni)** hanno conquistato il dominio linguistico pubblico, un dibattito—spesso molto acceso—non ha cessato di infuriare: **gli LLM sono una forma di intelligenza? Gli LLM pensano?**

I Negatori

Sostengono che le macchine sono solo un'imitazione realizzata attraverso processi matematici, senza vera intelligenza.

I Next-Tokenisti

Informatici ed esperti di IA che definiscono gli LLM come **pappagalli stocastici** che si limitano a indovinare la parola successiva.

- ❑ **Il Benaltrismo dell'IA:** L'IA vince a scacchi? *Il pensiero è ben altro!* L'IA traduce come un essere umano? *Il pensiero è ben altro!* L'IA trova collegamenti filosofici complessi? Ben altro! Ben altro!

Indice degli argomenti

01

Il fenomeno del bersaglio mobile

Come i criteri di intelligenza cambiano quando le macchine li superano

02

Oltre le motivazioni filosofiche

Un approccio scientifico alla questione dell'intelligenza artificiale

03

L'inesistenza del pensiero separato

Il pensiero non è un'essenza invisibile ma capacità adattiva

04

L'errore dei next-tokenisti

Il confronto scorretto tra meccanismi di base e comportamenti globali

05

Il cuore tecnologico

L'algoritmo Transformer e il meccanismo dell'attenzione

06

Analogie cerebrali

Neuroni e tensori a confronto nella costruzione dell'intelligenza

07

Multi-head attention

L'intelligenza gerarchica e le strutture multiple di elaborazione

08

Le critiche degli esperti

Floridi, Quattrococchi e la questione della vera intelligenza

09

La lezione di Wittgenstein

Contro il pensiero magico e l'idea occulta della mente

Il Bersaglio mobile dell'intelligenza

Nel caso dell'IA abbiamo assistito a un classico caso di **bersaglio mobile**, come hanno recentemente sostenuto Azeem Azhar and Nathan Warren sul *Washington Post*.

Tutti i criteri che, nell'essere umano e negli animali sono considerati prove di intelligenza, nel momento in cui vengono raggiunti e superati da una macchina, vengono dichiarati insufficienti.



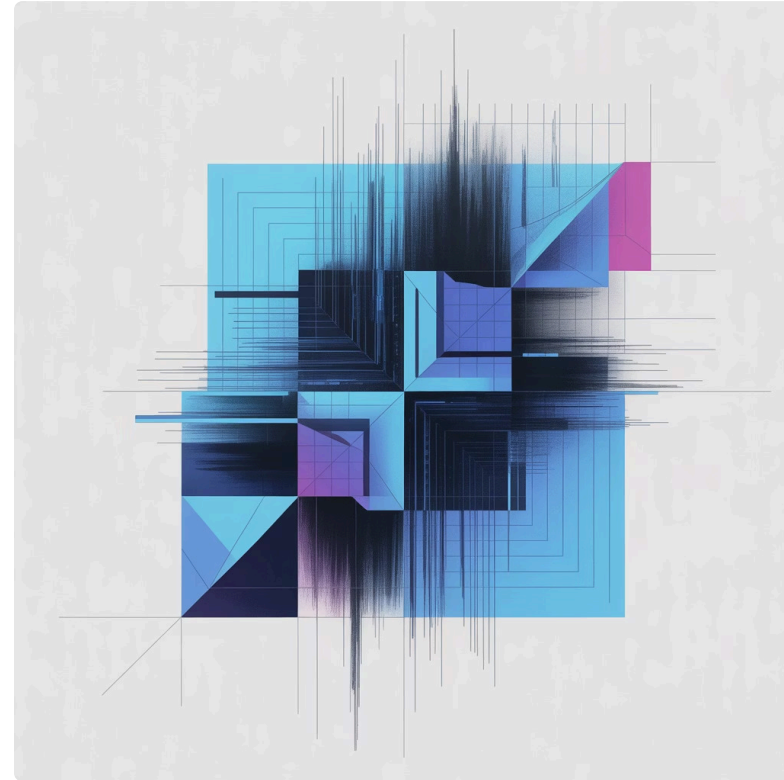
Finché era impossibile per una macchina, sembrava una soglia significativa, ma quando viene superato, ecco che non vale più nulla.

Oltre le motivazioni filosofiche

L'Obiettivo di Questo Articolo

Non dimostrerò che gli LLM pensano come gli esseri umani o che il percorso cognitivo è identico. Sarei assolutamente fuori strada.

Tuttavia, voglio chiarire che **le motivazioni informatico-scientifiche utilizzate dagli scettici sono infondate** e che l'intelligenza non è una qualità esclusiva degli esseri umani.



- ❑ Non considererò qui le motivazioni dei filosofi che presuppongono un ruolo metafisicamente privilegiato per gli esseri umani, ma soltanto i ragionamenti di chi—informatici ed esperti di IA—sostengono ci siano meccanismi speciali per «pensare veramente».

L'inesistenza del pensiero interno

Il pensiero non è un'essenza invisibile

Prima di tutto, l'obiezione fondamentale è che **non esiste un pensiero interno**. Se due sistemi producono lo stesso tipo di risposta, è inutile andare a guardare dentro per vedere se uno c'è arrivato pensando e l'altro c'è arrivato facendo calcoli. Quello che conta è l'esito finale.

Il Pensiero Come Capacità

Il pensiero non è un'essenza invisibile all'interno di un sistema cognitivo, ma è un modo per descrivere la **capacità adattiva** di certi sistemi.

Tra i Neuroni

Tra i neuroni non ci sono pensieri. I neuroni fanno quello che devono fare, organizzando stimoli in entrata e uscita e producendo il giusto comportamento.

Nessuna Cosa Aggiuntiva

In mezzo non c'è una cosa aggiuntiva e invisibile che chiamiamo pensiero. Non ci sono processi che producano il pensiero all'interno di un sistema cognitivo.

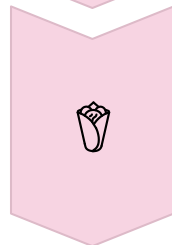
L'Errore concettuale dei Next-Tokenisti

Gli scettici (il mio termine favorito è **next-tokenisti**) fanno un errore concettuale ed empirico comparando il meccanismo di base dell'LLM (la probabilità condizionata) con il comportamento globale dell'essere umano.



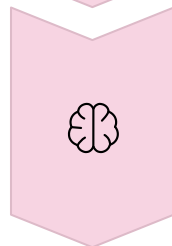
Confronto scorretto

Comparare il meccanismo di base dell'LLM con il comportamento globale dell'umano è fuorviante



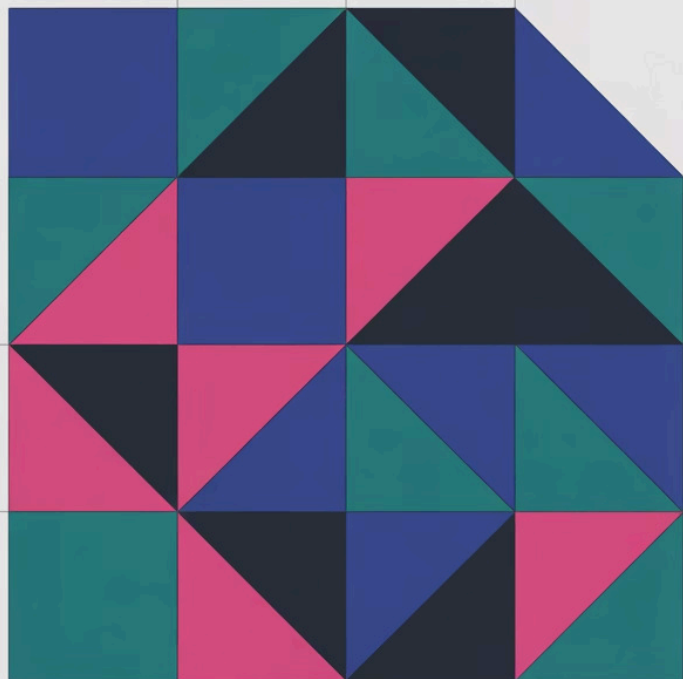
L'Errore simmetrico

Sarebbe come comparare il meccanismo di base del cervello (apprendimento hebbiano) con il comportamento globale dell'LLM



Conclusione assurda

Si concluderebbe che noi siamo meno intelligenti di un LLM, perché un neurone non sa esprimersi linguisticamente



Il Confronto corretto

Meccanismi di base

Il singolo neurone non è affatto più intelligente, intenzionale, cosciente, mentale, volitivo del singolo tensore di un LLM.

- Neuroni: soglia non lineare, apprendimento hebbiano
- LLM: tensori bayesiani, probabilità condizionata

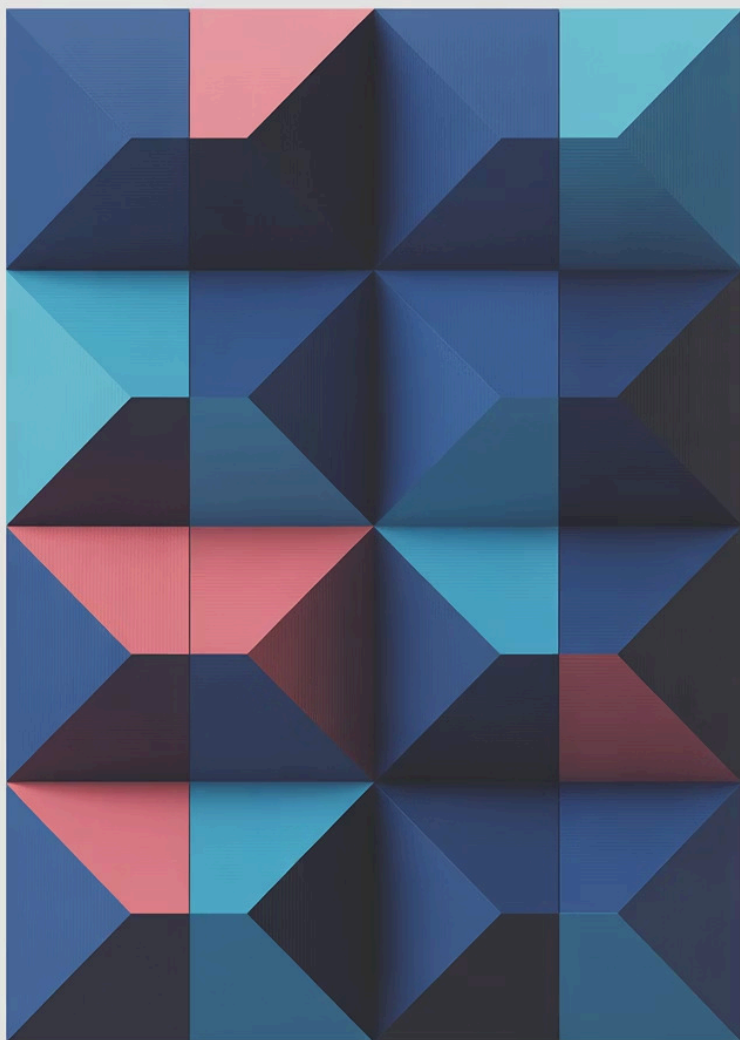
Comportamenti globali

La comprensione del testo non avviene a livello dei componenti di base, ma è frutto della **dinamica globale** tra la rete neurale e l'ambiente.

- Umani: interazione con ambiente fisico e sociale
- LLM: dataset di addestramento e prompt

□ Un LLM a livello di dinamica globale con l'ambiente esibisce un comportamento linguistico che—al netto di affermazioni metafisiche—permette di confrontare esseri umani e IA generative.

In questa prospettiva neutra, recentemente modellata dal cosiddetto **sistema zero**, è interessante considerare una progressiva integrazione tra intelligenza umana e macchinica, come espresso dal **principio di parità cognitiva**.



Il cuore tecnologico: l'algoritmo transformer

Anche considerando in maggiore dettaglio il funzionamento interno di un LLM non ci sono motivi per negargli cittadinanza cognitiva. Al suo cuore c'è l'**algoritmo Transformer** in qualche sua recente incarnazione.

Quelli che sostengono che gli LLM non fanno altro che indovinare la parola successiva, semplificano eccessivamente il loro meccanismo interno, limitandosi a fare un corto circuito tra il meccanismo di base e il comportamento complessivo del sistema.



Meccanismo di Base

Statistica bayesiana e probabilità condizionata



Livelli Intermedi

Tra il basso livello e l'output finale deve succedere qualcosa in mezzo

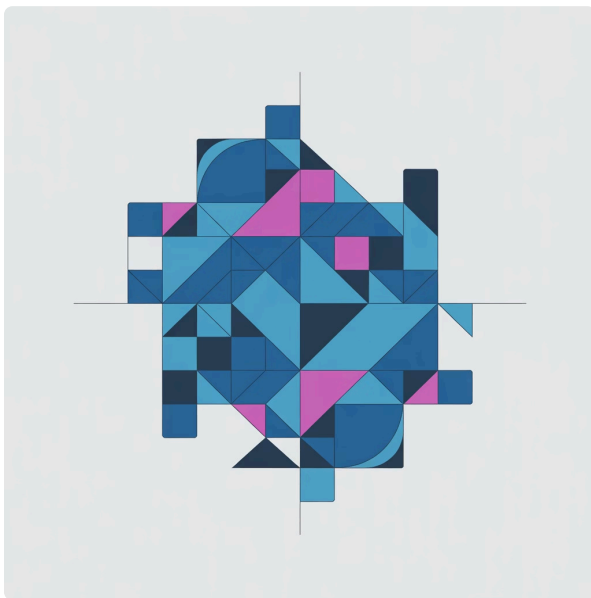


Comportamento Globale

Generazione di un flusso di token a partire da un prompt

Analogie cerebrali: neuroni e tensori

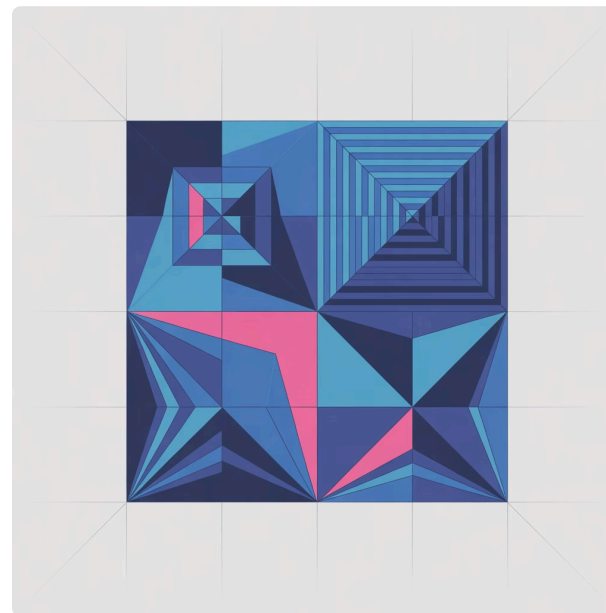
Questo modo di ragionare mostra tutti i suoi limiti se applicato al cervello. In fondo, si potrebbe dire, un **neurone** non è altro che una sommatoria a soglia non lineare con apprendimento hebbiano.



Il Cervello

Globalmente, il cervello esibisce un comportamento linguistico che qualcuno potrebbe ridurre a mettere una parola dietro l'altra.

È qui che l'argomento del pappagallo stocastico perde la sua forza.



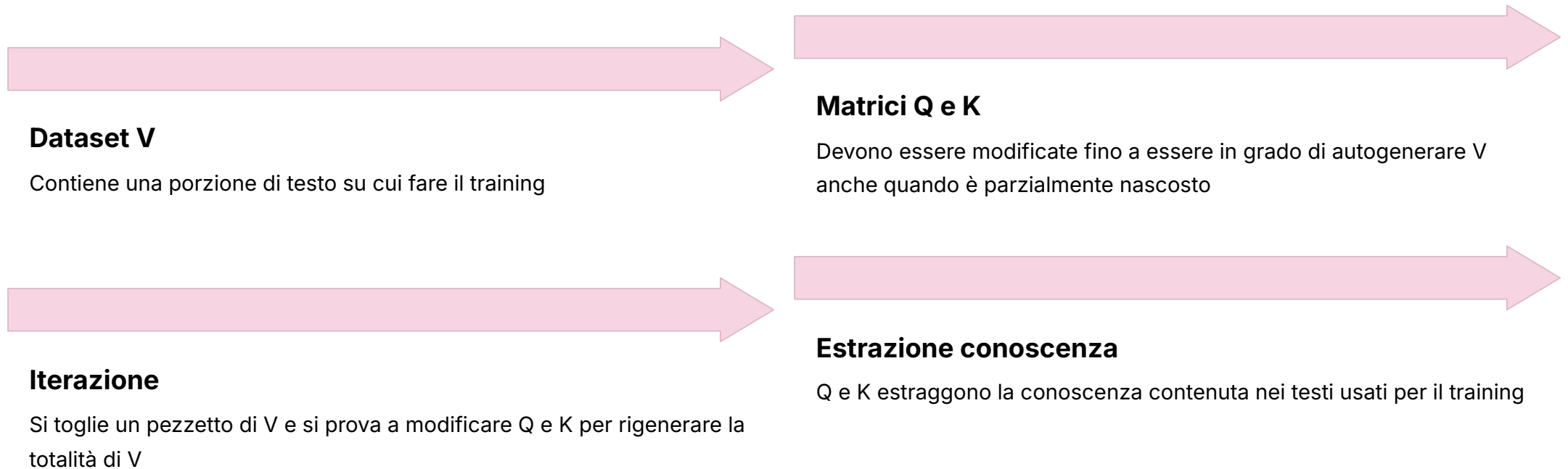
L'LLM

Per arrivare a prevedere quella famosa "parola successiva" o next token deve succedere qualcosa in mezzo.

Anatomia dell'attenzione

L'Equazione fondamentale

Entriamo nel dettaglio dell'**algoritmo Transformer**. Al suo cuore c'è l'equazione dell'attenzione che calcola le matrici Q e K sulla base del dataset V.

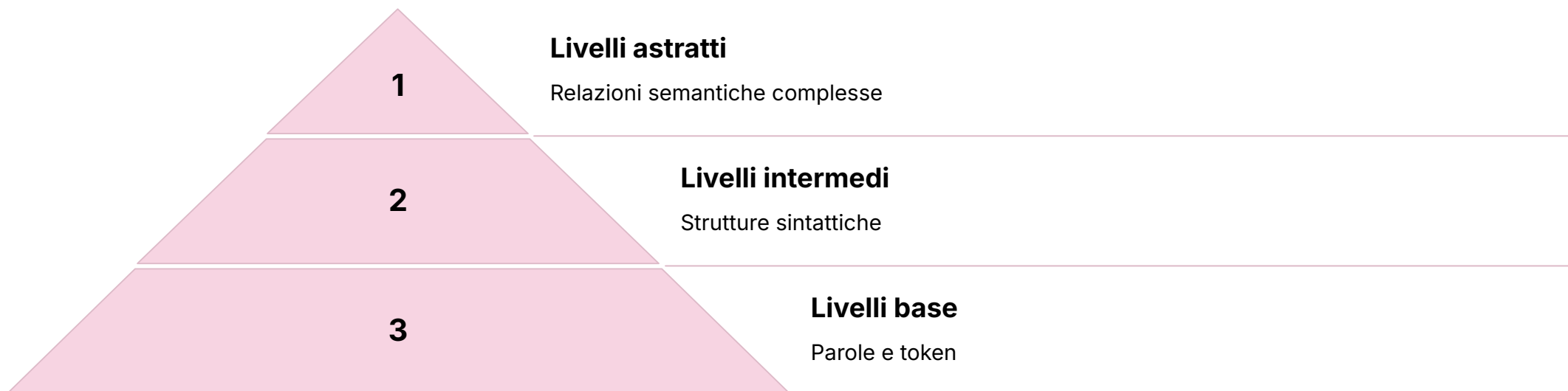


In questo modo Q e K costruiscono un **modello statistico** che incarna la conoscenza che nei testi si è tradotta in legami probabilistici tra le parole, diventando in grado di generare nuovi testi che riflettono la **struttura sintattico-semantic**a e, perché no, epistemica del testo di training.

Multi-Head Attention: l'intelligenza gerarchica

Ma non finisce qui. Un LLM non si limita a cercare legami tra parole, ma anche tra **strutture più astratte e più globali**.

I vari ChatGPT, Grok, Claude, DeepSeek o Gemini applicano l'equazione dell'attenzione sfruttando una **struttura gerarchica iterativa** in grado di creare collegamenti non solo tra parole, ma anche tra strutture bayesiane a molteplici livelli di astrazione.



- ❑ A livello astratto, causalità e probabilità condizionata sono cose diverse, ma è un dato di fatto che la **statistica bayesiana** è una delle migliori approssimazioni matematiche della causalità tra fenomeni. Un valore di $P(A|B)$ pari a 1 è un ottimo indicatore del fatto che B causi A.

Strutture multiple di elaborazione

L'amplificazione dell'attenzione a livelli gerarchici multipli è implementato dal cosiddetto **multi-head**. Il multi-head consiste nel procedere per vie multiple (in senso massivo) all'interno del dataset.

Cos'è una "testa"

Ogni head ("testa") è un meccanismo di attenzione indipendente, cioè un piccolo blocco che prende l'input, lo proietta in uno spazio ridotto, calcola le matrici Q e K in quel contesto e produce un output.

Elaborazione parallela

Il dataset viene così percorso a **livelli diversi di astrazione** e di complementarità. In questo modo si estraggono **relazioni sintattiche e semantiche** a livelli molteplici.

Non è conoscenza linguistica questa? Non è, a tutti gli effetti, l'implementazione del livello **epistemico e semiotico**?

A parità di altre condizioni, non è, in fondo, quello che facciamo noi esseri umani quando leggiamo e costruiamo una rete di relazioni tra "concetti"? Che cosa sono i **concetti** se non strutture condizionali nell'immensa rete di input linguistici che riceviamo fin da piccoli?



Le Critiche degli Esperti

“

Luciano Floridi

«L'IA ha intelligenza zero. È una sciocchezza credere che l'IA possa diventare intelligente anche solo come un gatto, che capisca o possa un giorno capire.»

”

“

Walter Quattrociocchi

«Gli LLM non capiscono il mondo: predicano la parola successiva in una sequenza, sulla base di correlazioni osservate in enormi corpus testuali. È un'operazione statistica, non epistemica.»

”

Gli scettici lamentano che non ci sono strutture che corrispondano a meccanismi deduttivi, pensiero o altro. È la posizione perfetta dei **next-tokenisti!**

Ma è sicuro che il cervello faccia qualcosa di diverso? È sicuro che un sistema di milioni di parametri che catturano le relazioni epistemiche multiple come l'attenzione e il multi-head non catturi la conoscenza dal linguaggio?