

24. aprile

Conoscenze imprescindibili per ragionare sull'intelligenza artificiale

La cultura è il superfluo indispensabile.
Gaetano Salvemini

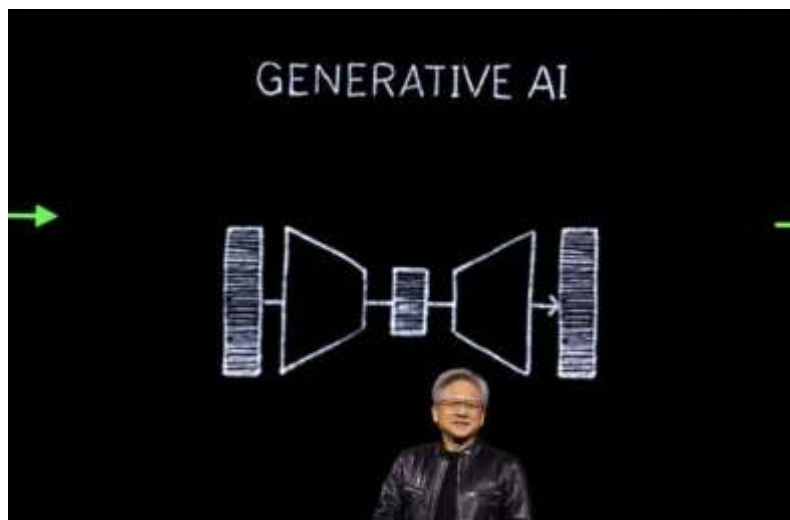
L'intelligenza artificiale generativa

L'intelligenza artificiale generativa è una categoria di intelligenza artificiale che utilizza i dati per creare contenuti originali. Al contrario, l'intelligenza artificiale classica potrebbe offrire solo previsioni basate su input di dati, non risposte nuove e uniche utilizzando l'apprendimento automatico.

Ma l'intelligenza artificiale generativa utilizza il "deep learning", una forma di apprendimento automatico che utilizza reti neurali artificiali (programmi software) simili al cervello umano, in modo che i computer possano eseguire analisi simili a quelle umane.

Tuttavia, l'intelligenza artificiale generativa non raccoglie risposte dal nulla. Genera risposte basate sui dati su cui è stato addestrato, che possono includere testo, video, audio e righe di codice. Immagina, diciamo, di svegliarti dal coma, bendato, e tutto ciò che riesci a ricordare sono 10 articoli di Wikipedia.

Tutte le tue conversazioni con un'altra persona su ciò che sai sono basate su quei 10 articoli di Wikipedia. È più o meno così, tranne per il fatto che l'intelligenza artificiale generativa utilizza milioni di articoli di questo tipo e molto altro ancora.



Jensen Huang, CEO di Nvidia

Cos'è un chatbot?

I chatbot AI sono programmi per computer che generano conversazioni simili a quelle umane con gli utenti, fornendo risposte uniche e originali alle loro domande.

I chatbot sono stati resi popolari dal ChatGPT di OpenAI e da allora ne hanno debuttati molti altri: Google Gemini, Microsoft CoPilot e Einstein di Salesforce guidano il gruppo, tra gli altri.

I chatbot non si limitano a generare risposte di testo: possono anche creare siti Web, creare visualizzazioni di dati, aiutare con la codifica, creare immagini e analizzare documenti.

A dire il vero, i chatbot basati sull'intelligenza artificiale non sono ancora infallibili: hanno già commesso molti errori .

Ma man mano che la tecnologia dell'intelligenza artificiale avanza rapidamente, aumenterà anche la qualità di questi chatbot.



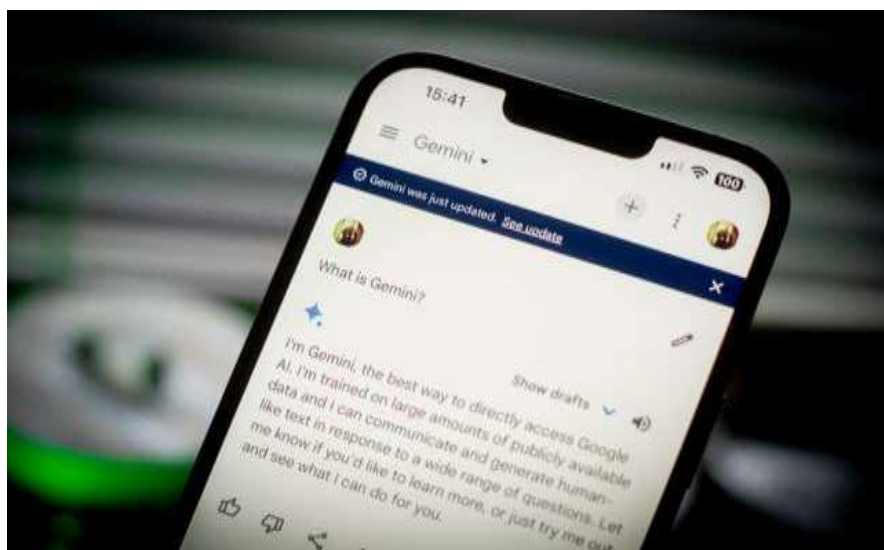
[Che cos'è un modello linguistico di grandi dimensioni \(LLM\)?](#)

I modelli linguistici di grandi dimensioni (LLM) sono un tipo di intelligenza artificiale generativa. Sono formati su grandi quantità di dati e testo, inclusi articoli di notizie ed e-book, per comprendere e generare contenuti, incluso testo in linguaggio naturale.

Fondamentalmente, vengono addestrati su tonnellate di testo in modo da poter prevedere quale parola verrà dopo. Prendi questa spiegazione da Google:

"Se iniziassi a digitare la frase "Maria ha calciato un...", un modello linguistico addestrato su dati sufficienti potrebbe prevedere "Maria ha calciato un pallone". Senza un addestramento sufficiente, potrebbe produrre solo un "oggetto rotondo" o solo il suo colore "giallo".

Chatbot popolari come ChatGPT di OpenAI e Gemini di Google, che hanno funzionalità come il riepilogo e la traduzione del testo, sono esempi di LLM.



Cos'è un semiconduttore?

I semiconduttori, noti anche come chip AI, vengono utilizzati nei circuiti elettrici di dispositivi come telefoni e computer. I dispositivi elettronici non esisterebbero senza i semiconduttori, costituiti da elementi puri come il silicio o composti come *l'arseniuro di gallio*, per condurre l'elettricità.

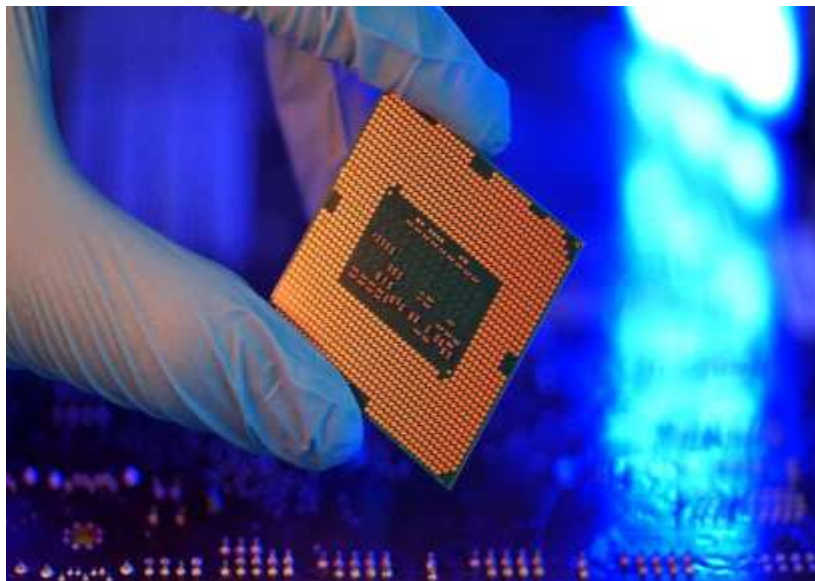
Il nome "semi" deriva dal fatto che il materiale può condurre più elettricità di un isolante, ma meno elettricità di un conduttore puro come il rame.

La più grande fonderia di semiconduttori al mondo, **Taiwan Semiconductor Manufacturing Company (TSMC)**, produce circa il 90% dei chip avanzati nel mondo e annovera tra i suoi clienti i migliori progettisti di chip Nvidia e Advanced Micro Devices (AMD).

Anche se i semiconduttori sono stati inventati negli Stati Uniti, ora producono circa il 10% dei chip mondiali, esclusi quelli avanzati necessari per modelli di intelligenza artificiale più grandi.

Il presidente Joe Biden ha firmato il CHIPS and Science Act nel 2022 per riportare la produzione di chip negli Stati Uniti, e l'amministrazione Biden ha già investito miliardi in aziende di semiconduttori tra cui Intel e TSMC per costruire fabbriche in tutto il paese.

Parte di questo sforzo ha anche a che fare con il contrastare i progressi della Cina nella produzione di chip e nello sviluppo dell'intelligenza artificiale.



Semiconduttore o microchip.

Cosa sono GPU e CPU?

Una GPU è un'unità di elaborazione grafica, un chip avanzato (o semiconduttore) che alimenta i grandi modelli linguistici dietro i chatbot AI come ChatGPT.

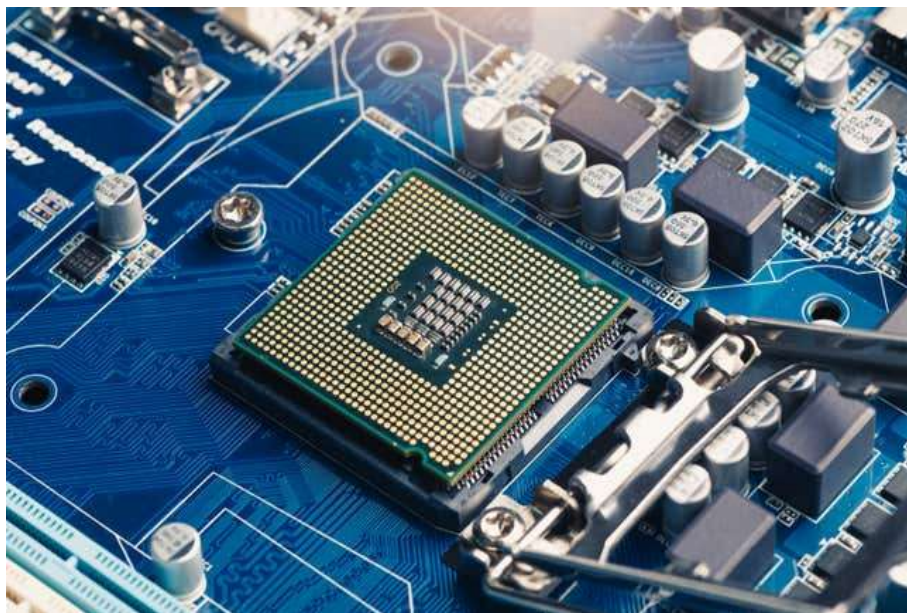
Tradizionalmente veniva utilizzato per realizzare videogiochi con immagini di qualità superiore.



Poi uno scienziato informatico ucraino-canadese, **Alex Krizhevsky**, ha mostrato come l'uso di una GPU potrebbe alimentare modelli di deep learning molto più velocemente di una CPU, un'unità di elaborazione centrale, o l'hardware principale che alimenta i computer.

Le CPU sono il "*cervello*" di un computer, che esegue le istruzioni affinché il computer funzioni. Una CPU è un processore *che* legge e interpreta le istruzioni del software per controllare le funzioni del computer. Ma una GPU è un acceleratore, un componente hardware progettato per far avanzare una funzione specifica di un processore.

Nvidia è il principale progettista di GPU, con i suoi chip H100 e H200 utilizzati nei data center delle principali aziende tecnologiche per alimentare il software di intelligenza artificiale. Altre aziende mirano a competere con gli acceleratori di Nvidia, tra cui Intel con il suo acceleratore Gaudi 3 e la GPU Azure Maia 100 di Microsoft.

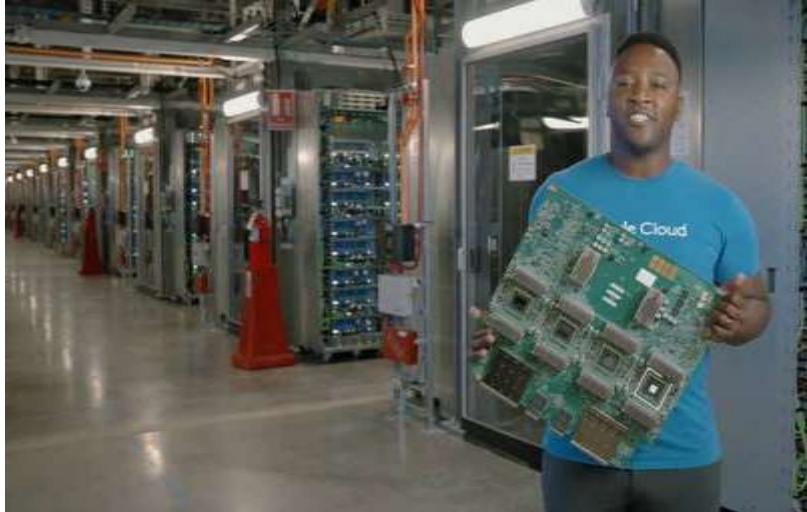


Unità di elaborazione principale (CPU) nella scheda madre

Cos'è un TPU?

TPU sta per "*unità di elaborazione tensore*". I chip di Google, a differenza di quelli di Microsoft e Nvidia, sono TPU: chip progettati su misura realizzati appositamente per l'addestramento di modelli IA di grandi dimensioni (mentre le GPU inizialmente erano realizzate per i giochi, non per l'intelligenza artificiale).

Mentre le CPU sono processori generici e le GPU sono un processore aggiuntivo che esegue attività di fascia alta, le TPU sono acceleratori personalizzati per eseguire servizi di intelligenza artificiale, rendendoli ancora più potenti.



Un TPU

Cos'è un'allucinazione?

Come accennato in precedenza, i chatbot con intelligenza artificiale sono in grado di svolgere molti compiti, ma commettono anche molti errori. Quando LLM come ChatGPT inventano informazioni false o prive di senso, si parla *di allucinazione*.

I chatbot *“hanno allucinazioni”* quando non dispongono dei dati di addestramento necessari per rispondere a una domanda, ma generano comunque una risposta che sembra un dato di fatto. Le allucinazioni possono essere causate da diversi fattori, come dati di addestramento ***imprecisi o distorti e overfitting***, ovvero quando un algoritmo non può fare previsioni o conclusioni da dati diversi da quelli su cui è stato addestrato.

Le allucinazioni sono attualmente uno dei maggiori problemi con i modelli di intelligenza artificiale generativa e non sono esattamente facili da risolvere. Poiché i modelli di intelligenza artificiale vengono addestrati su enormi quantità di dati, può essere difficile individuare problemi specifici nei dati.

A volte, i dati utilizzati per addestrare i modelli di intelligenza artificiale sono comunque imprecisi, perché provengono da luoghi come Reddit. Sebbene i modelli di intelligenza artificiale siano addestrati a non rispondere a domande di cui non conoscono la risposta, a volte non rifiutano queste domande e generano invece risposte imprecise.

Tuttavia :

Nonostante i progressi dell'intelligenza artificiale, la supervisione umana rimane essenziale !

*Domani 25 aprile:
Uno studio rivela i limiti della IA
nella codifica medica
to be continued...*

Ossitocina intranasale per l'obesità

Franziska Plessow, Ph.D. et al

Publicato il 23 aprile 2024

NEJM Evid 2024 ; 3 (5)

DOI: 10.1056/EVIDoa2300349

VOL. 3NO . 5

L'accumulo di prove precliniche e preliminari traslazionali mostra che il peptide ipotalamico ossitocina riduce l'assunzione di cibo, aumenta il dispendio energetico e promuove la perdita di peso. Al momento non è noto se la somministrazione di ossitocina sia efficace nel trattamento dell'obesità umana.

METODI

In questo **studio randomizzato, in doppio cieco**, controllato con placebo, abbiamo assegnato in modo casuale adulti con obesità 1:1 (stratificati per sesso e classe di obesità) a ricevere ossitocina intranasale (24 UI) o placebo quattro volte al giorno per 8 settimane.

L'endpoint primario era la variazione del peso corporeo (kg) dal basale alla settimana 8.

Gli endpoint secondari chiave includevano la variazione della composizione corporea (massa grassa totale [g], tessuto adiposo viscerale addominale [cm²] e frazione di grasso epatico [proporzione ; intervallo, da 0 a 1; valori più alti indicano una percentuale maggiore di grassi]), e dispendio energetico a riposo (kcal/giorno; aggiustato per massa magra) dal basale alla settimana 8 e apporto calorico (kcal) ad un pasto di prova sperimentale dal basale alla settimana 6.

RISULTATI

Sessantuno partecipanti (54% donne; età media \pm deviazione standard, 33,6 \pm 6,2 anni; indice di massa corporea [il peso in chilogrammi diviso per il quadrato dell'altezza in metri], 36,9 \pm 4,9) sono stati assegnati in modo casuale.

Non è stata riscontrata alcuna differenza nella variazione del peso corporeo dal basale alla settimana 8 tra i gruppi ossitocina e placebo (0,20 vs 0,26 kg; P=0,934).

L'ossitocina (rispetto al placebo) non è stata associata ad effetti benefici sulla composizione corporea o sul dispendio energetico a riposo dal basale alla settimana 8 (grasso totale: differenza [intervallo di confidenza al 95%], 196,0 g [da -1036 a 1428]; grasso viscerale: 3,1 cm² [da -11,0 a 17,2]; grasso epatico: -0,01 [da -0,03 a 0,01]; dispendio energetico a riposo: -64,0 kcal/giorno [da -129,3 a 1,4]).

L'ossitocina rispetto al placebo è stata associata a un ridotto apporto calorico al pasto di prova (-31,4 vs. 120,6 kcal; differenza [intervallo di confidenza al 95%], -152,0 kcal [da -302,3 a -1,7]).

Non ci furono dei seri eventi avversi. L'incidenza e la gravità degli eventi avversi non differivano tra i gruppi.

CONCLUSIONI

In questo studio randomizzato e controllato con placebo condotto su adulti obesi, l'ossitocina intranasale somministrata quattro volte al giorno per 8 settimane non ha ridotto il peso corporeo.

(Finanziato dal National Institute of Diabetes and Digestive and Kidney Diseases e altri; numero ClinicalTrials.gov, [NCT03043053](https://clinicaltrials.gov/ct2/show/study/NCT03043053) .)