



# Large Language Models e Applicazioni

Luca Bellanova, Data Scientist



# ● Indice

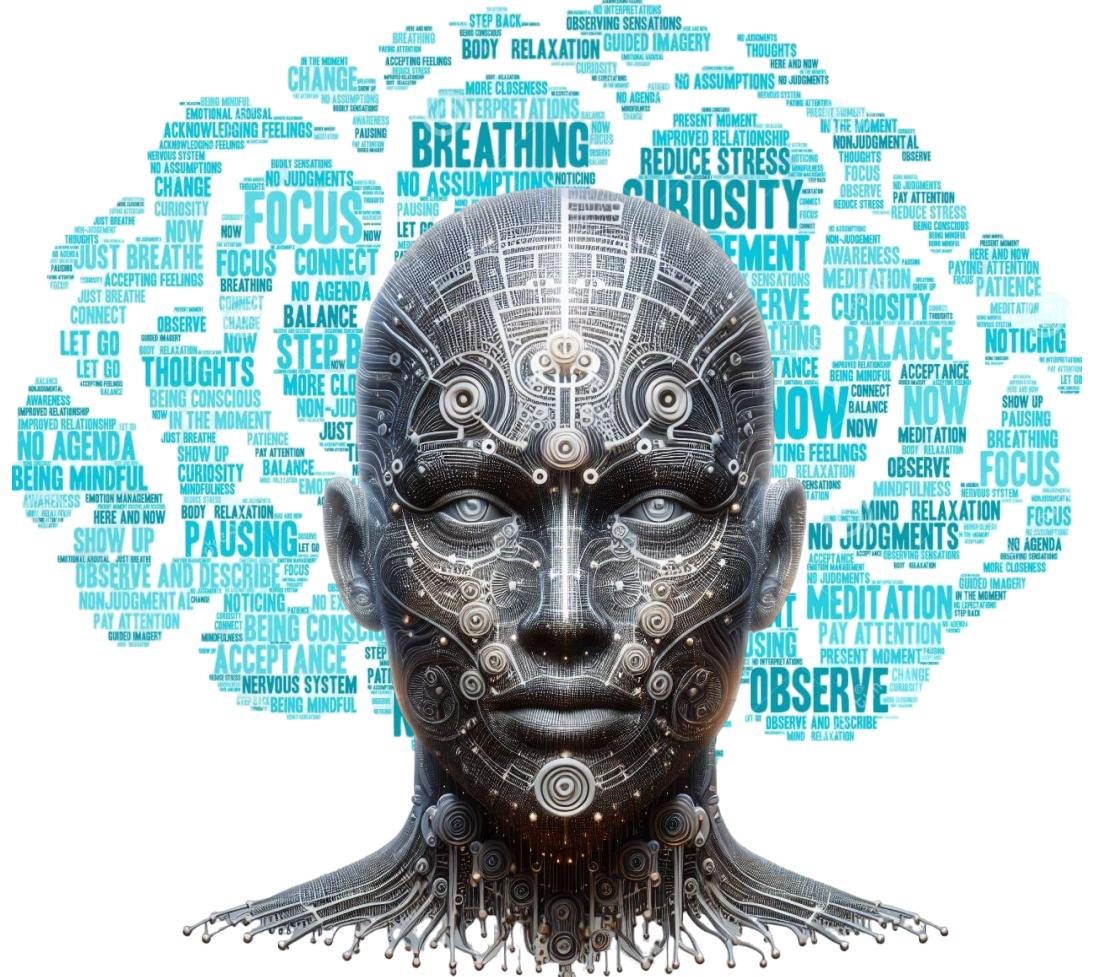
- Introduzione
- NLP
- L'Evoluzione dei Modelli
- Tecniche di Modeling
- Prompt/Flow Engineering
- Framework
- Rischi e Limitazioni
- Applicazioni nel Mondo Reale
- Modelli Multimodali
- Il Futuro dei LLM

# INTRODUZIONE

---

# • Che cos'è un Language Model?

*Modello di linguaggio avanzato, addestrato su enormi quantità di informazioni linguistiche per comprendere e generare linguaggio naturale.*



# ● Capacità



## Comprendere

Interpretare e attribuire significato al testo sfruttando le capacità di comprensione della semantica e estrazione di informazioni significative da un insieme di parole o frasi



## Generare

Produrre nuovo testo in risposta a una domanda o a una richiesta creando sequenze di parole che siano grammaticalmente corrette e coerenti nel contesto fornito

# • Definizione Classica

## *Distribuzione di probabilità su sequenze di token*

Supponiamo di avere un **vocabolario  $V$**  di un insieme di token

$$V = \{ate, ball, cheese, mouse, the\}$$

Un **modello linguistico  $p$**  assegna a ciascuna sequenza di token, appartenenti al vocabolario  $V$ , una probabilità (un numero compreso tra 0 e 1)

**$p$  più BASSA**  
Meno Coerente e Plausibile  
**Conoscenza Semantica**

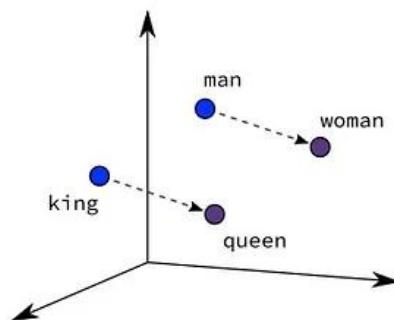
$$\begin{aligned} p(\text{the}, \text{mouse}, \text{ate}, \text{the}, \text{cheese}) &= 0.02, \\ p(\text{the}, \text{cheese}, \text{ate}, \text{the}, \text{mouse}) &= 0.01, \\ p(\text{mouse}, \text{the}, \text{the}, \text{cheese}, \text{ate}) &= 0.0001 \end{aligned}$$

**$p$  più BASSA**  
Grammaticalmente Scorretta  
**Conoscenza Sintattica**

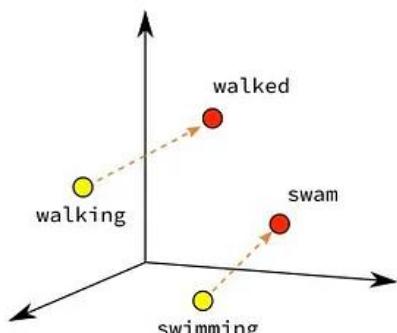
# • Embeddings

Gli **embeddings** sono vettori o matrici di numeri che rappresentano il significato e il contesto.

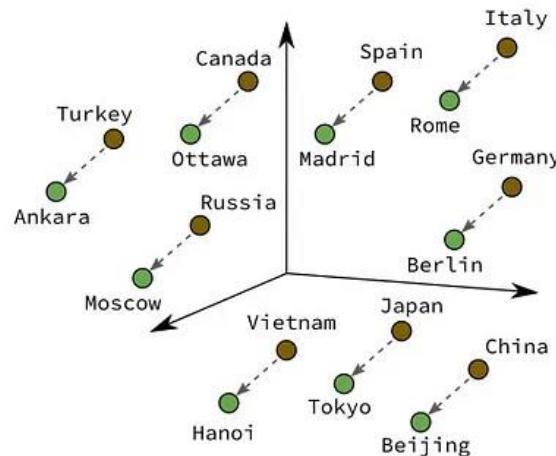
Gli embeddings possono aiutare il modello a comprendere le **relazioni semantiche** e **sintattiche** tra i token e a generare testi più pertinenti e coerenti. Inoltre, possono anche consentire al modello di gestire compiti **multimodali**, come la generazione di immagini e codici, convertendo diversi **tipi** di dati in una rappresentazione comune.



Male-Female



Verb Tense



Country-Capital

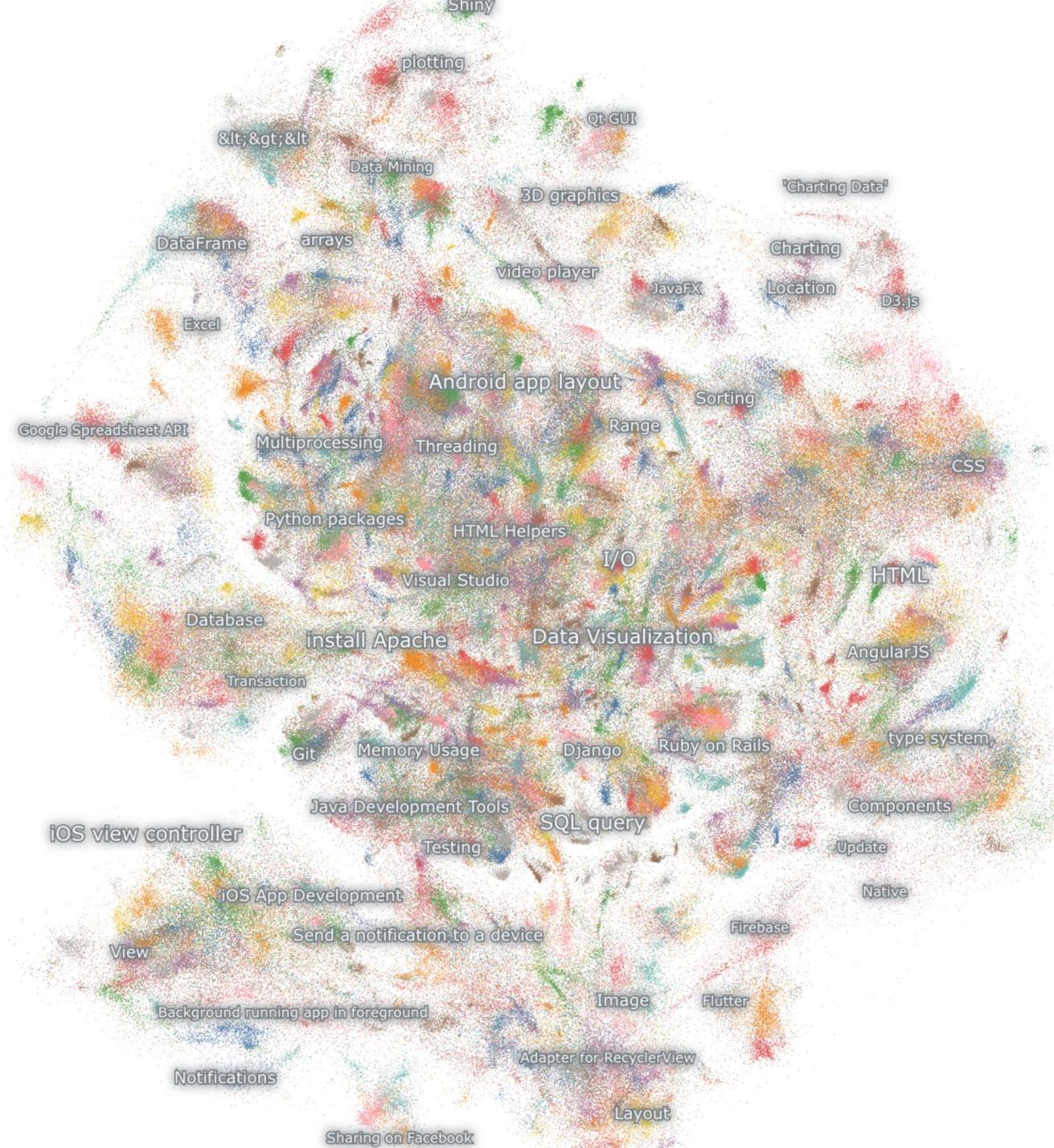
Le parole che possiedono **significati simili** avranno una rappresentazione vettoriale simile.

Le **relazioni geometriche** rappresentano **relazioni semantiche**.

# • Embeddings

Embedding di 8 milioni di domande di Stack Overflow

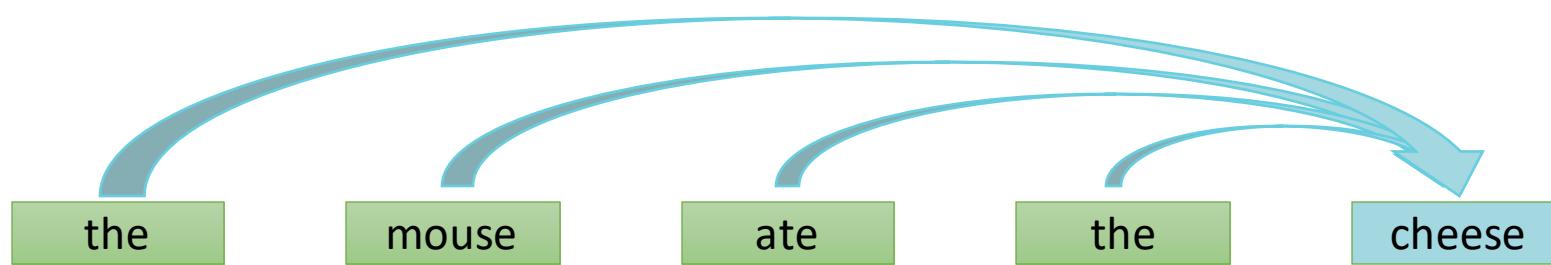
(<https://atlas.nomic.ai/map/edaff028-12b5-42a0-8e8b-6430c9b8222b/bcb42818-3581-4fb5-ac30-9883d01f98ec>)



# • Modelli Linguistici Autoregressivi

Tipologia di LM che operano **sequenzialmente**, generando un'uscita (output) un passo alla volta, in base alle informazioni **precedentemente generate**.

Questi modelli sono progettati per prevedere la probabilità della prossima parola nel contesto di ciò che è stato generato fino a quel momento.



 Famiglia di modelli **GPT** (Generative Pre-trained Transformer), come GPT-3 sviluppato da OpenAI.

# ● Temperatura

La **temperatura** in un Language Model (LM) è un parametro che controlla la **casualità** delle sue previsioni durante la generazione del testo.

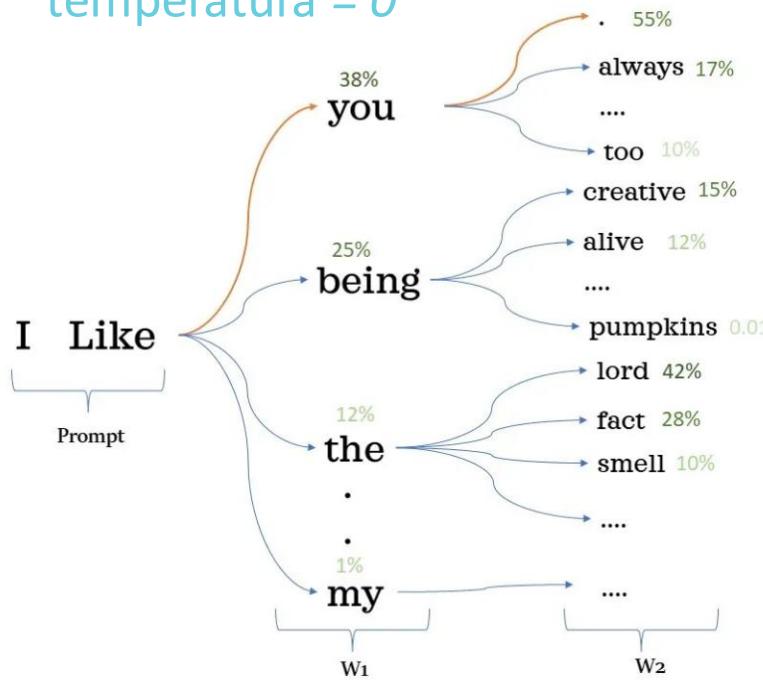
In sostanza, la temperatura influenza la diversità e l'incertezza delle risposte generate dal modello.

TEMPERATURA = 0	TEMPERATURA > 0
<p>Generazione <b>deterministica</b>. Il modello sceglierà sempre la parola più probabile in base al suo training.</p>	<p><b>Casualità</b> nella generazione. Un valore più alto di temperatura rende la distribuzione di probabilità più uniforme*, portando a una maggiore variabilità nelle scelte del modello durante la generazione.</p>

\*aumentando la temperatura si distribuisce la probabilità in modo più equo tra diverse possibilità.

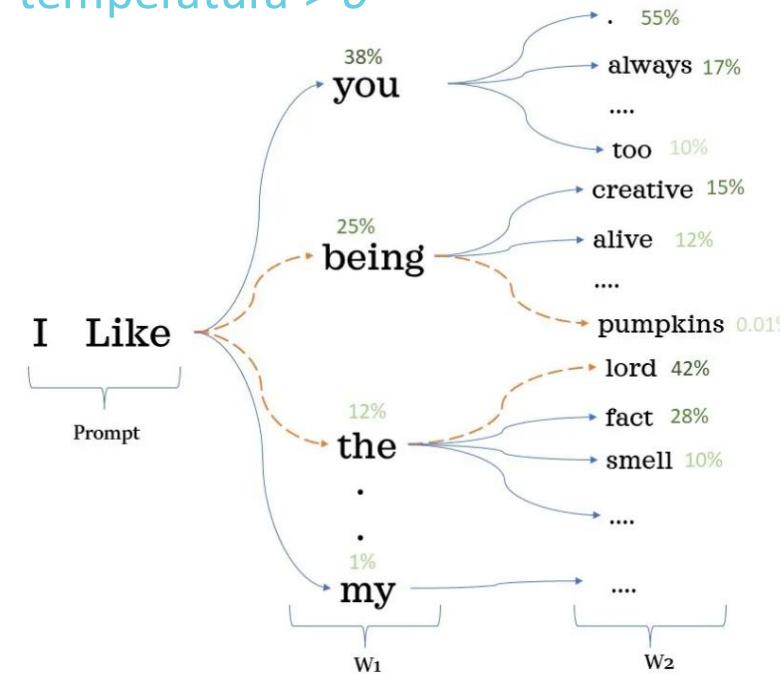
# ● Temperatura

temperatura = 0



Quando la **temperatura** è **0**, il modello selezionerà sempre la parola successiva più probabile, rendendo il processo completamente deterministico. La sequenza generata sarà sempre la stessa per uno specifico input. Se, ad esempio, il prompt è "I Like", la completamento sarà sempre "I Like you."

temperatura > 0



Quando la **temperatura** è **maggiori di 0**, durante l'inferenza il modello selezionerà casualmente la parola successiva, rendendo il processo non deterministico e più creativo. Con il prompt "I Like", il completamento potrebbe essere "I Like being pumpkins" oppure "I Like the lord". C'è un grado di casualità coinvolto nella selezione della parola successiva.

# NLP



# ● Evoluzione Storica del Natural Language Processing (NLP)

Le **Reti Neurali Ricorrenti (RNN)** rappresentano una tappa fondamentale nell'evoluzione del Natural Language Processing. La loro capacità di gestire sequenze di dati le ha rese fondamentali in molte applicazioni (traduzione automatica, generazione di testo creativo, analisi del sentimento), consentendo la modellazione del contesto e la comprensione di strutture sequenziali nel linguaggio naturale. Tuttavia, nonostante i successi iniziali, le RNN hanno mostrato limitazioni significative, particolarmente evidenti nell'apprendere **relazioni a lungo termine** nelle sequenze. In particolare, la difficoltà nella gestione di sequenze lunghe comportava una **perdita** di informazioni cruciali, compromettendo l'**efficacia** complessiva delle reti.

Per superare queste limitazioni, sono emerse soluzioni innovative come le **Long Short-Term Memory (LSTM)**. Le LSTM hanno introdotto meccanismi di memoria a lungo termine, consentendo al modello di **mantenere e aggiornare** informazioni importanti nel corso delle sequenze. Questi miglioramenti hanno rivoluzionato l'apprendimento sequenziale, permettendo alle reti di catturare relazioni complesse e di gestire sequenze più lunghe in modo più efficace.

Riconoscimento vocale, Traduzione automatica, Rilevamento di attività vocale.

# ● Evoluzione Storica del NLP

La vera rivoluzione è arrivata con l'introduzione dei **Transformers**.

Caratterizzati dalla loro struttura di **attenzione multi-head** e dalla gestione altamente efficiente del **contesto**, i Transformers hanno ridefinito il panorama dell'NLP.

3

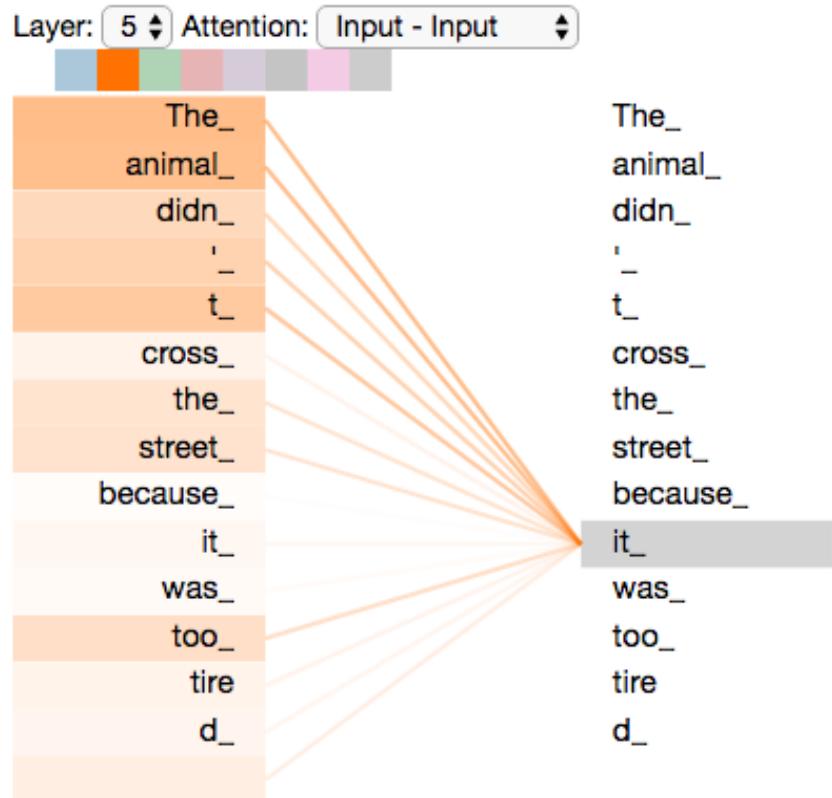
I **Transformers** analizzano pezzi di dati, chiamati *token*, e imparano a prevedere il *token successivo* in una sequenza, basandosi sui *token precedenti* e, se disponibili, su quelli *successivi*.

La possibilità di eseguire operazioni **parallele** su sequenze di dati ha portato a una significativa **accelerazione** nei tempi di addestramento dei modelli.



# • Evoluzione Storica del NLP

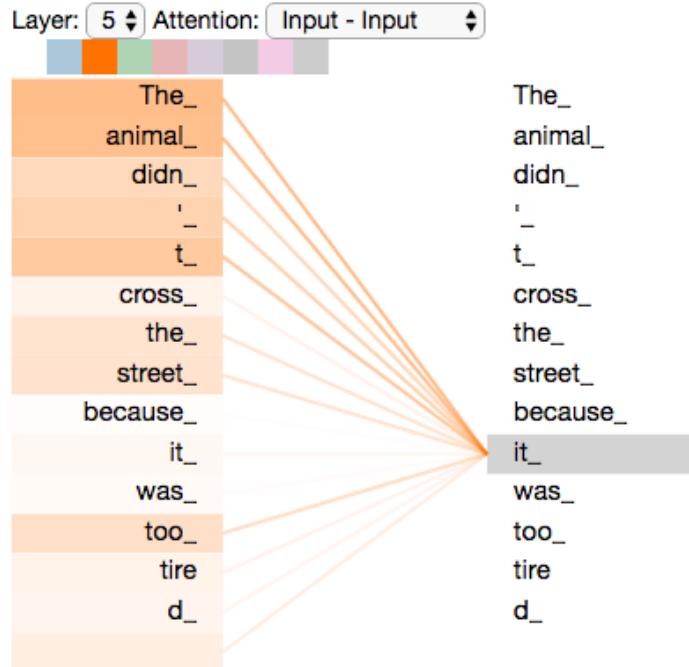
L'attenzione multi-head permette al modello di concentrarsi su diverse parti del contesto contemporaneamente. Questa capacità consente di catturare dipendenze a lungo termine nei dati, superando le limitazioni dei modelli ricorrenti che spesso lottano nel catturare relazioni distanti nel tempo.



**It** riferisce a street o animal?  
È una domanda semplice per un essere umano, ma non altrettanto per un algoritmo.

Quando il modello elabora la parola *it*, la **self-attention** gli permette di associare ***it* ad *animal***.

# ● Evoluzione Storica del NLP



## Senza attenzione

In un modello senza attenzione, ***it*** potrebbe essere interpretato in modo **generico**, senza considerare il contesto o le parole chiave più rilevanti.

## Con attenzione

Utilizzando l'attenzione, il modello può analizzare l'intera frase e attribuire maggiore importanza a determinate parole chiave, come ***animal***. Questo consente al modello di fornire una risposta più **precisa** e **personalizzata**, comprendendo che il focus principale è sull'animale e fornisce una risposta più mirata e dettagliata.

### ○ Come funziona:

- ✓ Durante l'elaborazione della richiesta, il transformer esamina attentamente ogni parola, assegnando pesi diversi alle parole più rilevanti. Quando genera una risposta, tiene conto di queste attenzioni per fornire una soluzione più accurata e specifica.

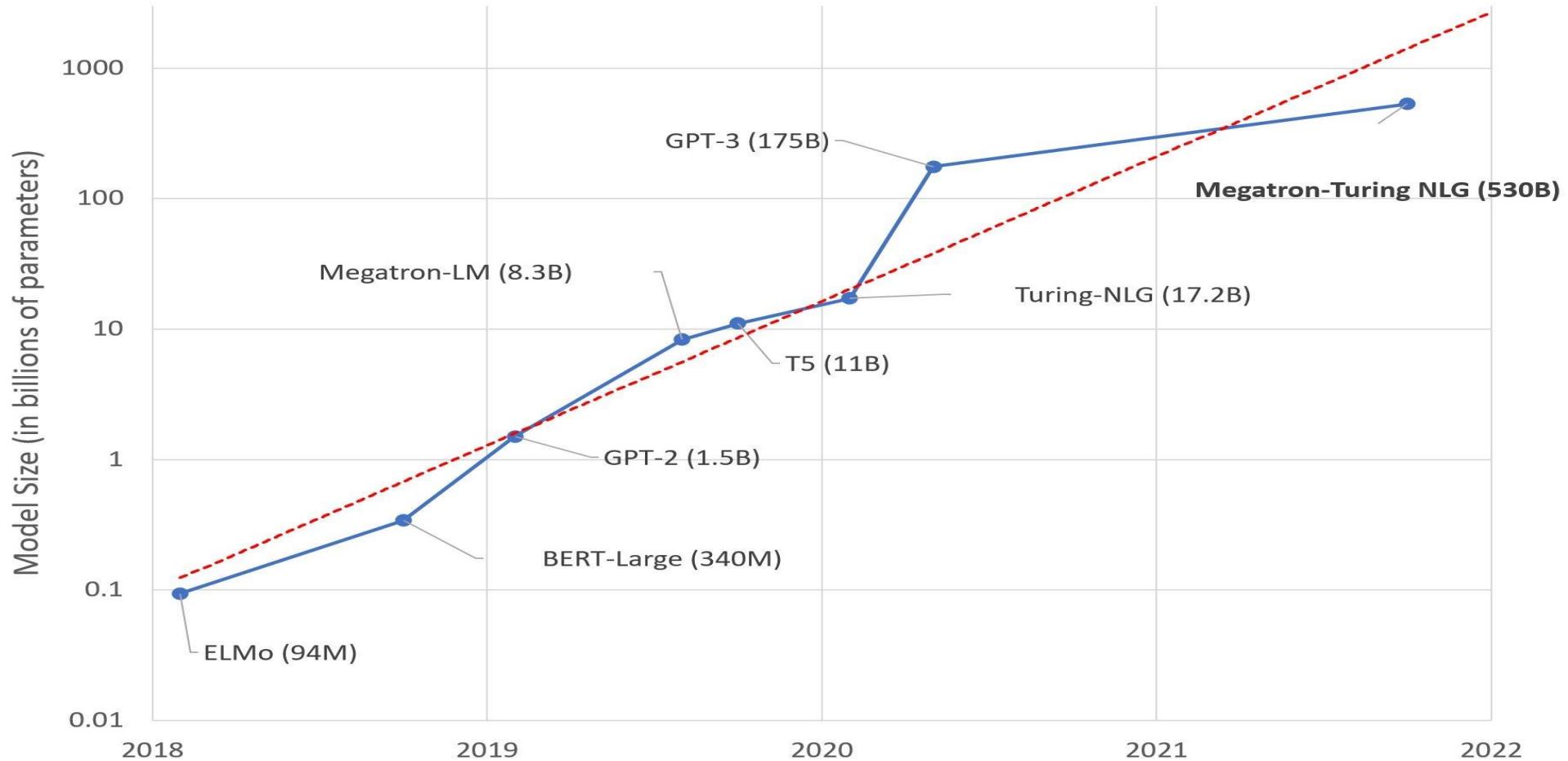
### ○ Vantaggi:

- ✓ L'uso dell'attenzione consente di comprendere meglio il contesto e di rispondere in modo più mirato alle esigenze, migliorando l'esperienza complessiva.

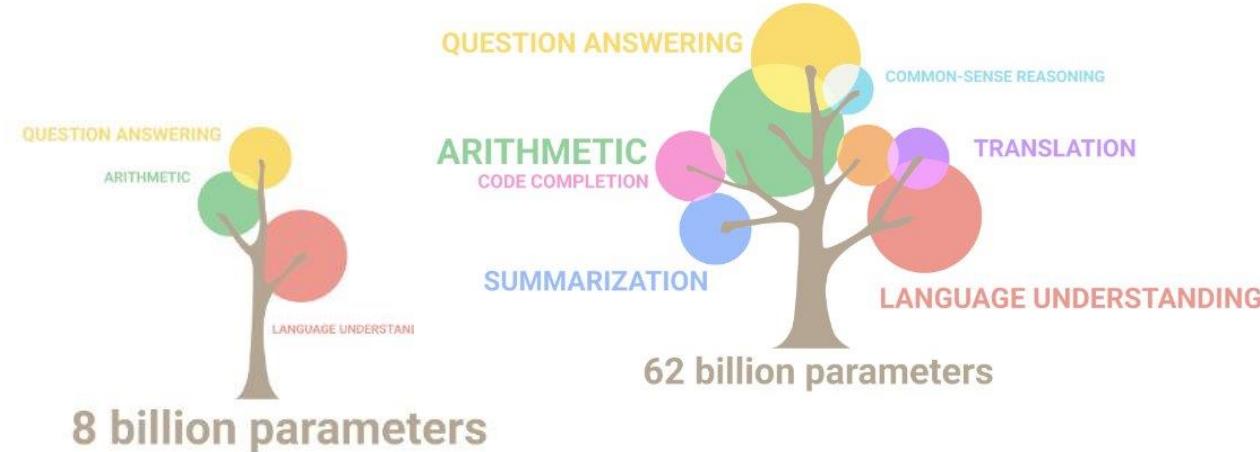
# L'EVOLUZIONE DEI MODELLI

---

# Le dimensioni dei modelli aumentano in modo esponenziale

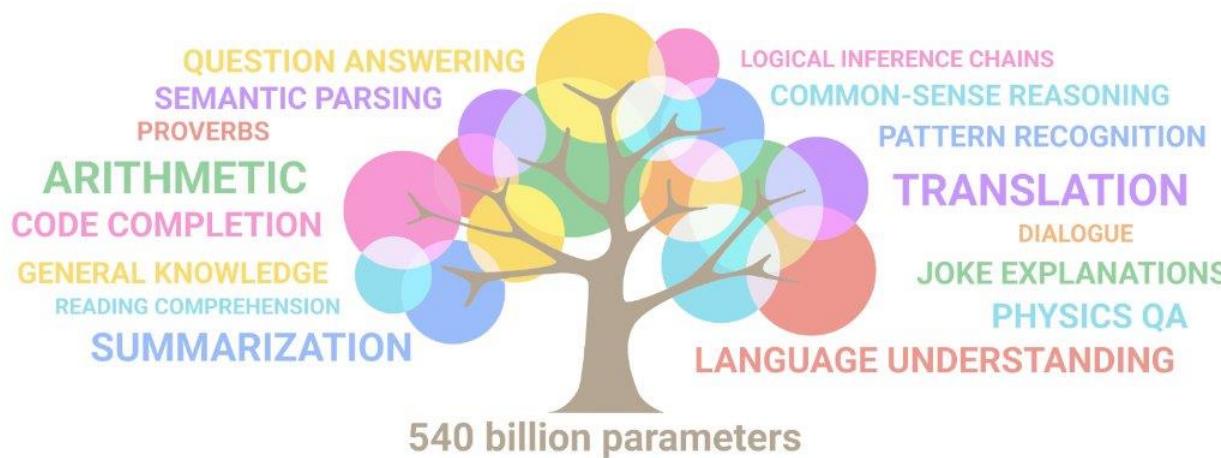


# Correlazione tra Capacità e Crescita dei Parametri

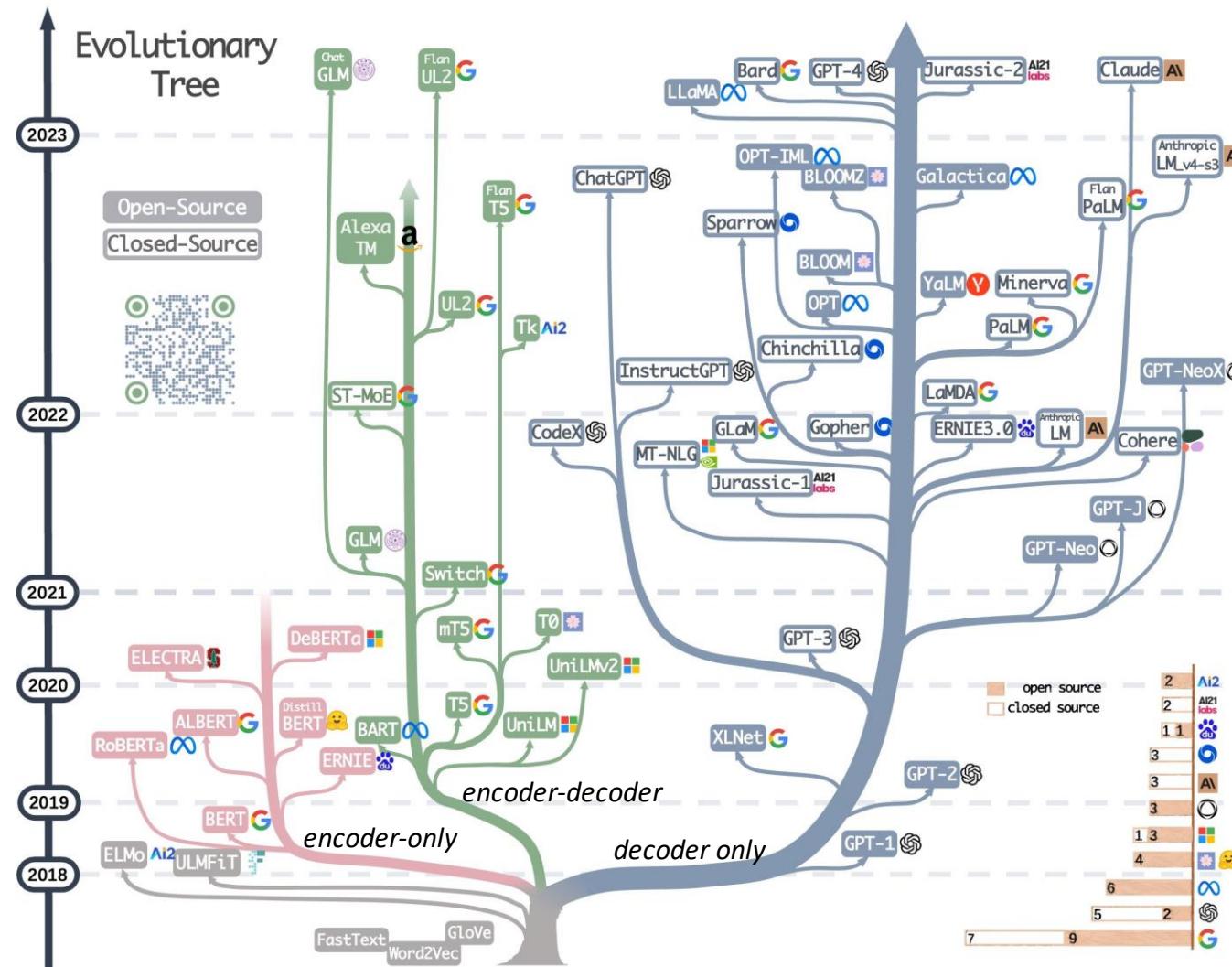


L'aumento dei parametri da 8 miliardi a 540 miliardi è correlato a un notevole incremento delle competenze del modello, che vanno dalla comprensione del linguaggio e l'aritmetica fino al ragionamento di senso comune, il riconoscimento dei modelli, le spiegazioni delle barzellette e le domande e risposte di fisica.

Questo sottolinea come l'aumento dei parametri possa portare a una **maggior versatilità e capacità** nei modelli di linguaggio.



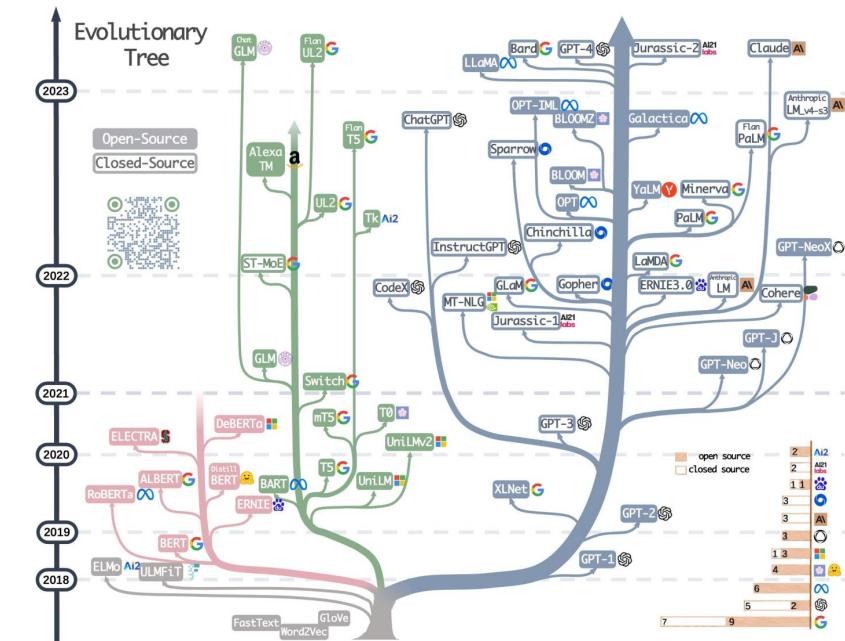
# • LLM Albero Evolutivo



# ● LLM Albero Evolutivo

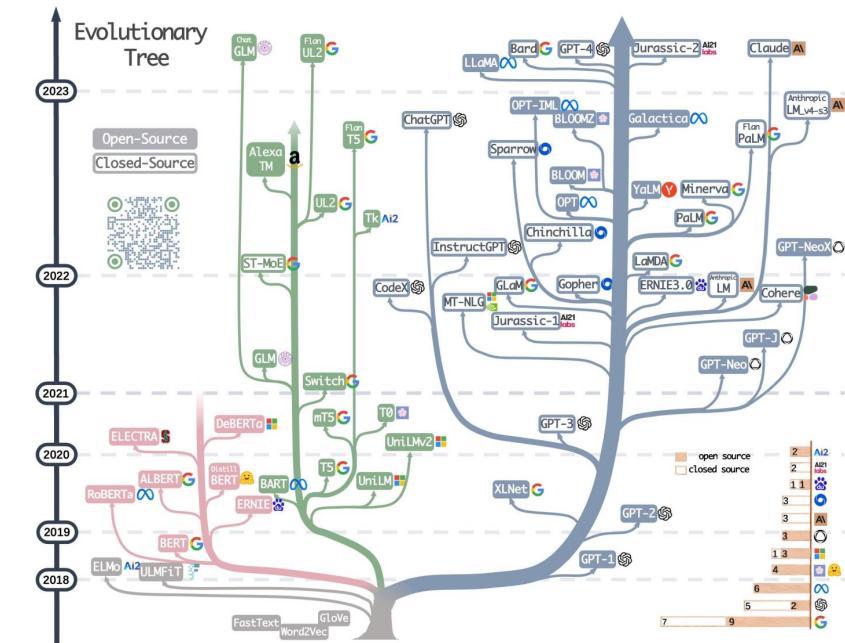
L'albero evolutivo dei moderni LLM traccia lo sviluppo dei modelli linguistici negli ultimi anni e mette in evidenza alcuni dei modelli più noti.

- I modelli sullo stesso ramo hanno relazioni più strette.
- I modelli basati su transformer sono mostrati con colori non grigi:
  - ✓ i modelli *decoder-only* nel ramo blu;
  - ✓ i modelli *encoder-only* nel ramo rosa;
  - ✓ i modelli *encoder-decoder* nel ramo verde.
- La posizione verticale dei modelli sulla timeline rappresenta le date di rilascio.
- I modelli *open-source* sono rappresentati da quadrati pieni, mentre i modelli *closed-source* sono rappresentati da quelli vuoti.
- Il diagramma a barre sovrapposte nell'angolo in basso a destra mostra il numero di modelli di varie aziende e istituzioni.



# ● LLM Albero Evolutivo

I modelli *decoder-only* hanno gradualmente dominato lo sviluppo degli LLM. All'inizio dello sviluppo degli LLM, i modelli di *decoder-only* non erano così popolari come i modelli *encoder-only* e *encoder-decoder*. Tuttavia, dopo il 2021, con l'introduzione del nuovo *GPT-3*, i modelli *decoder-only* hanno avuto un boom significativo. Nel frattempo, dopo l'iniziale crescita esplosiva determinata da *BERT*, i modelli *encoder-only* hanno iniziato gradualmente a scomparire.

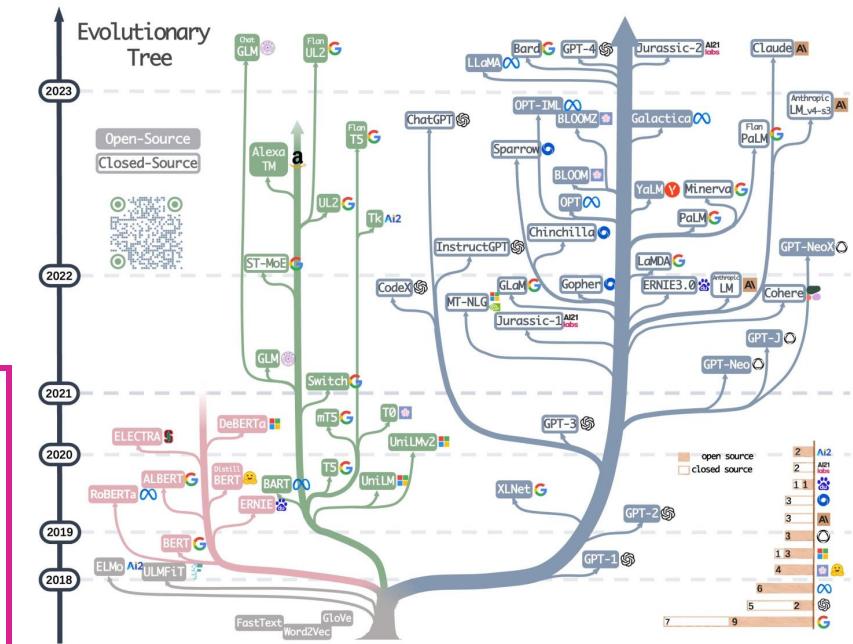


OpenAI mantiene costantemente la sua posizione di leadership nel settore LLM, sia attualmente che potenzialmente in futuro. Altre aziende e istituzioni stanno lottando per raggiungere OpenAI nello sviluppo di modelli comparabili a *GPT-3* e all'attuale *GPT-4*.

Meta contribuisce in modo significativo allo sviluppo di LLM *open-source* e ne promuove la ricerca. Meta si distingue come una delle aziende commerciali più generose; tutti gli LLM sviluppati da Meta sono *open-source*.

# • LLM Albero Evolutivo

I LLM mostrano una tendenza al *closed-sourcing*. Nelle prime fasi di sviluppo dei LLM (prima del 2020), la maggior parte dei modelli erano *open-sourced*. Con l'introduzione di *GPT-3*, le aziende hanno optato sempre più spesso per il *close-sourcing* dei loro modelli, come *PaLM*, il *LaMDA* e *GPT-4*.



I modelli *encoder-decoder* rimangono promettenti, in quanto questo tipo di architettura è ancora in fase di esplorazione attiva e la maggior parte di essi sono *open-sourced*. Google ha dato un contributo sostanziale alle architetture *encoder-decoder open-source*. Tuttavia, la flessibilità e la versatilità dei modelli decoder-only sembrano rendere meno promettente l'insistenza di Google in questa direzione.

# • Modelli a confronto

Model	Provider	Open-Source	Speed	Quality	Params	FINE-TUNEABILITY
gpt-4	OpenAI	No	★☆☆	★★★★	-	No
gpt-3.5-turbo	OpenAI	No	★☆☆	★★★★	175B	No
gpt-3	OpenAI	No	★☆☆	★★★★	175B	No
ada, babbage, curie	OpenAI	No	★☆☆	★★★★	350M - 7B	Yes
claude	Anthropic	Yes	★☆☆	★★★★	52B	No
claude-instant	Anthropic	Yes	★☆☆	★★★★	52B	No
command-xlarge	Cohere	No	★☆☆	★★★★	50B	Yes
command-medium	Cohere	No	★☆☆	★★★★	6B	Yes
BERT	Google	Yes	★☆☆	★★★★	345M	Yes
T5	Google	Yes	★☆☆	★★★★	11B	Yes
PaLM	Google	Yes	★☆☆	★★★★	540B	Yes
LLaMA	Meta AI	Yes	★☆☆	★★★★	65B	Yes
CTRL	Salesforce	Yes	★☆☆	★★★★	1.6B	Yes
Dolly 2.0	Databricks	Yes	★☆☆	★★★★	12B	Yes

# • ChatGPT



OpenAI:

*"Abbiamo addestrato un modello chiamato ChatGPT che interagisce in modo colloquiale. Il formato di dialogo consente a ChatGPT di rispondere a domande successive, ammettere i propri errori, contestare premesse errate e rifiutare richieste inappropriate."*

## ChatGPT

Examples	Capabilities	Limitations
"Explain quantum computing in simple terms" →	Remembers what user said earlier in the conversation	May occasionally generate incorrect information
"Got any creative ideas for a 10 year old's birthday?" →	Allows user to provide follow-up corrections	May occasionally produce harmful instructions or biased content
"How do I make an HTTP request in Javascript?" →	Trained to decline inappropriate requests	Limited knowledge of world and events after 2021

# • ChatGPT



**Plugins:** strumenti progettati per i LLM, i quali aiutano ChatGPT ad accedere a informazioni aggiornate, a eseguire calcoli o a utilizzare servizi di terze parti.



**Expedia**

Bring your trip plans to life—get there, stay there, find things to see and do.



**FiscalNote**

Provides and enables access to select market-leading, real-time data sets for legal, political, and regulatory data and information.



**Instacart**

Order from your favorite local grocery stores.



**KAYAK**

Search for flights, stays and rental cars. Get recommendations for all the places you can go within your budget.



**Klarna Shopping**

Search and compare prices from thousands of online shops.



**Milo Family AI**

Giving parents superpowers to turn the manic to magic, 20 minutes each day. Ask: Hey Milo, what's magic today?



**OpenTable**

Provides restaurant recommendations, with a direct link to book.



**Shop**

Search for millions of products from the world's greatest brands.



**Speak**

Learn how to say anything in another language with Speak, your AI-powered language tutor.



**Wolfram**

Access computation, math, curated knowledge & real-time data through Wolfram|Alpha and Wolfram Language.



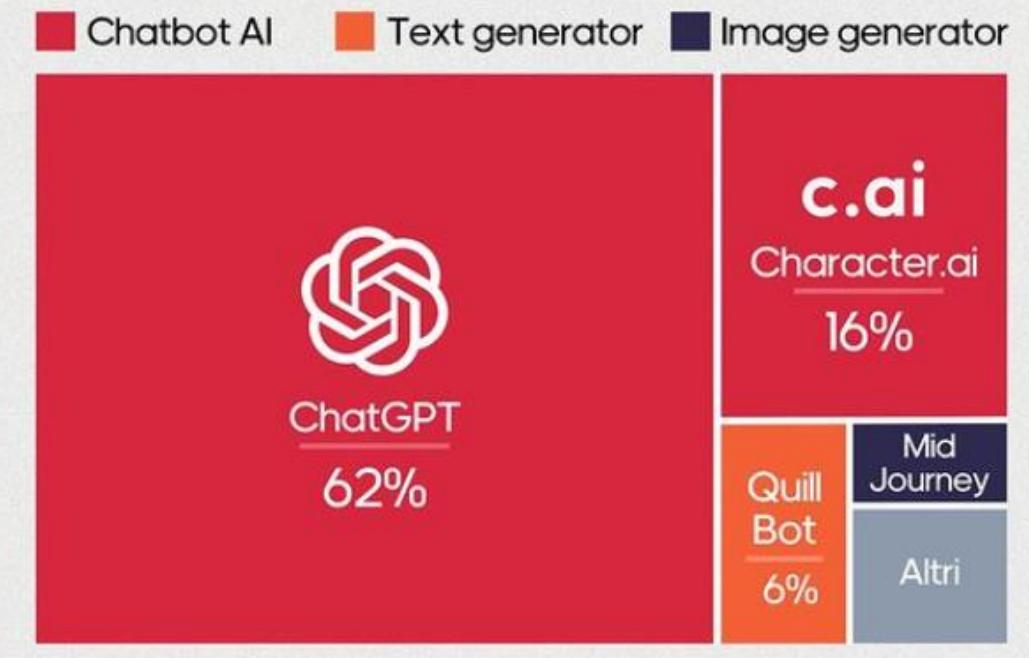
**Zapier**

Interact with over 5,000+ apps like Google Sheets, Trello, Gmail, HubSpot, Salesforce, and more.

## Il Traffico verso Tool AI nel 2023

# ChatGPT domina: nel 2023 è suo più della metà del traffico verso Tool AI

Il Chatbot di Open AI ha ricevuto quasi **15 miliardi di visite**, mentre character.ai e QuillBot si fermano rispettivamente a 4 miliardi e a 1,1 miliardi



# **TECNICHE DI MODELING**

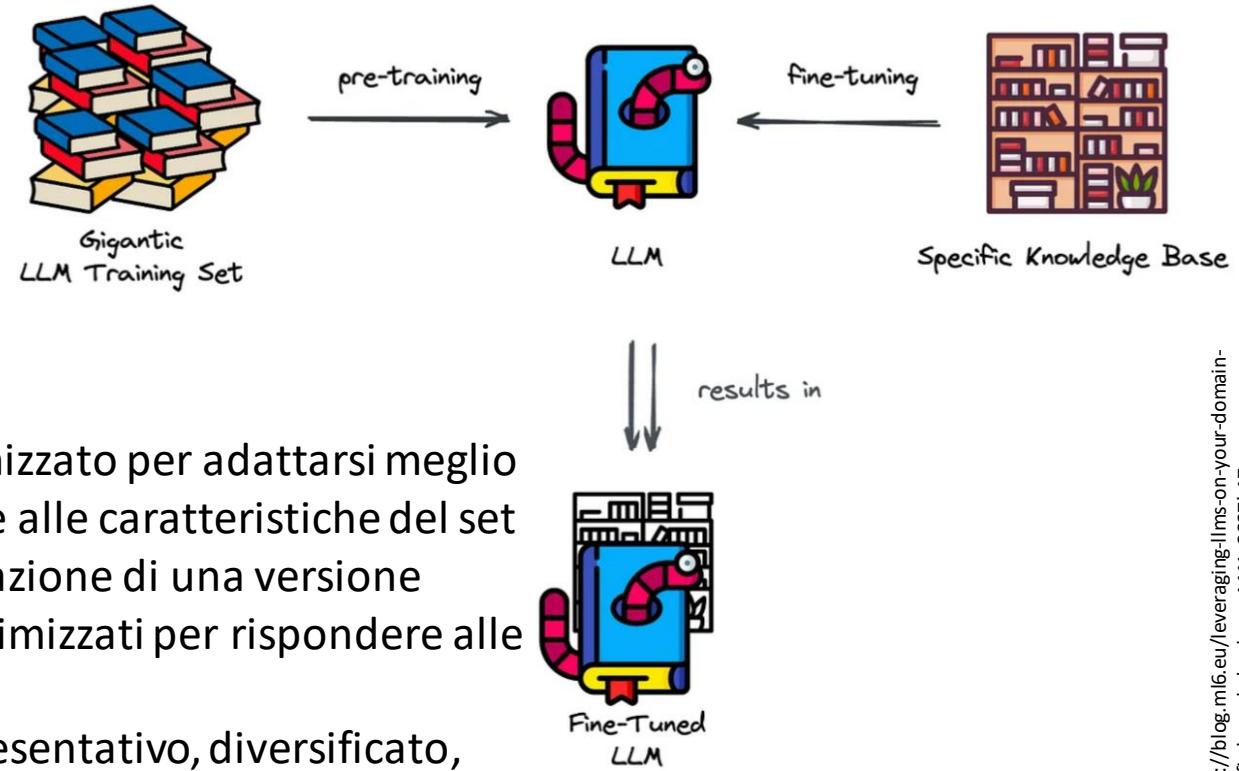


# • Fine Tuning e Apprendimento In-Context

Il **fine-tuning** è un processo che implica l'ulteriore addestramento di un LLM per **specializzarlo** su un task/dominio specifico. Questo metodo parte dall'LLM **pre-addestrato** sottoponendolo a una fase di addestramento **successiva** su un insieme di dati etichettati, specifici per il contesto d'uso desiderato.

Durante il fine-tuning, il modello LLM viene ottimizzato per adattarsi meglio ai nuovi dati, **regolando i pesi** del modello in base alle caratteristiche del set di dati di addestramento. Ciò si traduce nella creazione di una versione personalizzata del modello originale, con pesi ottimizzati per rispondere alle esigenze specifiche del task/dominio.

Il set di dati di addestramento deve essere rappresentativo, diversificato, equilibrato e coerente per garantire una specializzazione efficace del modello LLM.

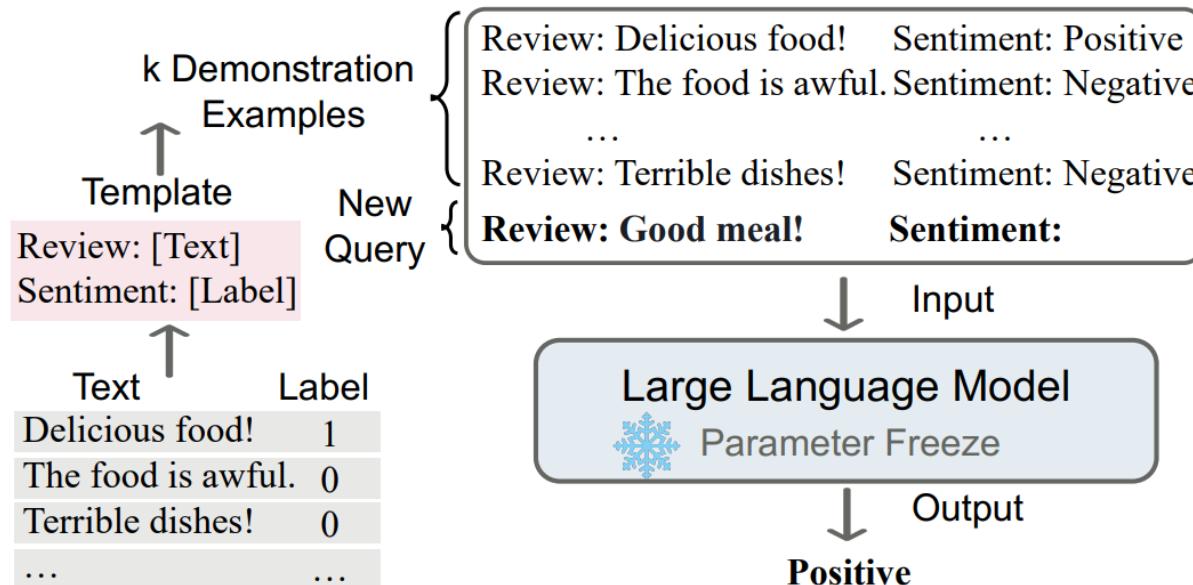


# • Fine Tuning e Apprendimento In-Context

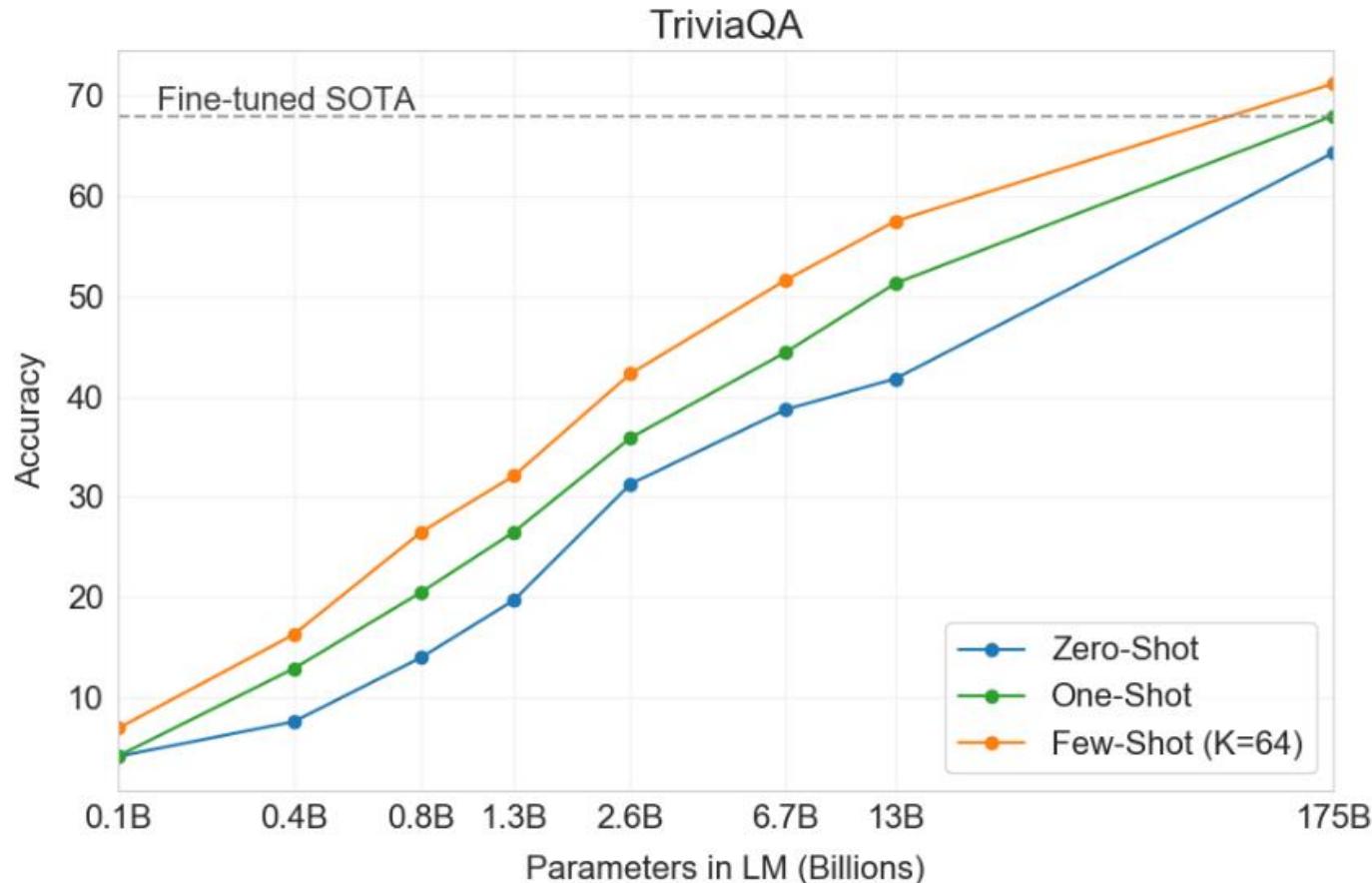
L'apprendimento in-context è un processo in cui un LLM pre-addestrato viene guidato nelle risposte a compiti specifici attraverso la fornitura di un **prompt** o **contesto**.

A differenza del fine-tuning, **non** richiede ulteriore addestramento e consente al modello di essere riutilizzato per diversi compiti con richieste varie.

I prompt possono assumere diverse forme, come **istruzioni**, un **singolo esempio** o **alcuni esempi** di output previsto in base all'input fornito. Questo approccio sfrutta la **conoscenza preesistente** del modello e la sua **capacità di generalizzazione** per eseguire compiti specifici in base ai segnali contestuali forniti.



# • Fine Tuning e Apprendimento In-Context



Si tratta di una sorta di *apprendimento effimero supervisionato*.

- **Zero-Shot Learning**: non viene fornito alcun esempio.
- **One-Shot Learning**: viene fornito un esempio.
- **Few-Shot Learning**: vengono forniti pochi esempi.

# • Fine Tuning e Apprendimento In-Context

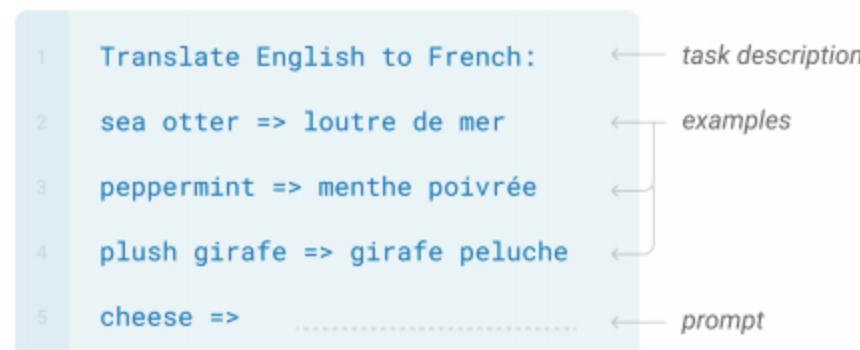
## Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.



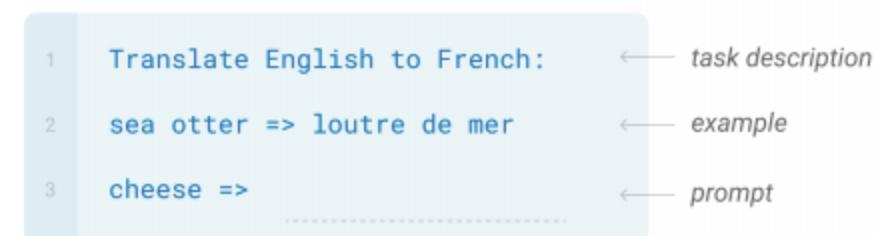
## Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.



## One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.



# • Fine Tuning e Apprendimento In-Context

*Quale approccio funziona meglio?*



## Numero di tasks distinti

Il fine-tuning richiede la creazione di dati di addestramento separati e specifici per ogni task. L'apprendimento in-context, invece, è più facilmente adattabile a un'ampia gamma di task senza bisogno di modelli separati.



## Disponibilità dei dati

Il fine-tuning richiede dati di alta qualità e specifici per il dominio, la cui creazione può richiedere tempo e denaro. L'apprendimento in-context non richiede dati di addestramento, ma avere alcuni esempi è utile per lo One o Few Shot Learning.



## Prompt engineering skills

Un apprendimento in-context efficace si basa spesso sull'arte di creare prompt efficaci. La qualità dell'output dipende dalla qualità del prompt. C'è anche una limitazione alla quantità di contesto che può essere fornita a causa dei limiti di token nell'architettura del modello. Il fine-tuning non richiede prompt engineering skills, ma buoni dati.



## Accuracy

Il fine-tuning garantisce un'accuratezza e un'affidabilità maggiore grazie a un controllo più preciso sui dati utilizzati. Al contrario, le risposte dell'apprendimento in-context possono variare di qualità a seconda dell'efficacia del contesto.



## Tempistica per il Deployment

Il fine-tuning richiede tempo e impegno per raccogliere i dati di addestramento, mettere a punto il modello e iterare per ottenere la migliore precisione. L'apprendimento in-context non richiede una formazione aggiuntiva: più rapido.



## Problemi di privacy

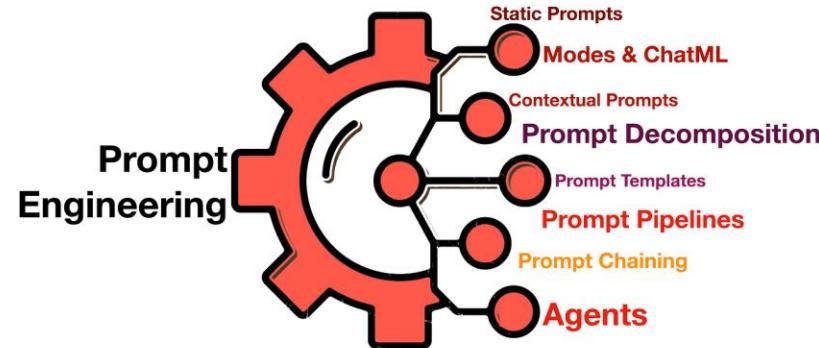
Nelle applicazioni in cui la privacy dei dati è fondamentale, il fine-tuning può aiutare il modello ad anonimizzare meglio le informazioni sensibili. L'apprendimento in-context si basa sull'ingegnerizzazione immediata, quindi anche se è possibile anonimizzare, il contesto potrebbe non essere trasmesso correttamente.

# PROMPT/FLOW ENGINEERING

---

# • Prompt Engineering

Il **Prompt Engineering** è il processo che coinvolge la creazione e l'ottimizzazione di **prompt** per guidare un modello di intelligenza artificiale generativa (IA Generativa) a generare risposte desiderate.



Sebbene l'IA generativa cerchi di **imitare** gli esseri umani, richiede istruzioni dettagliate per creare risultati pertinenti e di alta qualità. Nel prompt engineering, sei tu a scegliere i formati, le frasi, le parole e i simboli più appropriati che guidano l'IA a interagire in modo più significativo. I prompt engineer utilizzano la creatività unita a tentativi ed errori per creare una raccolta di testi di input, in modo che l'IA generativa di un'applicazione funzioni come previsto.

# ● Prompt Engineering

## Prompt

Un prompt è un testo in linguaggio naturale che richiede all'IA generativa di eseguire un'attività specifica.

## Sviluppo del Prompt



# ● Prompt Engineering

## Sviluppo del Prompt

- **Creazione di Prompt**

Questo implica la formulazione di un input iniziale o di una domanda che guida il modello IA a produrre una risposta utile.

- **Ottimizzazione del Prompt**

Questo processo coinvolge l'aggiustamento del prompt per migliorare la qualità delle risposte del modello IA. Questo può includere la modifica della formulazione del prompt, l'aggiunta di più contesto o la specificazione del formato desiderato per la risposta.

- **Generazione e Valutazione delle Risposte**

Dopo che il modello IA ha generato una risposta, il Prompt Engineering comporta anche la valutazione della qualità della risposta. Questo può aiutare a identificare le aree in cui il prompt può essere ulteriormente ottimizzato.

- **Rivalutazione del Prompt**

In base alla valutazione della risposta, il prompt viene rivalutato. Se la risposta non è soddisfacente, il prompt viene riformulato o ottimizzato ulteriormente, creando un ciclo di feedback.

- **Ripetizione del Processo**

Il processo di ottimizzazione, generazione/valutazione e rivalutazione del prompt viene ripetuto finché non si ottiene una risposta di qualità soddisfacente.

# • Prompt Engineering

## Esempi di Prompt

### 1. Riassunto di un libro:

Riassumi il libro "Il Nome della Rosa" di Umberto Eco, mettendo in evidenza i seguenti punti:

- Trama principale
- Personaggi chiave e le loro caratteristiche
- Temi principali trattati nel libro
- Il tuo parere personale sul libro

### 2. Storia creativa:

Scrivi una storia che inizia con "Era una notte buia e tempestosa", che coinvolge i seguenti elementi:

- Un astronauta di nome Marco
- Un pianeta sconosciuto con una civiltà aliena
- Un misterioso artefatto alieno
- Un finale sorprendente

Sei un esperto nell'identificare informazioni chiave. Per favore, analizza il seguente testo, che troverai tra i due set di tre apici (''''), e individua le entità più importanti. Formatta le tue scoperte nel seguente modo:

- Nome:
- Cognome:
- Età

Ecco il testo da analizzare:

'''  
'''

# • Prompt Engineering

## Best Practice

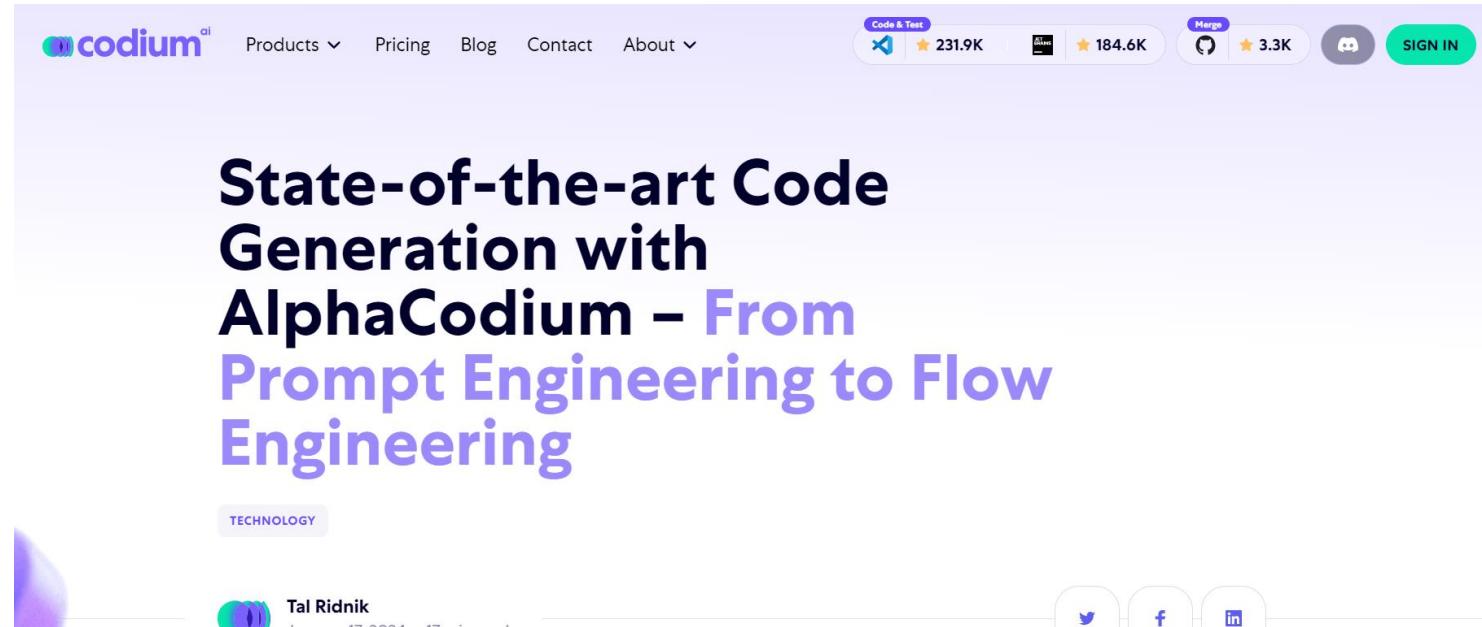
<p><b>Istruzioni inequivocabili</b> Definisci chiaramente la risposta desiderata nel tuo prompt per evitare interpretazioni errate da parte dell'IA. Ciò aiuta l'IA a concentrarsi solo sulla tua richiesta e a fornire una risposta in linea con il tuo obiettivo.</p>	<p><b>Contesto adeguato</b> Fornisci un contesto adeguato all'interno del prompt e includi i requisiti di output nell'input del prompt, limitandolo a un formato specifico.</p>
<p><b>Equilibrio di informazioni</b> Equilibra semplicità e complessità nel prompt per evitare risposte vaghe, non correlate o inaspettate. Un prompt troppo semplice può mancare di contesto, mentre un prompt troppo complesso può confondere l'IA.</p>	<p><b>Sperimenta e perfeziona il prompt</b> Il prompt engineering è un processo iterativo. Per ottimizzare la precisione e la pertinenza potrebbero essere necessari più tentativi. La flessibilità e l'adattabilità sono essenziali.</p>



# • Flow Engineering

L'AI generativa è ampiamente utilizzata per la generazione automatica di **codice** di programmazione, rappresentando una delle applicazioni più comuni in questo campo. Strumenti come **GitHub Copilot** sono esempi di questa tendenza, consentendo ai programmatore di **velocizzare** il processo di scrittura del codice. Questi in genere sono dei potenti **autocomplete**, ma un nuovo approccio sta emergendo:

Un'innovazione in questo settore è rappresentata da **AlphaCodium**, che adotta un approccio distintivo delegando **completamente** la creazione del codice al modello di AI generativa.

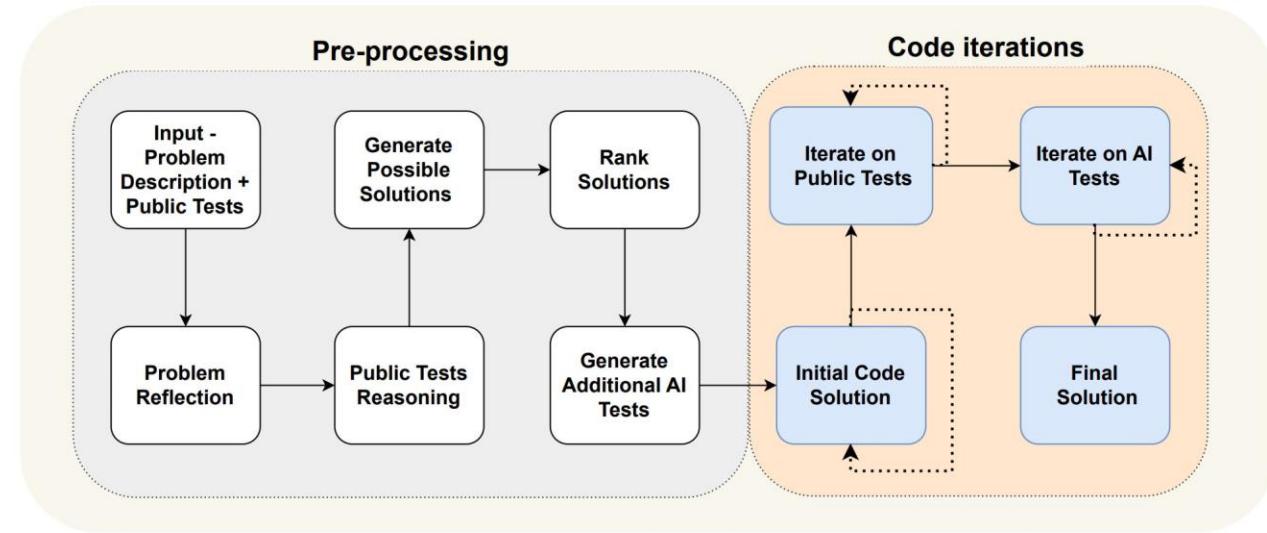


# • Flow Engineering

## Come funziona?

Si costruisce una catena di prompt:

1. "Ragiona sul problema e **crea bullet points**"
2. "Proponi possibili **soluzioni**"
3. "**Esamina le condizioni di successo** (i test)"
4. "Scrivi una **prima soluzione**"
5. "**Testala** contro le condizioni di successo"



(a) The proposed AlphaCodium flow.

Questo nuovo approccio è detto **Flow Engineering**.

Esso non si concentra su l'*ottimizzazione del singolo comando* che viene dato al modello (prompt), ma costruisce un **workflow intelligente** per sfruttare le capacità dei modelli di linguaggio.

Il Flow Engineering è stato testato con vari modelli e ha raddoppiato le performance rispetto al singolo prompt.

# FRAMEWORK

---

# Che dire di LangChain



LangChain è un *framework* progettato per sviluppare applicazioni basate su modelli linguistici.

Le applicazioni più potenti e differenziate basate su LLM non si limitano a richiamare un modello linguistico, ma combinano *diversi principi*:

- **Essere data-aware:** collegare un modello linguistico ad altre fonti di dati;
- **Essere agentic:** permettere a un modello linguistico di interagire con il suo ambiente.

# ● LangChain

*Cosa sono le catene e perché ne abbiamo bisogno?*

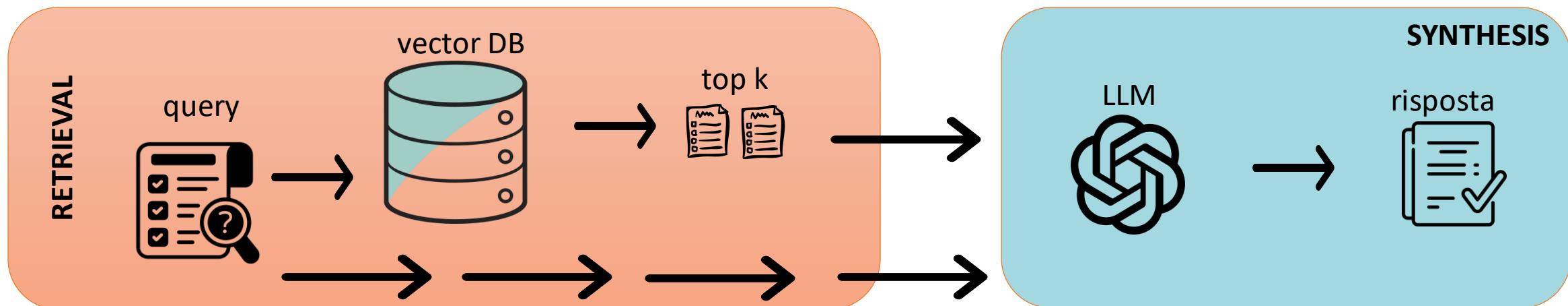
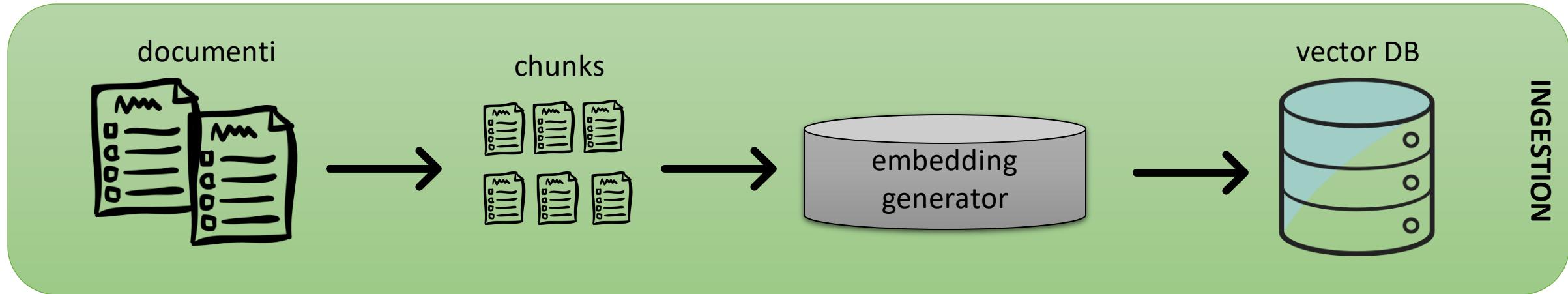
Le catene  permettono di implementare applicazioni complesse che richiedono la concatenazione di un LLM con diversi tipi di componenti che comportano sequenze di chiamate a un LLM o ad altre utilità.

Casi d'uso standard:

- Question Answering su documenti specifici
- Agenti
- ChatBot

Il [vantaggio](#) dell'utilizzo di LangChain consiste nel creare applicazioni complesse sfruttando la [stretta connessione](#) tra catene che lavorano insieme per ottenere la soluzione desiderata.

# • RAG (Retrieval-Augmented Generation)



# RISCHI E LIMITAZIONI

---

# • Bias

Nonostante le notevoli capacità di generare testo simile a quello umano, gli LLM possono ereditare e persino aggravare i **pregiudizi** presenti nei dati dai quali sono stati addestrati. Questo fenomeno può tradursi in **rappresentazioni distorte** e in un **trattamento ingiusto** di diverse categorie demografiche, come quelle basate su razza, genere, lingua e gruppi culturali.

Il termine **bias** si riferisce a una **distorsione** che inclina il sistema (sia esso il cervello umano o software) verso una specifica opzione o interpretazione.

## BIAS DI SELEZIONE

La distorsione causata dalla scelta selettiva di determinati dati.

*Esempio: Un sondaggio politico che intervista solo gli elettori di una specifica area geografica potrebbe non riflettere accuratamente l'opinione dell'intera popolazione.*

## BIAS DI CONFERMA

La tendenza a cercare, interpretare e ricordare informazioni in modo che confermino le proprie convinzioni preesistenti.

*Esempio: Se un utente chiede all'LLM informazioni su un argomento controverso, il modello potrebbe produrre una risposta che conferma le preesistenti opinioni dell'utente anziché offrire un'analisi neutrale.*

## BIAS DI GENERE

Il trattamento ingiusto o la valutazione discriminatoria basata sul genere di una persona.

*Esempio: Se si chiede all'LLM di completare la frase "Il medico è...", un modello con bias di genere potrebbe più facilmente suggerire "Il medico è un uomo" anziché "Il medico è una persona".*

# Bias

## BIAS CULTURALE

La tendenza a valutare le cose in base ai propri valori culturali, portando a una visione distorta di altre culture.

*Esempio: Se si chiede all'LLM di descrivere le caratteristiche di una famiglia ideale, il modello potrebbe incorporare stereotipi culturali specifici, come la struttura tradizionale di genere.*

## BIAS DI DISPONIBILITÀ'

La tendenza a giudicare la probabilità di un evento in base alla facilità con cui vengono richiamati esempi o informazioni correlati.

*Esempio: Se si chiede all'LLM di fornire statistiche sulla criminalità, il modello potrebbe essere influenzato da informazioni mediatiche sensazionalistiche, sovrastimando la gravità del problema.*

## BIAS DI RAZZIALE ED ETNICO

Trattamento ingiusto o discriminazione basata sulla razza o sull'origine etnica di un individuo.

*Esempio: Se si chiede all'LLM di descrivere un gruppo etnico specifico, il modello potrebbe generare risposte che incorporano stereotipi negativi presenti nei dati di addestramento.*

## BIAS DI POSITIVITÀ'

La preferenza per informazioni positive rispetto a quelle negative, influenzando la percezione e la valutazione degli eventi.

*Esempio: Se si chiede all'LLM di valutare l'impatto ambientale di un'industria, un modello con bias di positività potrebbe enfatizzare gli aspetti positivi minimizzando i problemi ambientali.*

I tentativi nella ricerca per ridurre questa problematica sono diversificati e coinvolgono svariati campi di studio, tra cui:

- **Bias Detection**
- **Bias Mitigation**
- **Interpretabilità dell'AI**

# • Limitazioni

*Comprensione  
del Contesto*

*Affidabilità*

*Controllo Output*

*Dati di  
Addestramento*

*Sicurezza*

*Interpretabilità*

*Risorse  
Computazionali*



# ● Limitazioni e Rischi

## ○ Comprensione del Contesto:

- ✓ Gli LLM possono avere difficoltà a comprendere e rispondere in modo appropriato al contesto più ampio di una conversazione o di un testo. Questo può portare a risposte che non sono pertinenti o che non tengono conto delle informazioni precedenti nella conversazione.

## ○ Dati di Addestramento:

- ✓ Gli LLM necessitano di ingenti quantità di dati per l'addestramento, ma questo processo può essere problematico a causa della presenza di dati sbilanciati, obsoleti o inadatti, con conseguente introduzione di potenziali bias nel modello.

## ○ Risorse Computazionali:

- ✓ Gli LLM sono modelli molto grandi che richiedono una notevole quantità di risorse computazionali per l'addestramento e l'esecuzione. Questo può limitare la loro applicabilità in ambienti con risorse limitate o aumentare i costi di utilizzo.

## ○ Affidabilità:

- ✓ Gli LLM possono generare output che sembrano plausibili ma che sono completamente falsi o ingannevoli. Questo può essere particolarmente problematico in contesti in cui l'accuratezza e l'affidabilità delle informazioni sono cruciali, come la medicina o il giornalismo.

# • Limitazioni e Rischi

## ○ Controllo dell'output & Proprietà Intellettuale:

- ✓ Può essere difficile controllare l'output di un LLM. Ad esempio, potrebbe essere difficile garantire che un LLM non generi contenuti inappropriati, offensivi o dannosi. Questo può richiedere ulteriori misure di sicurezza o moderazione.  
Inoltre, i modelli linguistici hanno la capacità di generare contenuti creativi, sollevando interrogativi sui diritti di proprietà intellettuale e sul plagio. Determinare l'originalità del contenuto generato e stabilire un'attribuzione adeguata diventa una sfida in tali scenari.

## ○ Interpretabilità:

- ✓ Gli LLM sono modelli complessi e non lineari, e può essere molto difficile capire perché hanno generato un certo output. Questo può rendere difficile la diagnosi e la correzione degli errori.

## ○ Sicurezza:

- ✓ Gli LLM possono essere sfruttati per scopi malevoli, come la generazione di disinformazione o contenuti dannosi. Questo richiede misure di sicurezza robuste e attenzione alla responsabilità etica nell'uso degli LLM.  
In aggiunta, interagire con gli LLM potrebbe avere implicazioni psicologiche ed emotive, soprattutto per gli individui vulnerabili.

# • Allucinazioni

Un'*allucinazione* si verifica quando un LLM genera informazioni *errate* presentandole come se fossero un dato di fatto.

---



# • Allucinazioni

 ANSA

[Ha usato ChatGpt per ricerche legali, avvocato in tribunale a NY - Hi-tech](#)

Un avvocato di New York è comparso in tribunale per rispondere dell'uso da parte della piattaforma di intelligenza artificiale ChatGPT in...

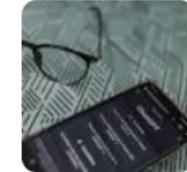
29 mag 2023



 Open

[La prima volta di ChatGPT in tribunale: a New York cita sentenze fasulle e il giudice gli dà torto](#)

Davanti ai giudici di Manhattan, Steven Schwartz ha deciso di affidare la propria difesa al chatbot. Ma... Se nel campo della medicina i primi...



29 mag 2023

 Open

[New York, multa di 5mila dollari a due avvocati per aver citato casi legali falsi consigliati da ChatGPT](#)

Schwartz ci ha tenuto a far sapere di sentirsi «mortificato» per non aver capito che non stava usando «un motore di ricerca,...

23 giu 2023



# ● Allucinazioni

## ○ Cause

- ✓ **Dati di Addestramento insufficienti, obsoleti o di bassa qualità** - Un LLM è **valido** quanto lo sono i dati su cui è stato addestrato. Se l'LLM non comprende la tua richiesta o non dispone di informazioni sufficienti, farà affidamento sul set di dati limitato su cui è stato addestrato per generare una risposta, anche se imprecisa.
- ✓ **Lacune Semantiche** - mancanza di capacità di ragionamento basate sul *buon senso*.
- ✓ **Adattamento eccessivo** - Addestramento su un set di dati limitato, può comportare la memorizzare degli input e degli output appropriati. Ciò rende l'LLM incapace di **generalizzare** in modo efficace i nuovi dati, con conseguenti allucinazioni.
- ✓ **Domande Vaghe** - se la domanda o il suggerimento di input sono ambigui, il modello potrebbe generare una risposta basata su quella che considera l'**interpretazione** più probabile, che potrebbe non essere in linea con le intenzioni dell'utente.

# • Allucinazioni

## ○ Rimedi

- ✓ **Set di dati curati** - utilizza set di dati **verificati** di alta **qualità** per perfezionare il modello. Più accurati sono i dati di addestramento, meno è probabile che il modello generi allucinazioni.
- ✓ **Filtraggio dell'output** - implementa meccanismi per **filtrare** o **contrassegnare** output potenzialmente errati o allucinanti in base a determinati criteri come la probabilità statistica o l'aderenza a un set di regole specifico del dominio.
- ✓ **Meccanismo di feedback** - Stabilire un sistema di **feedback** degli utenti in tempo reale, come l'apprendimento per rinforzo. Se il modello produce un'allucinazione, gli utenti possono segnalare le informazioni errate, che possono essere utilizzate per un'ulteriore messa a punto.
- ✓ **Definizione dell'ambito** - definire chiaramente l'**ambito** delle attività per le quali il modello è progettato per assistere e avvisare gli utenti di non fare affidamento su di esso per attività al di fuori di tale ambito.

# • Gli LLM e la Matematica

GPT-4 ha un tasso di **accuratezza** del 2% nella moltiplicazione a 5 cifre.

La maggior parte dei modelli open source **non riesce** nemmeno a fare **semplici somme**.

The screenshot shows a conversation between a user and ChatGPT. The user, identified by a blue circular icon with 'BE' and the text 'You', asks: 'Quanto fa 3367487\*2334?'. ChatGPT, identified by a green circular icon with a brain-like symbol and the text 'ChatGPT', responds: 'Il risultato della moltiplicazione 3367487 \* 2334 è 7,850,080,758.' To the right of the screenshot, the text 'Invece fa 7.859.714.658' is displayed, indicating that ChatGPT provided an incorrect result.

Il motivo è che gli LLM sono **modelli di linguaggio** e sono addestrati per capire e comprendere il **linguaggio naturale**, non fare operazioni aritmetiche.

$$2 + 2 = 5$$



@datapizza

# • Gli LLM e la Matematica

Nello specifico, il problema risiede nella **tokenizzazione**.

## La tokenizzazione

Un token è un rappresentazione numerica di una stringa di testo.

Quando si addestra un LLM, tutto il testo su cui viene allenato è **tokenizzato** e inserito in un enorme **vocabolario**, da cui si creano poi i **vettori numerici** che rappresentano parole, concetti, e così via.

Quando si prepara un tokenizzatore da zero, si prende un ampio testo e si trova la codifica a coppie di byte (sequenze di due byte consecutive presenti nel testo) minima per una dimensione di vocabolario scelta. Ciò significa che quasi certamente i numeri non avranno rappresentazioni uniche di token.

Tokens	Characters
13	39

Lasciate ogni speranza, voi ch'entrate.

[43, 5171, 6629, 55474, 47116, 27564, 11, 68752, 523, 6, 306, 7853, 13]

# • Gli LLM e la Matematica

Ciò significa che quasi certamente i numeri **non avranno rappresentazioni uniche di token**.

**21**

Potrebbe essere un singolo  
token [21] oppure [2,1]

**143**

Potrebbe essere [143], [14,3]  
o qualsiasi altra combinazione

Questo porta ad una **rappresentazione fittizia e frammentata** del concetto di numero.

Una possibile soluzione sarebbe quella di **forzare la tokenizzazione a cifra singola**.

**143**

Verrebbe tokenizzato come  
[1,4,3]

*Non è una soluzione robusta.*

# • Gli LLM e la Matematica

Gli LLM quando ricevono un input aritmetico non utilizzano strumenti di calcolo matematici, ma tentano di **predire** il prossimo token più probabile tentando di **indovinare** il risultato corretto.

## Come risolvere?

1. Usa la modalità ***Advanced Data Analysis*** di GPT4
2. Scrivi il **problema** matematico
3. Il modello scriverà il **codice Python** e ti darà il risultato

# APPLICAZIONI NEL MONDO REALE

---

# • Applicazioni



# • Applicazioni nel Mondo Reale

## ○ Chatbot/Assistenti virtuali

- ✓ Un assistente virtuale basato su LLM può migliorare l'efficienza aziendale fornendo un servizio clienti rapido, gestendo le attività di routine, ottimizzando la produttività e riducendo i costi operativi.

## ○ Generazione e cura dei contenuti:

- ✓ Gli LLM creano rapidamente contenuti di alta qualità per blog e social media, garantendo pertinenza e correttezza grammaticale. Semplificano la produzione di contenuti e migliorano l'engagement del pubblico online per le aziende.

## ○ Traduzione linguistica:

- ✓ Gli LLM traducono senza interruzioni testi tra lingue, essenziali per comunicazioni globali aziendali come marketing e comunicazioni con i clienti, garantendo precisione e adattamento culturale nei contenuti tradotti.

## ○ Generazione di codice:

- ✓ Gli LLM consentono la generazione di codice in diversi linguaggi, rendendo accessibile la creazione di base anche a chi ha un background tecnico limitato.

## ○ Analisi del sentimento:

- ✓ Gli LLM trasformano l'analisi dei sentimenti aziendali, guidando strategie di marketing, miglioramenti di prodotti e servizi, e gestione del marchio attraverso l'interpretazione delle emozioni del pubblico sui social media e recensioni dei clienti.

# • Applicazioni nel Mondo Reale

## ○ Rilevamento e prevenzione degli attacchi informatici:

- ✓ Gli LLM vengono utilizzati in tempo reale per la sicurezza informatica, consentendo il rilevamento di attacchi informatici attraverso l'analisi di ampi set di dati aziendali, identificando schemi sospetti ed emettendo avvisi tempestivi.

## ○ Rivoluzione nella trascrizione:

- ✓ Gli LLM convertono rapidamente audio e video in testi precisi. Diverse aziende utilizzano gli LLM per trascrivere e riassumere i contenuti, eliminando la necessità di trascrizione manuale e risparmiando tempo.

## ○ Narrazione:

- ✓ Gli LLM hanno rivoluzionato lo storytelling, fungendo da architetti narrativi capaci di creare o adattare storie per temi, stimoli e pubblico specifici.

## ○ Automazione delle vendite:

- ✓ Gli LLM guidano la rivoluzione dell'automazione delle vendite, personalizzando i processi dalla fase iniziale all'assistenza post-vendita attraverso l'analisi dei dati dei clienti.

## ○ Reclutamento delle risorse umane e screening dei candidati:

- ✓ Gli LLM stanno trasformando il reclutamento, facilitando lo screening iniziale dei candidati attraverso l'analisi automatizzata di curriculum e lettere di presentazione. Questo non solo risparmia tempo, ma aumenta anche l'obiettività nella valutazione, riducendo il rischio di pregiudizi umani.

# • Question Answering Aziendale

Lo **scopo** è creare un assistente virtuale AI la cui missione è:

- agire come partner collaborativo per i colleghi nelle fasi iniziali dell'assistenza e
- Facilitare una comprensione immediata della documentazione aziendale.

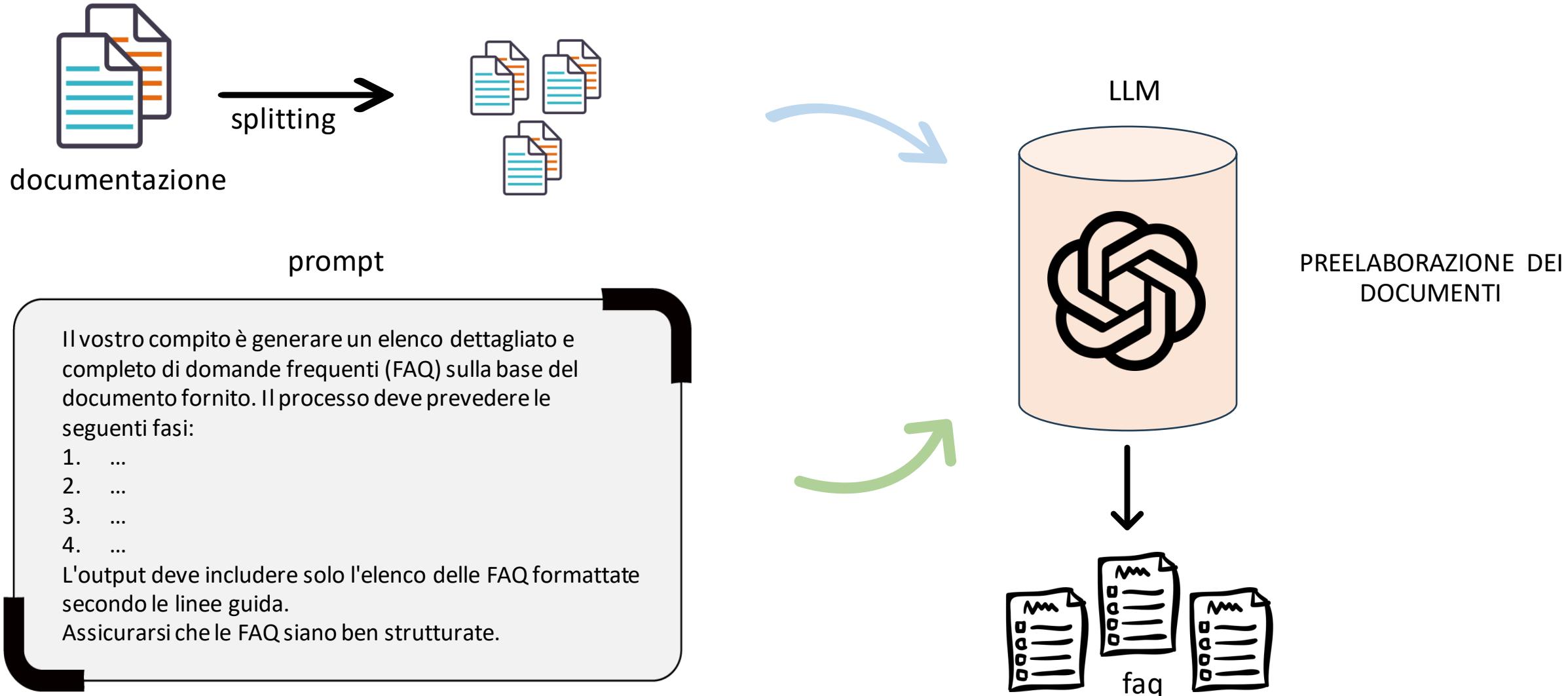
L'assistente virtuale sarà in grado di fornire risposte tempestive, mirate e costruttive, ottimizzando il flusso di lavoro e aumentando il livello di **autonomia** e **competenza**.

*Ciao!  
È un piacere  
conoserti.  
Come posso migliorare  
la tua giornata?*

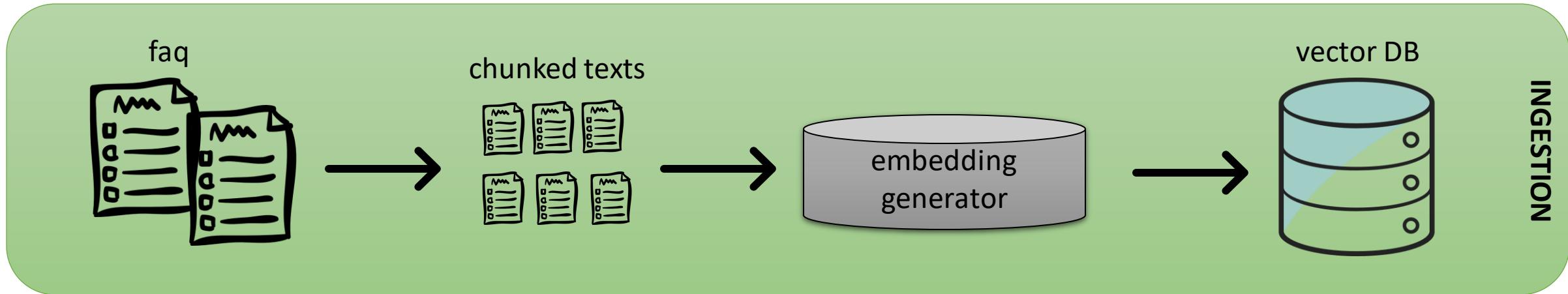


(AI GENERATED)

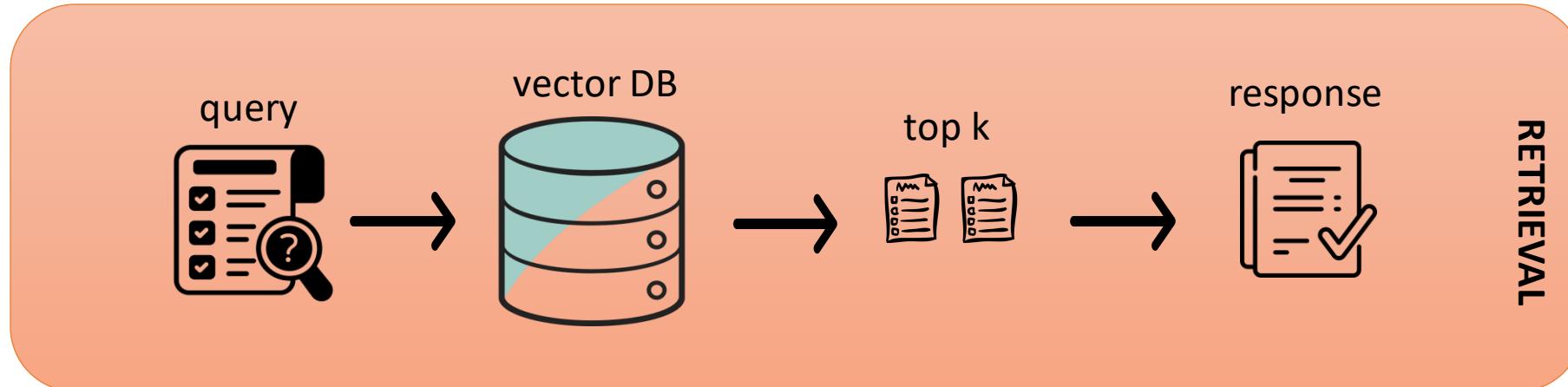
# • Question Answering Aziendale



# • Question Answering Aziendale

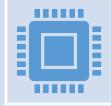


RAG PIPELINE



# • Question Answering Aziendale

## VANTAGGI



**Riduzione dei costi:** Evitando l'uso continuo di LLM per la generazione delle risposte, si riducono notevolmente i costi di calcolo e di manutenzione.



**Risposta immediata:** L'accesso diretto al database degli embeddings consente di ottenere risposte immediate senza i tempi di attesa necessari per la generazione del testo.



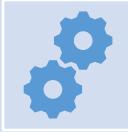
**Accuratezza nel Matching:** gli embeddings consentono un matching semantico più preciso tra le query degli utenti e le informazioni esistenti, migliorando l'accuratezza delle risposte.



**Scalabilità:** Il sistema basato su embeddings è facilmente scalabile, consentendo di gestire un volume maggiore di interazioni senza un aumento proporzionale dei costi.

# • Question Answering Aziendale

## VANTAGGI



**Semplicità operativa:** Il nostro approccio riduce la complessità del backend, semplificando la manutenzione e l'aggiornamento del sistema.



**Miglioramento della privacy:** Poiché non è necessario elaborare ogni richiesta attraverso un LLM esterno, si riduce il rischio di esposizione dei dati sensibili.



**Maggiore velocità di implementazione:** La costruzione dei set di dati e gli aggiornamenti del sistema sono più rapidi, garantendo un adattamento più veloce alle esigenze aziendali.

# **MODELLI MULTIMODALI**

---

# • Modelli Multimodali

Gli LLM **multimodali** offrono la possibilità di ampliare l'impatto dei sistemi di solo linguaggio con nuove interfacce e capacità, che consentono loro di risolvere nuovi tasks e di fornire nuove esperienze ai loro utenti, come l'analisi delle immagini fornite.

GPT-4V(ision) System Card

OpenAI

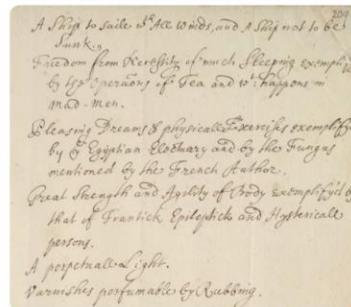
September 25, 2023

## 1 Introduction

GPT-4 with vision (GPT-4V) enables users to instruct GPT-4 to analyze image inputs provided by the user, and is the latest capability we are making broadly available. Incorporating additional modalities (such as image inputs) into large language models (LLMs) is viewed by some as a key frontier in artificial intelligence research and development [1, 2, 3]. Multimodal LLMs offer the possibility of expanding the impact of language-only systems with novel interfaces and capabilities, enabling them to solve new tasks and provide novel experiences for their users.

# • Modelli Multimodali

ETHAN



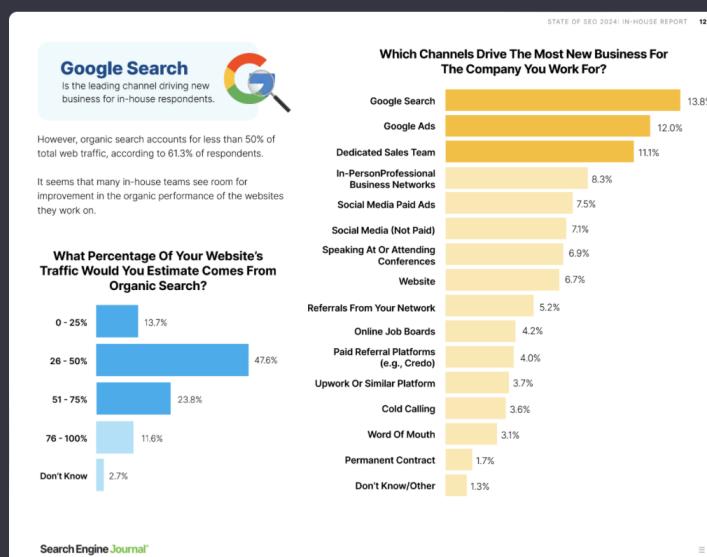
Can you read this?

CHATGPT

Yes, I can read the text in the image. Here's a transcription:

A Ship to saile without winds, and a Ship not to be sunk.  
Freedom from Necessity of much Sleeping exemplified by the Operations of Tea and what happens in Mad men.  
Pleasing Dreams & physicall Exercises exemplified by the Egyptian Electuary and by the Fungus mentioned by the French Author.  
Great Strength and Agility of Body exemplified by that of Frantick Epileptick and Hysterical persons.  
A perpetuall Light.  
Varnishes performable by Rubbing.

Please note that some of the words, especially older terms or specific names, might be challenging to decipher due to the old-fashioned handwriting, but I did my best to provide an accurate transcription.



Can you write an article about the data represented on this page?



## Understanding the State of SEO in 2024: Key Insights and Analysis

Search engine optimization (SEO) remains a pivotal aspect of the digital landscape. However, how businesses perceive and harness SEO's potential is continually evolving. A recent report from Search Engine Journal offers invaluable insights into this dynamic domain. Below, we'll delve into the report's crucial takeaways and what they might signify for businesses.

### 1. Google Search Dominates New Business Acquisition

According to the data, Google Search emerges as the leading channel for driving business, with 13.8% of in-house respondents naming it as their top source. This is followed by Google Ads at 12.0%.

Send a message

...

ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT September 25 Version

# IL FUTURO DEI LLM

---

# • Qual è il futuro dei LLM?

- **Capacità aumentate**

Per quanto siano notevoli, l'attuale livello di tecnologia non è perfetto e gli LLM non sono infallibili. Tuttavia, le versioni più aggiornate avranno una maggiore precisione e funzionalità migliori, man mano che gli sviluppatori imparano a potenziare le loro prestazioni, riducendo i pregiudizi ed eliminando le risposte errate.

- **Formazione audiovisiva**

Mentre gli sviluppatori addestrano la maggior parte degli LLM utilizzando testo, alcuni hanno iniziato ad addestrare modelli utilizzando input video e audio. Questo tipo di training dovrebbe portare a uno sviluppo più rapido del modello e aprire nuove possibilità in termini di utilizzo di LLM per veicoli autonomi.

- **Trasformazione dell'ambiente di lavoro**

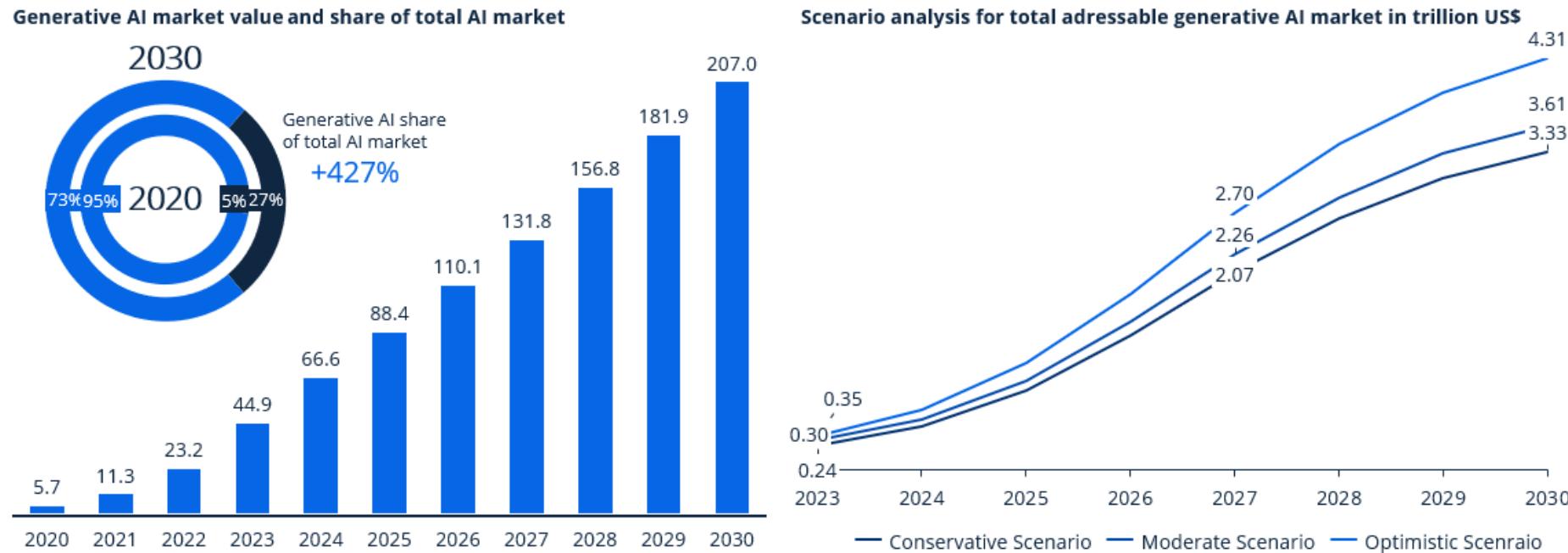
I LLM sono un fattore dirompente che cambierà l'ambiente di lavoro. I LLM ridurranno probabilmente le attività monotone e ripetitive, nello stesso modo in cui lo hanno fatto i robot per le attività di produzione ripetitive. Le possibilità includono attività d'ufficio ripetitive, chatbot per il servizio clienti e semplici operazioni di copywriting automatizzato.

- **IA conversazionale**

Gli LLM miglioreranno senza dubbio le prestazioni degli assistenti virtuali automatizzati come Alexa, Google Assistant e Siri. Saranno in grado di interpretare meglio l'intento dell'utente e rispondere a comandi sofisticati.

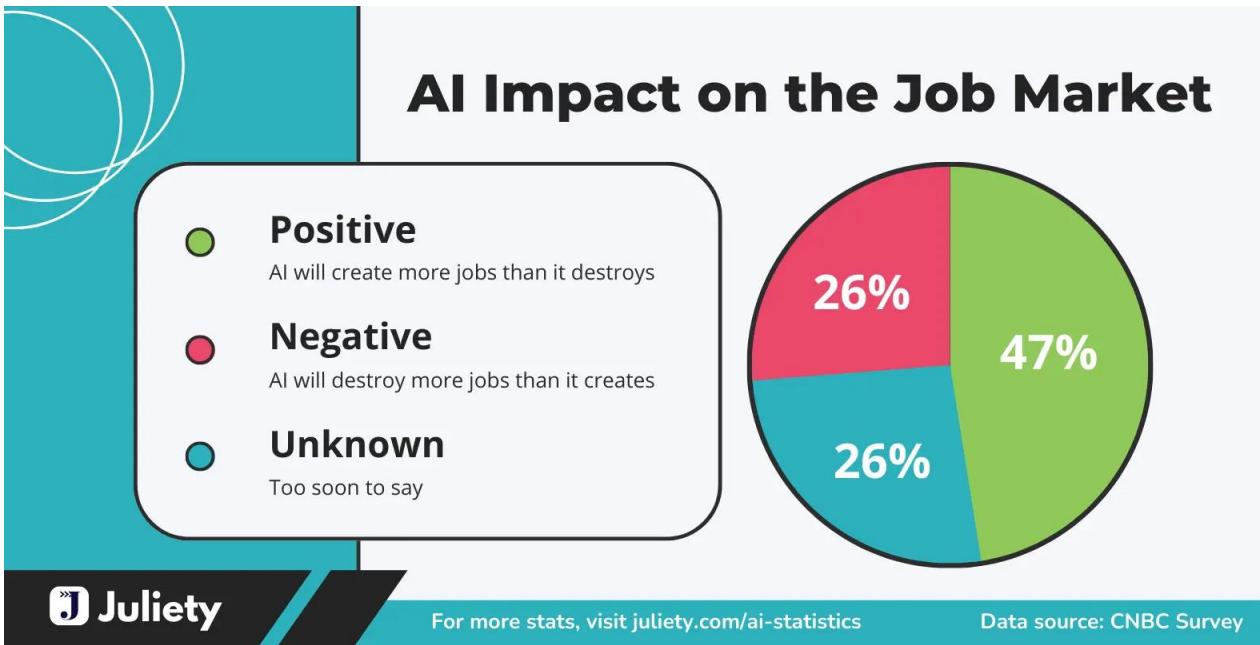
<https://aws.amazon.com/it/what-is/large-language-model/>

# ● Previsioni di Mercato



Nel panorama in continua evoluzione dei progressi dell'IA, un campo si distingue per la sua rapida ascesa e il suo potenziale di trasformazione: l'**IA generativa**. I modelli linguistici di grandi dimensioni (LLM), al centro di questa trasformazione, continuano a crescere in dimensioni e costi. Il GPT-4 non solo ha superato più del 90% degli avvocati praticanti all'esame di abilitazione, ma ha anche mostrato *prestazioni di livello umano* in una serie di valutazioni professionali. Le conseguenze sono sorprendenti, nel 2020, rappresentava il **5%** del mercato totale dell'IA, ma si prevede che arriverà al **27%** entro il 2030. L'aumento dei finanziamenti nel settore dell'IA generativa testimonia il suo promettente futuro. Inoltre, il suo valore di mercato aumenterà da **5,7 miliardi** di dollari nel 2020 a **207 miliardi** di dollari nel 2030.

# • Impatto sul mercato del Lavoro



Molte persone sono preoccupate dalla possibilità che l'IA sostituisca i posti di lavoro umani. Secondo diverse fonti di statistiche sull'intelligenza artificiale, si prevede che l'IA creerà sia nuove opportunità che sfide nel mercato del lavoro, man mano che l'automazione diventerà più diffusa. Poiché nel 2023-2024 assisteremo a un balzo unico nello sviluppo dell'IA, molti dirigenti del settore tecnologico hanno una visione ottimistica dell'impatto dell'intelligenza artificiale. Secondo il sondaggio della CNBC sui dirigenti del settore tecnologico, il 47% degli intervistati ritiene che le tecnologie AI creeranno più posti di lavoro di quanti ne elimineranno. La metà degli altri, il 26%, prevede che l'IA porterà a una maggiore perdita di posti di lavoro e un'altra metà ritiene che sia troppo presto per stabilirlo.

# • Crescita della Produttività

Agenzia Nova

IA, von der Leyen: "Può determinare un ulteriore balzo alla produttività delle imprese europee"

L'intelligenza artificiale (IA) può determinare un ulteriore balzo alla produttività delle imprese europee. Lo ha detto la presidente della...

1 settimana fa

# Matrice Digitale

**ChatGPT aumenta la produttività dei professionisti e migliora la qualità del lavoro**

Un recente studio condotto su professionisti che utilizzano ChatGPT per scrivere documenti aziendali ha evidenziato come il tempo impiegato...

3 apr 2023

Il Fatto Quotidiano

**"Con ChatGPT la mia produttività è aumentata di dieci volte". Ecco gli entusiasti dell'AI**

"Con ChatGPT la mia produttività è aumentata di dieci volte". Ecco gli entusiasti dell'AI .

"ChatGPT ti ruba la vita, è come una droga. Per...

18 mag 2023

Wired Italia

**Intelligenza artificiale: avrà un impatto sul 40% dei lavori**

Lo evidenzia una ricerca del Fondo monetario internazionale (Fmi), secondo cui la tecnologia avrà conseguenze molti forti nelle economie...

3 settimane fa



BusinessCommunity.it

**Pietrobon (Microsoft): Copilot per Microsoft 365 aumenta la produttività e migliora la vita professionale**

Pietrobon (Microsoft): Copilot per Microsoft 365 aumenta la produttività e migliora la vita professionale.

5 giorni fa

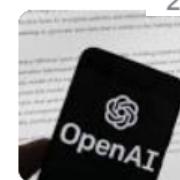


AGI

**Con l'intelligenza artificiale 1 persona su 3 potrebbe lavorare solo 4 giorni**

Secondo il rapporto del Think tank Autonomy riguarderebbe 8,8 milioni di persone in Gran Bretagna e 35 milioni negli Stati Uniti,...

22 nov 2023

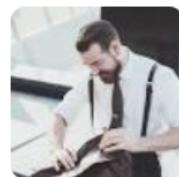


C Collettiva

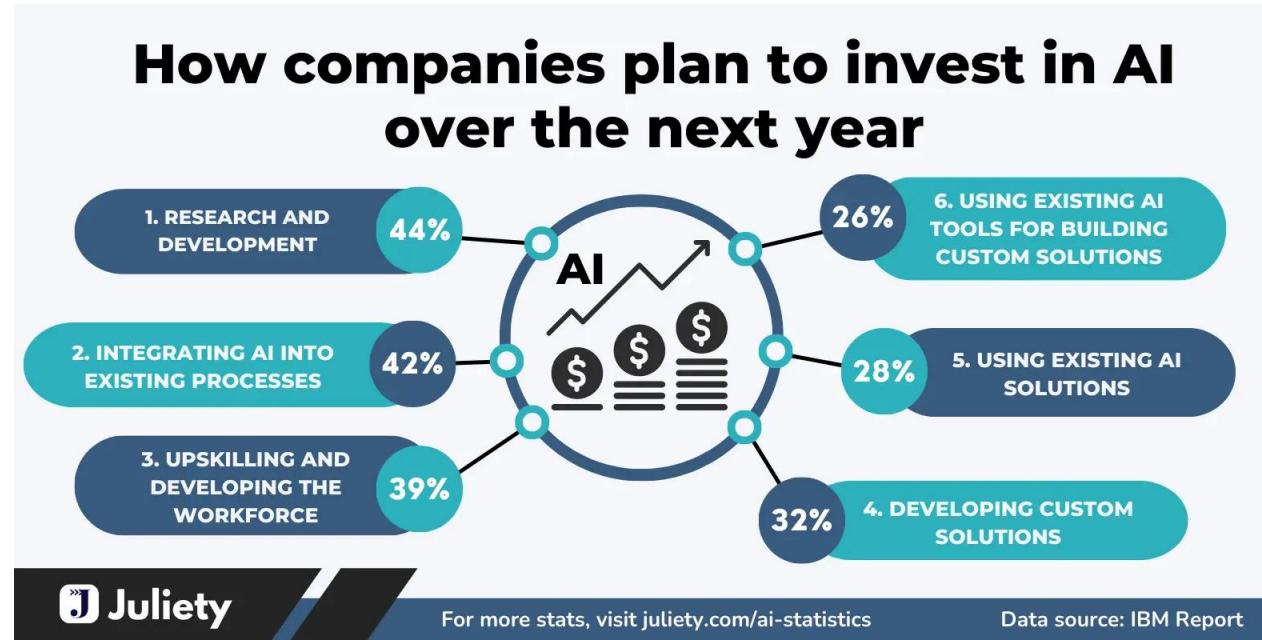
**Per lavorare abbiamo (già) bisogno di ChatGPT?**

In modo interessante, gli autori sostengono che l'uso di ChatGPT riduce le disuguaglianze tra i lavoratori, poiché beneficia maggiormente coloro...

26 lug 2023



# • Investimenti in AI



L'80% dei CEO intende incorporare le tecnologie AI nelle proprie attività aziendali entro due anni (Xometry). Il 31% delle organizzazioni intervistate ha dichiarato di voler aumentare la spesa per l'IA e l'apprendimento automatico nel 2023 e oltre, alla luce dell'attuale contesto economico (BusinessWire). L'adozione dell'IA da parte dei leader aziendali è aumentata in modo significativo, con il 61% che sta abbracciando attivamente o sta già incorporando l'IA, segnando un notevole aumento rispetto al 51% riportato all'inizio del 2023 (Xometry). La percentuale di aziende che vedono un ritorno sugli investimenti in IA è aumentata notevolmente dal 2017 al 2023, raggiungendo il 92,1% rispetto al 48,1% (NewVantage).

- **Licenziamenti dovuti all'AI**

Tech

## **Centinaia di licenziamenti a Google, pesa la concorrenza dell'Intelligenza artificiale**

*L'annuncio dopo una decisione simile di Amazon. L'avvocato della Corte Ue conferma la multa da 2,4 miliardi per l'azienda di Mountain View. Google: "Investiamo in modo responsabile sulle nostre priorità"*

# GRAZIE

Luca Bellanova

