

Creare prompt efficaci con ChatGPT

Che cos'è ChatGPT

LLM OVVERO MODELLI LINGUISTICI DI GRANDI DIMENSIONI

I LLM, acronimo di *Large Language Models*, sono modelli di intelligenza artificiale progettati per **analizzare** e **generare** il linguaggio umano in modo avanzato. Questi modelli sono addestrati su vasti *corpora* di testi e usano **reti neurali profonde** per processare e generare testi in modo coerente e semantico.

Questi modelli hanno una vasta gamma di applicazioni, tra cui assistenza nella scrittura, traduzione automatica, chatbot, generazione di contenuti e molto altro.

ChatGPT è un esempio di LLM e rappresenta una delle implementazioni specifiche basate sulla tecnologia *GPT-3.5* oppure *GPT-4*. Si tratta di un modello di conversazione avanzato che può analizzare e generare testi in modo coerente e naturale, in risposta a domande o istruzioni fornite in linguaggio umano.

COME CHATGPT OPERA SUI TESTI

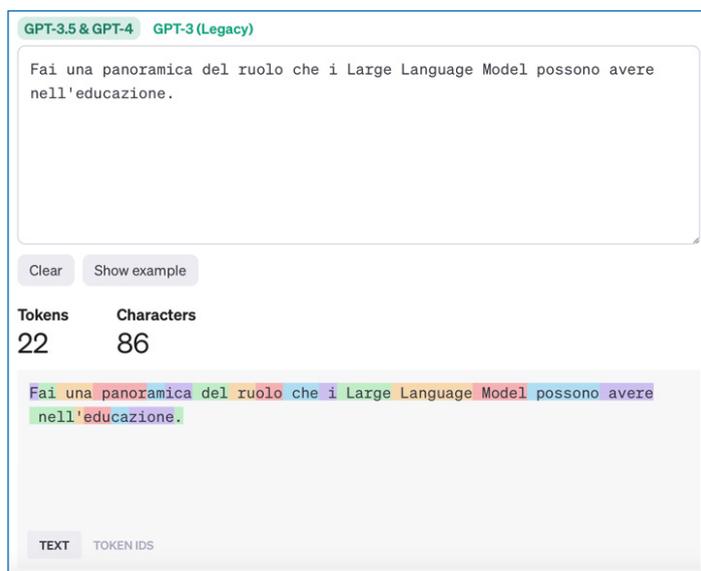
ChatGPT opera sui testi suddividendoli in particelle dette **token**. I token non sono né parole, né sillabe, né lettere; possiamo vedere un esempio di frase spezzata in token se osserviamo la figura a fianco. La frase *Fai una panoramica del ruolo che i Large Language Model possono avere nell'educazione.* viene suddivisa in token, ciascuno dei quali corrisponde a un blocchetto colorato.

Il concetto di token è importante per capire come *ChatGPT* operi per generare nuovo testo mediante l'autoregressione.

L'**autoregressione** è un concetto fondamentale nei modelli linguistici. Consiste nell'idea che il modello generi testo in modo sequenziale, un passo alla volta, in un processo in cui ogni passo dipende dalla generazione dei passi precedenti. In altre parole, il modello crea il testo un token alla volta, utilizzando i token generati in precedenza per guidare la generazione futura.

Quando forniamo un input a *ChatGPT*, il modello inizia a generare una risposta token per token. A ogni passo, il modello sceglie il token successivo sulla base del testo disponibile fino a quel momento nella conversazione. La generazione di nuovo testo non può considerare l'intera conversazione se questa è troppo lunga; infatti il testo considerato è costituito soltanto dagli ultimi 4096 (o 16384 o 32768) token, a seconda della versione di *ChatGPT*.

Il processo di generazione continua fino a quando il modello completa la generazione della risposta.



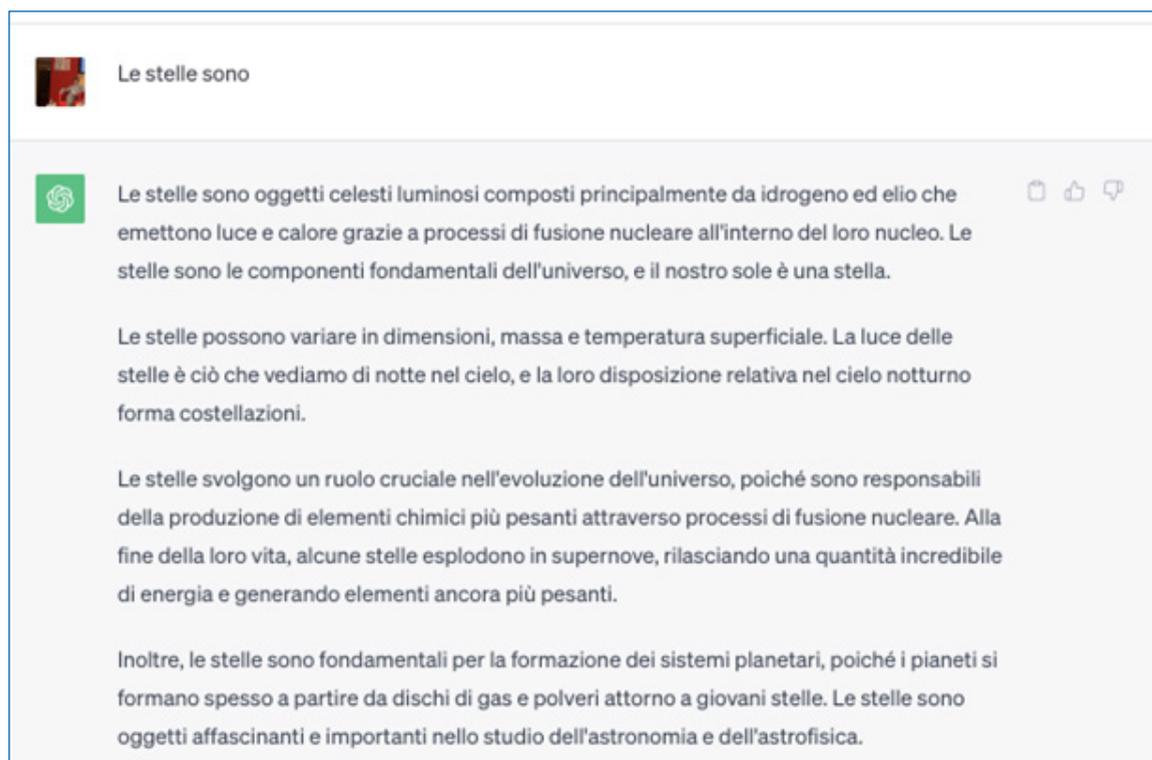
CHATGPT COMPRENDE IL TESTO?

L'autoregressione consente a *ChatGPT* di generare testo in modo coerente e contestualmente rilevante. Tuttavia, è importante notare che il modello non ha una comprensione completa del contesto o della conversazione in quanto non "ricorda" nel senso umano, ma si basa sulle informazioni immediate a sua disposizione e quelle apprese durante il suo addestramento. Ciò significa che può essere sensibile all'ordine e alla formulazione delle frasi nell'input, e talvolta potrebbe non generare risposte sempre perfette o precise, specialmente in conversazioni molto lunghe o complesse.

Possiamo affermare che i LLM come GPT mostrano una straordinaria competenza linguistica formale, cioè la capacità di generare linguaggio grammaticalmente fluente e simile a quello umano, ma mancano ancora della comprensione concettuale necessaria per abilità linguistiche funzionali simili a quelle umane, cioè la capacità di comprendere e utilizzare il linguaggio in modo robusto nel mondo reale, di cui non hanno percezione. Per questo motivo, nonostante *ChatGPT* sappia generare testi capaci di sorprenderci, è improprio parlare di creatività nel senso che attribuiamo agli esseri umani.

Regole generali per creare prompt efficaci

Le domande o più in genere il testo che sottoponiamo a *ChatGPT* sono dette **prompt**. Il testo generato da *ChatGPT* è detto **output**. Nella figura seguente abbiamo un esempio di un semplice prompt (*Le stelle sono*) che genera un output articolato.



The screenshot shows a chat window with a user prompt and a detailed AI response. The prompt is "Le stelle sono". The response is a comprehensive paragraph about stars, their composition, and their role in the universe.

Prompt: Le stelle sono

Response:

Le stelle sono oggetti celesti luminosi composti principalmente da idrogeno ed elio che emettono luce e calore grazie a processi di fusione nucleare all'interno del loro nucleo. Le stelle sono le componenti fondamentali dell'universo, e il nostro sole è una stella.

Le stelle possono variare in dimensioni, massa e temperatura superficiale. La luce delle stelle è ciò che vediamo di notte nel cielo, e la loro disposizione relativa nel cielo notturno forma costellazioni.

Le stelle svolgono un ruolo cruciale nell'evoluzione dell'universo, poiché sono responsabili della produzione di elementi chimici più pesanti attraverso processi di fusione nucleare. Alla fine della loro vita, alcune stelle esplodono in supernove, rilasciando una quantità incredibile di energia e generando elementi ancora più pesanti.

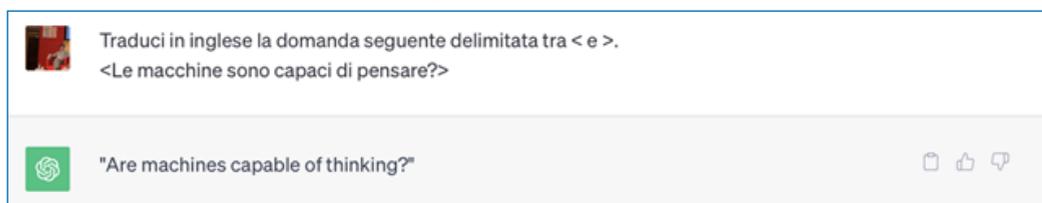
Inoltre, le stelle sono fondamentali per la formazione dei sistemi planetari, poiché i pianeti si formano spesso a partire da dischi di gas e polveri attorno a giovani stelle. Le stelle sono oggetti affascinanti e importanti nello studio dell'astronomia e dell'astrofisica.

L'output di *ChatGPT* dipende fortemente dai prompt che gli sottoponiamo, per cui è importante scrivere prompt efficaci. In questo paragrafo vedremo alcuni principi guida che stanno alla base delle buone pratiche nella scrittura dei prompt.

Il testo che usiamo come prompt può contenere i seguenti componenti:

- **Istruzione:** si tratta del compito o dell'istruzione specifica che *ChatGPT* deve eseguire.
- **Contesto:** è costituito da tutte le informazioni esterne che possono indirizzare il modello verso risposte migliori; queste informazioni possono definire una situazione specifica che *ChatGPT* deve considerare per rispondere in maniera ottimale.
- **Dati in input:** si tratta della domanda per la quale ci interessa trovare una risposta oppure di dati necessari a *ChatGPT* per generare la sua risposta.
- **Indicatore dell'output:** definisce le caratteristiche dell'output, oppure il suo formato.

Ognuno di questi componenti è descritto da una o più parole che inseriamo nel prompt. Non tutti questi componenti sono necessari per un prompt: talvolta può bastare un'istruzione, talvolta occorre usarli tutti. Vediamo alcuni esempi.

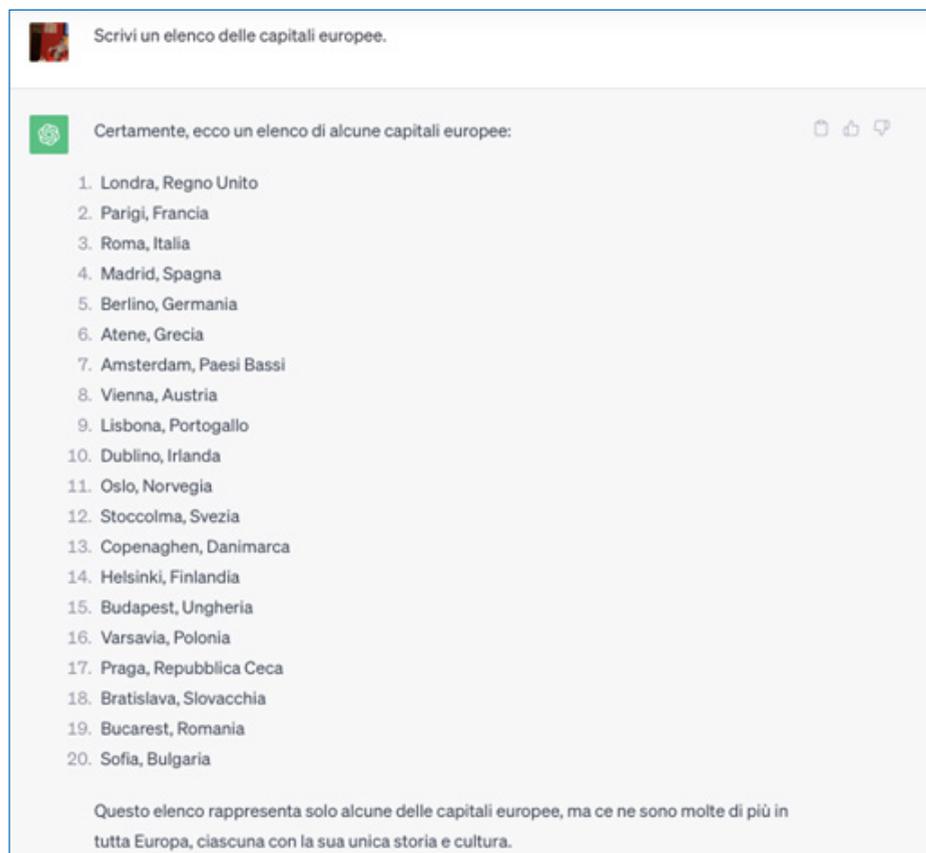


In questo esempio abbiamo usato l'**istruzione** *Traduci* e abbiamo fornito come **dati in input** la frase *Le macchine sono capaci di pensare?*

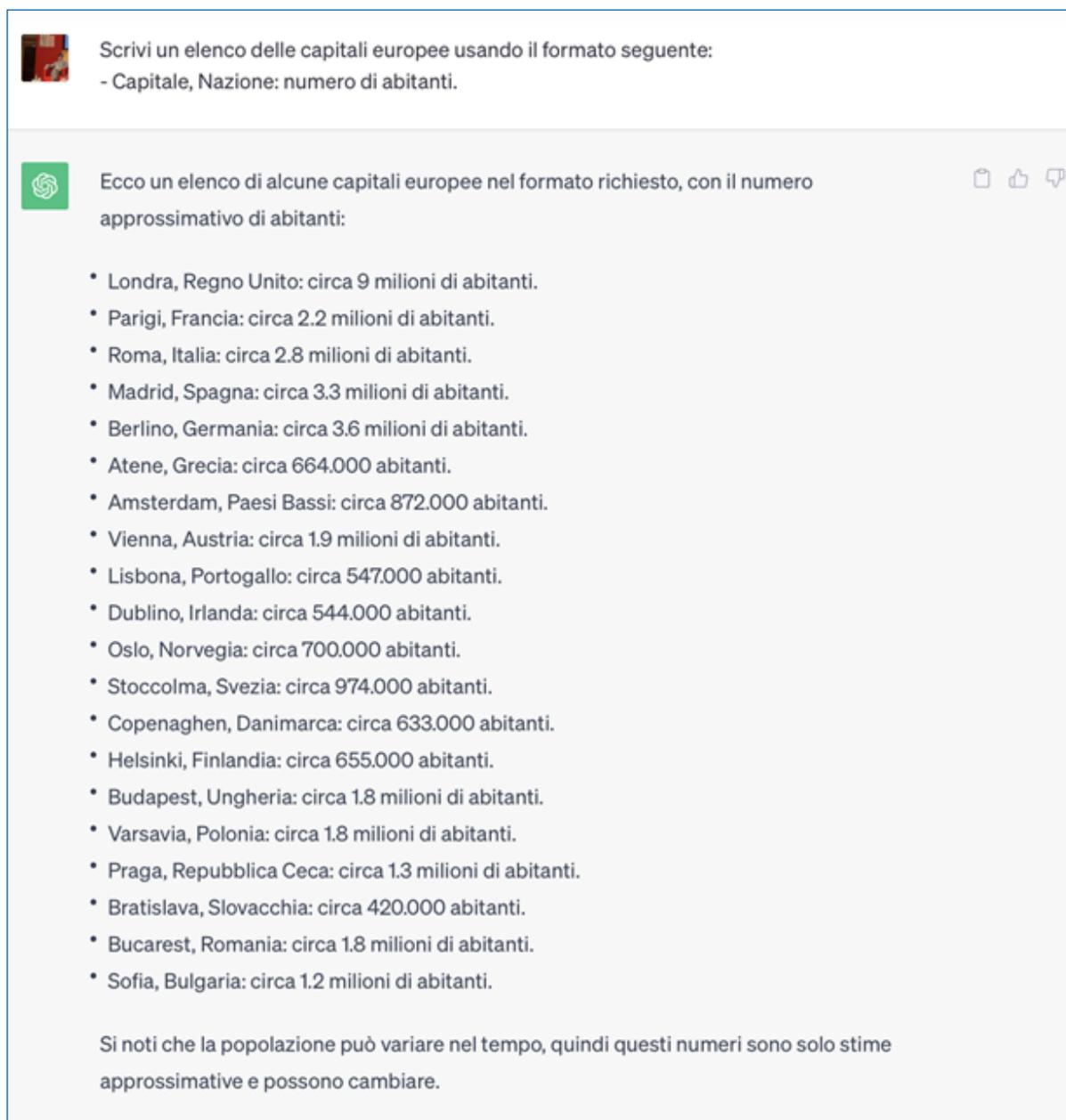
Abbiamo delimitato questa frase con i caratteri < e > per suddividere in maniera chiara la nostra richiesta dalla frase data in input.

Possiamo delimitare i dati in input con caratteri delimitatori diversi, per esempio "" o ||, a condizione che essi non siano presenti nella parte rimanente del prompt.

Nell'esempio seguente chiediamo a *ChatGPT* l'elenco delle capitali europee.



Possiamo modificare il prompt specificando un **indicatore di output** che definisca il **formato** e le **informazioni** che vogliamo ottenere.



Scrivi un elenco delle capitali europee usando il formato seguente:
- Capitale, Nazione: numero di abitanti.

Ecco un elenco di alcune capitali europee nel formato richiesto, con il numero approssimativo di abitanti:

- Londra, Regno Unito: circa 9 milioni di abitanti.
- Parigi, Francia: circa 2.2 milioni di abitanti.
- Roma, Italia: circa 2.8 milioni di abitanti.
- Madrid, Spagna: circa 3.3 milioni di abitanti.
- Berlino, Germania: circa 3.6 milioni di abitanti.
- Atene, Grecia: circa 664.000 abitanti.
- Amsterdam, Paesi Bassi: circa 872.000 abitanti.
- Vienna, Austria: circa 1.9 milioni di abitanti.
- Lisbona, Portogallo: circa 547.000 abitanti.
- Dublino, Irlanda: circa 544.000 abitanti.
- Oslo, Norvegia: circa 700.000 abitanti.
- Stoccolma, Svezia: circa 974.000 abitanti.
- Copenaghen, Danimarca: circa 633.000 abitanti.
- Helsinki, Finlandia: circa 655.000 abitanti.
- Budapest, Ungheria: circa 1.8 milioni di abitanti.
- Varsavia, Polonia: circa 1.8 milioni di abitanti.
- Praga, Repubblica Ceca: circa 1.3 milioni di abitanti.
- Bratislava, Slovacchia: circa 420.000 abitanti.
- Bucarest, Romania: circa 1.8 milioni di abitanti.
- Sofia, Bulgaria: circa 1.2 milioni di abitanti.

Si noti che la popolazione può variare nel tempo, quindi questi numeri sono solo stime approssimative e possono cambiare.

Possiamo notare che *ChatGPT* si attiene fedelmente al formato di output che gli abbiamo indicato.

Vediamo ora un ultimo esempio in cui inseriamo un **contesto**. Nei due prompt della chat seguente abbiamo chiesto a *ChatGPT* di scrivere la definizione di mole, prima impersonando un docente di chimica, poi uno studente alle prime armi con la chimica.

Possiamo osservare che la forma linguistica è molto diversa tra i due output ottenuti.

 Sei un insegnante di chimica, scrivi una definizione di mole.

 La mole è una unità di misura fondamentale in chimica che rappresenta una quantità di sostanza. Essa è definita come la quantità di sostanza di un sistema che contiene un numero di entità elementari, come atomi, molecole, ioni o altre particelle, pari al numero di Avogadro, che è approssimativamente $6,022 \times 10^{23}$. Questo numero è noto come il "numero di Avogadro".

La mole fornisce un modo conveniente di esprimere le quantità di sostanze chimiche in laboratorio, consentendo di stabilire relazioni tra la massa di una sostanza e il numero di particelle che essa contiene. Ad esempio, una mole di atomi di idrogeno conterrà circa $6,022 \times 10^{23}$ atomi di idrogeno e avrà una massa pari alla massa atomica dell'idrogeno in grammi (circa 1 grammo). Quindi, la mole è una parte fondamentale della teoria atomica e della chimica moderna, consentendo ai chimici di eseguire calcoli e misure più agevolmente.

 Sei uno studente alle prime armi con la chimica: scrivi una definizione di mole.

 Certamente! Se sei uno studente alle prime armi con la chimica, ecco una definizione semplice della mole:

La mole è un'unità di misura in chimica che ci aiuta a contare il numero di particelle (come atomi o molecole) in una sostanza. Un mole rappresenta una quantità molto grande: è il numero di particelle in una massa di una sostanza uguale alla sua massa atomica o molecolare espressa in grammi. Il numero di particelle in un mole è chiamato "numero di Avogadro" ed è approssimativamente $6,022 \times 10^{23}$.

In sostanza, la mole è come un "numero speciale" che ci aiuta a fare calcoli e comprendere meglio le quantità di sostanze chimiche con cui lavoriamo in laboratorio.

Per concludere elenchiamo alcune **buone pratiche per scrivere prompt efficaci**.

- Possiamo usare un **processo iterativo** per arrivare a ottenere prompt ottimali sperimentando: una strategia è iniziare con prompt semplici e migliorarli progressivamente, per esempio aggiungendo alcuni dei componenti visti poc'anzi.
- Dobbiamo impartire sempre **istruzioni chiare**, per esempio: *scrivi, classifica, riassumi, traduci*.
- Proviamo sempre a essere molto specifici riguardo all'istruzione e al **compito** che vogliamo che *ChatGPT* esegua; cerchiamo quindi di ripulire i nostri prompt da eventuali imprecisioni.
- Evitiamo di scrivere nei prompt cosa non fare, scriviamo invece **cosa fare**.