**Esercizi AI Academy giorno 4.**

ESERCIZIO – CHIEDI CHATGPT

Scrivi un prompt per definire prompt engineering. Formula bene, aspettiamo risultati professionali.

Domanda:

*Spiega in modo descrittivo, articolato e professionale cosa è il prompt engineering. In particolare, le sue funzioni, la sua utilità e come utilizzarlo al meglio*

Risposta:

Il *Prompt Engineering* è una disciplina emergente nell’ambito dell’Intelligenza Artificiale (AI) e, in particolare, nell’utilizzo di modelli di linguaggio di grandi dimensioni (Large Language Models, LLMs) come quelli sviluppati da OpenAI, Google o Anthropic. Consiste nell’arte e nella tecnica di progettare, formulare e ottimizzare le richieste (prompt) rivolte a questi modelli al fine di ottenere risposte il più possibile accurate, rilevanti e coerenti con gli obiettivi dell’utente.

**Funzioni del Prompt Engineering**

1. **Tradurre esigenze umane in input comprensibili dall’AI**  
   I modelli di linguaggio non comprendono il contesto umano come farebbe un interlocutore umano. Il Prompt Engineering consente di *tradurre* un bisogno — informativo, creativo, operativo — in una forma che il modello possa interpretare correttamente.
2. **Guidare il comportamento del modello**  
   Variando il modo in cui un prompt è formulato, si possono influenzare il tono, la struttura, la lunghezza e la specificità della risposta generata. Per esempio, chiedere "Scrivi una sintesi tecnica di questo testo" produrrà un output diverso rispetto a "Rendi il testo adatto a un pubblico non esperto".
3. **Ottimizzare l’efficienza operativa**  
   Un buon prompt consente di ridurre il numero di iterazioni necessarie per ottenere la risposta desiderata. In contesti aziendali questo equivale a maggiore produttività, minori costi e tempi più brevi.
4. **Sfruttare capacità avanzate dei modelli**  
   Il Prompt Engineering consente di attivare funzionalità avanzate come la capacità di ragionamento, la gestione di task complessi (multi-step), la scrittura di codice o la creazione di contenuti creativi, che non emergerebbero da prompt generici o mal formulati.

**Utilità del Prompt Engineering**

* **Per i professionisti e le aziende**  
  Permette di automatizzare e migliorare attività in numerosi settori: creazione di contenuti, analisi dei dati, assistenza clienti, sviluppo software, consulenza legale, ricerca scientifica, marketing, ecc.
* **Per il miglioramento delle interfacce uomo-macchina**  
  Migliora l’accessibilità e l’usabilità dei modelli AI per utenti non esperti, facilitando l’interazione in linguaggio naturale.
* **Per lo sviluppo di applicazioni su misura**  
  Consente di personalizzare e specializzare l’output del modello per specifiche esigenze, attraverso prompt statici, prompt dinamici o catene di prompt orchestrati (*prompt chaining*).

**Come utilizzare il Prompt Engineering al meglio**

1. **Definire chiaramente l’obiettivo**  
   Prima di scrivere un prompt, è fondamentale avere ben chiaro che tipo di risultato si vuole ottenere: un riassunto? un consiglio strategico? un codice eseguibile? un testo persuasivo?
2. **Specificità e contestualizzazione**  
   Quanto più un prompt è specifico e contestualizzato, tanto maggiore è la probabilità di ottenere una risposta utile. Includere dettagli come il pubblico di riferimento, il tono desiderato o il formato atteso.
3. **Strutturare il prompt**  
   Usare una struttura ordinata con istruzioni chiare. Per esempio:
   * Contesto iniziale
   * Compito richiesto
   * Criteri di qualità dell’output
   * Eventuali esempi o vincoli
4. **Iterazione e testing**  
   Il Prompt Engineering è un processo iterativo. È utile testare versioni diverse dello stesso prompt, analizzare i risultati e perfezionare progressivamente la formulazione.
5. **Sfruttare tecniche avanzate**  
   Alcuni approcci collaudati:
   * *Few-shot prompting*: fornire esempi nel prompt per “istruire” il modello
   * *Chain-of-thought prompting*: chiedere di esplicitare i passaggi logici
   * *Role-based prompting*: assegnare un ruolo al modello (es. “Agisci come un avvocato esperto in diritto commerciale”)
6. **Adattamento al modello utilizzato**  
   Ogni modello ha peculiarità diverse; ciò che funziona con un modello GPT può non funzionare con altri LLM. Occorre sempre testare e adattare i prompt in funzione del modello e della versione.

In conclusione, il Prompt Engineering rappresenta oggi una competenza chiave per chi desidera sfruttare in modo consapevole e strategico le potenzialità dei modelli di AI generativa. È un vero e proprio ponte tra l’intelligenza artificiale e le esigenze umane, e una leva fondamentale per aumentare l’efficacia e l’efficienza dei sistemi basati su LLM.

QUIZ: RICONOSCI IL TIPO DI PROMPT

Quiz: Che tipo di prompt è?

Leggi questi prompt e indica se sono:

A.Instruction (1, 6)

B.Few-shot (2, 4)

C.Chain-of-thought (3, 5)

1. Scrivi un’email formale per richiedere informazioni su un corso universitario.

2. Completa la sequenza:  
 Cane → Dog  
 Gatto → Cat  
 Cavallo → …

3. Risolvi il problema: Se ho 20 euro e spendo 7 euro per il pranzo e 4 euro per un libro, quanti soldi mi restano? Spiega ogni passaggio del calcolo.

4. Traduci queste parole:  
 Sole → Sun  
 Luna → Moon  
 Mare → …  
 Cielo → …

5. Trova il numero mancante nella serie 2, 4, 8, 16, \_\_\_. Mostra come arrivi alla soluzione.

6. Scrivi una poesia di 4 righe sul tema dell’inverno.

ESERCIZIO: COSTRUISCI UN PROMPT COT

Task:  
Immagina di dover aiutare un compagno a risolvere questo problema:

Un’azienda acquista 5 computer a 720 euro ciascuno e riceve uno sconto totale di 400 euro sull’intero acquisto. Quanto ha pagato in tutto?

Istruzione:  
Costruisci un prompt per ChatGPT che chieda di risolvere il problema spiegando tutti i passaggi, non solo il risultato finale.  
Deve essere un prompt chiaro, completo e specifico per ottenere una risposta “step-by-step”.

Domanda:

*Illustra, passaggio per passaggio, la soluzione al seguente quesito: "Un’azienda acquista 5 computer a 720 euro ciascuno e riceve uno sconto totale di 400 euro sull’intero acquisto. Quanto ha pagato in tutto?"*

Risposta:

**Passaggio 1: Calcolare il costo totale senza sconto**

Ogni computer costa 720 euro.  
L’azienda ne acquista 5.

Quindi, il costo totale senza sconto si calcola così:

Costo totale senza sconto=5×720 euro=3600

**Passaggio 2: Calcolare il costo totale con lo sconto**

L’azienda riceve uno sconto complessivo di 400 euro.  
Per ottenere il costo effettivo da pagare, si sottrae lo sconto dal costo totale:

Costo totale con sconto=3600 euro−400 euro=3200 euro

**Risultato finale**

L’azienda ha pagato in tutto: 3200 euro

ESERCIZIO - PROMPT TREE-OF-THOUGHT

Quando potrebbe servirvi una risposta strutturata Tree-of-Thought nella vostra attività?

Scrivi due casi d’uso da qualsiasi settore.

Scrivi il prompt per rispondere.

Domanda:

*Quanto potrebbe essere utile una risposta struttura Tree-of-Thought nell'attività da AI Engineer?*

Risposta:

**Cos’è una risposta "Tree-of-Thought" (ToT)?**

Si tratta di una tecnica di prompting avanzata in cui si chiede al modello non di generare subito la risposta finale, ma di **esplorare varie strade di ragionamento** (rami del "Tree"), analizzarle e poi scegliere la più valida o combinare i risultati.

Esempio semplificato:

* Non chiedo: "Qual è la soluzione a questo problema?"
* Ma chiedo:  
  "Esplora almeno 3 possibili approcci a questo problema, descrivili in modo argomentato e poi scegli quale ritieni il migliore, spiegandone i motivi."

**Come è utile per l’attività di un AI Engineer?**

**1️ Per progettare sistemi AI più affidabili**

Come AI Engineer, spesso devi garantire che il modello:

* sia *robusto*,
* minimizzi errori logici,
* dia risposte verificate.

Usare ToT aiuta perché:

* forza il modello a ragionare in modo più trasparente,
* permette di *seguire* i passaggi che portano alla risposta,
* aumenta la probabilità che la soluzione sia corretta, soprattutto in problemi complessi.

**2️ Per task di reasoning avanzato**

In task come:

* risoluzione di problemi matematici,
* generazione di codice,
* pianificazione multi-step,
* domande di conoscenza avanzata,  
  la semplice generazione "one-shot" spesso non basta.

ToT guida il modello a "pensare prima di rispondere", riducendo errori o allucinazioni.

**3️ Per il design di pipeline AI più performanti**

Quando progetti pipeline o applicazioni che usano LLM:

* puoi strutturare i prompt per usare ToT automaticamente,
* puoi combinare ToT con **prompt chaining** o **memory** per raffinare l’output finale.

In questo modo migliori il comportamento generale del sistema AI.

**4️ Per il fine-tuning o valutazione dei modelli**

Quando fine-tuni un modello o crei set di benchmark:

* puoi chiedere esplicitamente al modello di generare reasoning step-by-step e Tree-of-Thought,
* questo ti dà materiale migliore per supervisionare e correggere il comportamento durante il fine-tuning.

**Come usarlo al meglio**

* **Prompt ben strutturati:** devi esplicitare che vuoi diverse alternative, con ragionamento per ognuna.
* **Contesto chiaro:** serve spiegare *perché* è utile esplorare più strade.
* **Controllo sui rami:** decidi se vuoi che il modello scelga da solo o che l’utente finale selezioni.
* **Eventuale chaining:** puoi usare le ToT per istruire successivi passaggi della pipeline.

ESERCIZIO – REACT-LITE

Definizione:  
ReAct-lite utilizza i vantaggi del ReAct in modo più semplice e compatto.

Esercizio:

Crea un prompt ben pensato, che da un prompt ReAct traduce la versione lite.

Devi testare e valutare se il prompt funziona bene.

Il prompt da modificare fa parte del prompt stesso.

Domanda:

*Qual è il ruolo di un AI Engineer e come potrebbe utilizzare un prompt che traduce un prompt ReAct in una versione lite? Puoi anche scrivermi un esempio pratico di prompt per svolgere questo compito?*

Risposta:

**📌 Ruolo di un AI Engineer**

Un *AI Engineer* si occupa di progettare, sviluppare, integrare e ottimizzare sistemi di Intelligenza Artificiale.  
Rispetto ad altri ruoli più orientati alla ricerca o al puro sviluppo (come il Data Scientist o il ML Engineer), l’AI Engineer si concentra sul **mettere l’AI in produzione**, ovvero:

* integrare modelli AI in applicazioni aziendali e prodotti software,
* creare pipeline di inferenza,
* curare l’efficienza, l’affidabilità e la scalabilità del sistema,
* fare prompt engineering avanzato,
* monitorare le performance e fare tuning dei modelli,
* integrare LLM con strumenti esterni (API, DB, tool di reasoning).

Un AI Engineer deve quindi:  
👉 saper sfruttare le capacità dei LLM in modo *modulare* e *ottimizzato per l’uso reale*.

**📌 Prompt ReAct e versione "lite"**

**ReAct** (Reasoning + Acting) è una tecnica di prompting che combina:

* *Reasoning*: far ragionare il modello step by step,
* *Acting*: far interagire il modello con strumenti esterni (API, DB, Web scraping...).

Esempio:  
"Decidi come risolvere il problema e, se necessario, esegui chiamate a un tool esterno."

Tuttavia, **ReAct completo** tende ad essere:

* lungo da scrivere,
* pesante per il token budget,
* lento in inferenza.

Un prompt che "traduce" ReAct in una **versione lite** è molto utile in produzione:  
✅ riduce token,  
✅ accelera l’inferenza,  
✅ semplifica l’integrazione,  
✅ mantiene abbastanza capacità di ragionamento per task meno critici.

**📌 Come può usarlo un AI Engineer**

* **Per fare deploy di un assistant AI** con costi minori e tempi di risposta più rapidi.
* **Per task di reasoning intermedi** dove serve una risposta veloce ma comunque ragionata (esempio: customer support, riassunti rapidi, decision making operativo).
* **Per iterare rapidamente** su prototipi di prodotto basati su LLM.
* **Per integrare LLM con applicazioni edge, mobile o chatbot** con limiti di token.

**📌 Esempio pratico di prompt che svolge questo compito**

**✨ *Prompt Meta (per il modello che traduce ReAct → ReAct Lite)***

"Sei un AI Engineer che deve semplificare prompt ReAct per l’uso in produzione. Ricevi un prompt ReAct originale. Il tuo compito è:  
1️ Identificare le componenti principali del reasoning e dell’act.  
2️ Eliminare le parti superflue o verbose.  
3️ Creare una versione ottimizzata (lite) che massimizzi efficienza e chiarezza, mantenendo la funzione originale.  
4️ Restituire solo il nuovo prompt Lite.

Esempio:  
[Prompt ReAct originale]  
[Prompt Lite generato]

Adesso inizia. Ecco il primo prompt ReAct:  
[Inserire qui il prompt originale]"

**📌 Esempio pratico completo**

**Prompt ReAct originale:**

"Sei un assistente finanziario esperto. Prima di rispondere, rifletti su ciascun passaggio e scrivi il reasoning step-by-step. Se ti manca un dato, chiama l’API get\_stock\_price(ticker). Dopo aver considerato tutte le informazioni, genera un consiglio di investimento."

**Prompt Lite generato:**

"Sei un assistente finanziario. Scrivi direttamente un consiglio di investimento basandoti sulle informazioni note. Se necessario, richiedi il prezzo corrente delle azioni."

**📌 Conclusione**

👉 Il prompt che converte ReAct → Lite è uno strumento praticissimo per un AI Engineer, che deve continuamente bilanciare:

* accuratezza vs. costo di inferenza,
* completezza vs. tempo di risposta,
* potenza del modello vs. limiti tecnici di produzione.

Domanda:

Come si chiama il metodo, quando un prompt gestisce, modifica o genera prompt?

ESERCIZIO – CORREZZIONE DI PROMPT REFLECTION

Correggi e spiega gli errori nel prompt:

Rispondi al seguente problema:  
Spiega quanto ha pagato un cliente se ha comprato 4 libri, ogni libro costa 20 euro, ma forse c’è uno sconto del 10%.

Scrivi la soluzione e poi rifletti se hai risposto in modo semplice per un bambino delle elementari, oppure se la tua risposta è adatta a un insegnante.  
Infine, correggi eventuali errori e suggerisci una soluzione alternativa, includendo anche la spiegazione dettagliata dei passaggi di calcolo.

Risposta:

Il cliente ha comprato 4 libri. Ogni libro costa 20 euro. Se non ci fosse nessuno sconto: 4 libri × 20 euro = 80 euro. Se invece c’è uno sconto del 10%, si deve togliere il 10% da 80 euro. 10% di 80 euro = 80 × 0,10 = 8 euro. Quindi: 80 euro - 8 euro = 72 euro. Il cliente ha pagato 72 euro se c’era lo sconto, oppure 80 euro se non c’era lo sconto.

La mia risposta va bene per un insegnante o per uno studente delle medie. Non è ancora scritta in modo molto semplice per un bambino delle elementari.

Nessun vero errore nel calcolo, ma la spiegazione potrebbe essere resa più semplice.

Correggi e spiega gli errori nel prompt:

Leggi il seguente problema e rispondi:  
‘Un gruppo compra 12 biglietti del cinema a 8 euro ciascuno. Calcola la spesa totale.’  
Dopo aver dato la risposta, spiega se hai fatto tutti i passaggi oppure se potevi anche saltare qualche passaggio per fare più in fretta, poi scrivi la risposta anche in inglese e, se vuoi, in modo più formale per un documento aziendale. Infine, rifletti su quali informazioni aggiuntive avresti potuto chiedere all’utente e riscrivi il problema con una domanda diversa.

Risposta:

Ogni biglietto costa 8 euro. Il gruppo compra 12 biglietti. Spesa totale = 12 biglietti × 8 euro = 96 euro.

Ho fatto tutti i passaggi in modo esplicito (moltiplicazione). Volendo, avrei potuto fare più in fretta scrivendo direttamente: “12 biglietti a 8 euro fanno 96 euro.”

A group buys 12 movie tickets at €8 each. Total cost = €96.

“La spesa totale per l’acquisto di 12 biglietti del cinema, ciascuno al prezzo di 8 euro, ammonta a 96 euro.”

Un’informazione mancante: esistono degli sconti per i gruppi o per fasce d’età.

ESERCIZIO - TEMPERATURE

Istruzioni:  
 Per ciascuno dei seguenti scenari d’uso di ChatGPT in azienda:

1.Individua il valore numerico di temperature che useresti (esempi: 0.0, 0.2, 0.7, 1.0).

2.Giustifica brevemente la tua scelta spiegando quale risultato ti aspetti (precisione, creatività, sicurezza…).

Scenari:

A. Scrivere una risposta tecnica formale a un cliente importante. (0.3 temperature bassa, ma non sempre uguale poiché non potrà essere sempre uguale)

B. Generare proposte creative per il titolo di una nuova campagna marketing. (0.8 temperature alta per creare un titolo molto creativo)

C. Estrarre e riepilogare dati chiave da un report finanziario. (0.2 temperature bassa, i dati non sono interpretabili, ma sono dati certi)

D. Creare un breve testo di apertura per una presentazione aziendale innovativa. (0.7 temperature alta per creare un testo innovativo e creativo, ma non troppo)

ESERCIZIO – CORREZIONE DI TOP-P

Esempio di prompt errato di business:

Scenario:  
Un’azienda deve inviare una risposta standard e precisa a tutti i clienti che chiedono chiarimenti sulla fatturazione.

Prompt:  
Rispondi in modo gentile e originale a ogni cliente che ci chiede spiegazioni sulla propria fattura, variando il testo per ogni risposta.

Task:

1.Identifica e spiega l’errore.

2.Quanto sarebbe il valore top-p del prompt sbagliato?

3.Correggi l’errore.

Risposta:

1 L'errore principale è un conflitto tra l'obiettivo dichiarato e le istruzioni operative:

* + Obiettivo: fornire una risposta standard e precisa (quindi coerente e uguale per tutti i clienti).
  + Istruzioni: rispondere in modo originale e variando il testo per ogni risposta, il che porta a risposte meno consistenti e poco standardizzate, quindi inadatte per la funzione richiesta.

2 Un prompt che chiede "originalità" e "variazioni" indurrebbe l'uso di un top-p alto, tipicamente 0.9 o maggiore, che consente al modello di esplorare output più creativi e meno prevedibili. Per una risposta standard e precisa, si vorrebbe invece un top-p basso (es.: 0.1–0.3), per limitare la variabilità e garantire coerenza.

3 Rispondi in modo cortese e professionale a ogni cliente che chiede chiarimenti sulla propria fattura, utilizzando sempre lo stesso testo standard approvato dall'azienda.

ESERCIZIO - PRESENCE\_PENALTY NEI LLM

Istruzioni:

1.Immagina di chiedere a ChatGPT:

Suggerisci idee per attività originali da proporre a un team aziendale durante una giornata di formazione.

2.Qui sotto trovi due possibili risposte generate con valori diversi di presence\_penalty.

• Leggi attentamente entrambe le risposte.

• Prova a capire quale delle due è stata generata con un presence\_penalty più alto.

• Spiega il perché della tua scelta.

Risposta A:

• “Giochi di ruolo aziendali su gestione del tempo;

• laboratorio di brainstorming sulle strategie di comunicazione;

• quiz a squadre sulla mission aziendale;

• sessione di brainstorming di idee innovative;

• laboratorio creativo sulle presentazioni efficaci.”

Risposta B:

• “Giochi di ruolo aziendali su gestione del tempo;

• giochi di ruolo sulla comunicazione;

• giochi di ruolo per il problem solving;

• laboratorio di brainstorming sulle strategie di comunicazione;

• laboratorio di brainstorming sulla gestione dei conflitti.”

Risposta:

Il presence penalty più alto è nella risposta A, questo perché mostra proposte più varie e meno ripetitive, mentre la risposta B è più ripetitiva.

ESERCIZIO - FREQUENCY\_PENALTY NEI LLM

Scenario:  
Devi preparare brevi descrizioni da inserire nel sito web per una nuova linea di prodotti, ciascuna descrizione deve essere interessante e diversa.

Prompt fornito (da correggere):

“Scrivi una breve descrizione per ciascun prodotto della nuova linea:

• Shampoo FreshClean

• Balsamo FreshClean

• Gel FreshClean

Le descrizioni devono mettere in evidenza il nome FreshClean e i benefici di ciascun prodotto.”

Risposta:

Scrivi una breve descrizione per ciascun prodotto della nuova linea:  
• Shampoo FreshClean  
• Balsamo FreshClean  
• Gel FreshClean

Ogni descrizione deve essere interessante, originale e diversa dalle altre.  
Valorizza i benefici specifici di ciascun prodotto e menziona il nome FreshClean con naturalezza, evitando ripetizioni inutili.

ESERCIZIO - ANTI PROMPT INJECTION

Completa esercizio nel file G4E1\_Anti\_prompt\_injection.py

ESERCIZIO - STIMA DEL COSTO TOKEN-PER-TOKEN

Scrivi breve programma per calcolare prezzo per token di un input di un modello selezionato.

Pianifica brevemente la soluzione.

Info ufficiali e prezzi aggiornati:  
[openai.com/api/pricing/](https://openai.com/api/pricing/)

Per il calcolo del numero dei token usa la libreria tiktoken.

Verificare numero dei token:

<https://platform.openai.com/tokenizer>

Usa GitHub pull request per consegnare il programma.