



# Introduzione al prompting prt 2

*Programmazione*

**Il Prompt Engineering è l'insieme di tecniche e principi utilizzati per progettare, strutturare e ottimizzare le istruzioni fornite a un modello di Intelligenza Artificiale generativa, con l'obiettivo di ottenere output coerenti, affidabili e allineati allo scopo desiderato.**

**Nei modelli linguistici di grandi dimensioni, il prompt non è una semplice domanda, ma rappresenta il contesto operativo completo entro cui il modello genera le risposte: include istruzioni, vincoli, esempi, ruolo assegnato al modello e formato dell'output.**

**Poiché questi sistemi funzionano in modo probabilistico e non deterministico, la qualità del prompt diventa il principale fattore di controllo del comportamento del modello, spesso più rilevante della scelta dello strumento stesso.**



**Le tecniche di prompting nascono come risposta diretta all'evoluzione dei modelli di linguaggio di grandi dimensioni (LLM) e, in particolare, alla loro capacità di generalizzare comportamenti complessi a partire da semplici istruzioni testuali.**

**Inizialmente, con i primi modelli NLP, l'interazione era rigida e fortemente guidata dal codice; con l'avvento dei modelli pre-addestrati e successivamente dei modelli foundation, è emerso che il modo in cui una richiesta viene formulata influenza drasticamente l'output.**

**Da qui nasce il prompting come pratica empirica: sperimentare formulazioni diverse per “indirizzare” il modello.**



**Con il tempo, queste pratiche si sono strutturate in tecniche riconoscibili, trasformandosi da semplice abilità d'uso a vera e propria disciplina metodologica.**

**La NLP (Natural Language Processing) è il ramo dell'intelligenza artificiale che si occupa di permettere alle macchine di comprendere, interpretare, generare e manipolare il linguaggio naturale umano.**

**Combina informatica, linguistica e statistica per trasformare testi e frasi in dati strutturati su cui gli algoritmi possono operare, affrontando problemi come l'analisi del significato, il riconoscimento delle intenzioni, l'estrazione di informazioni e la generazione automatica di testo.**

**La NLP è alla base di tecnologie come chatbot, assistenti vocali, traduttori automatici, analisi del sentiment e motori di ricerca avanzati, ed è diventata centrale con l'evoluzione dei modelli di linguaggio di grandi dimensioni.**

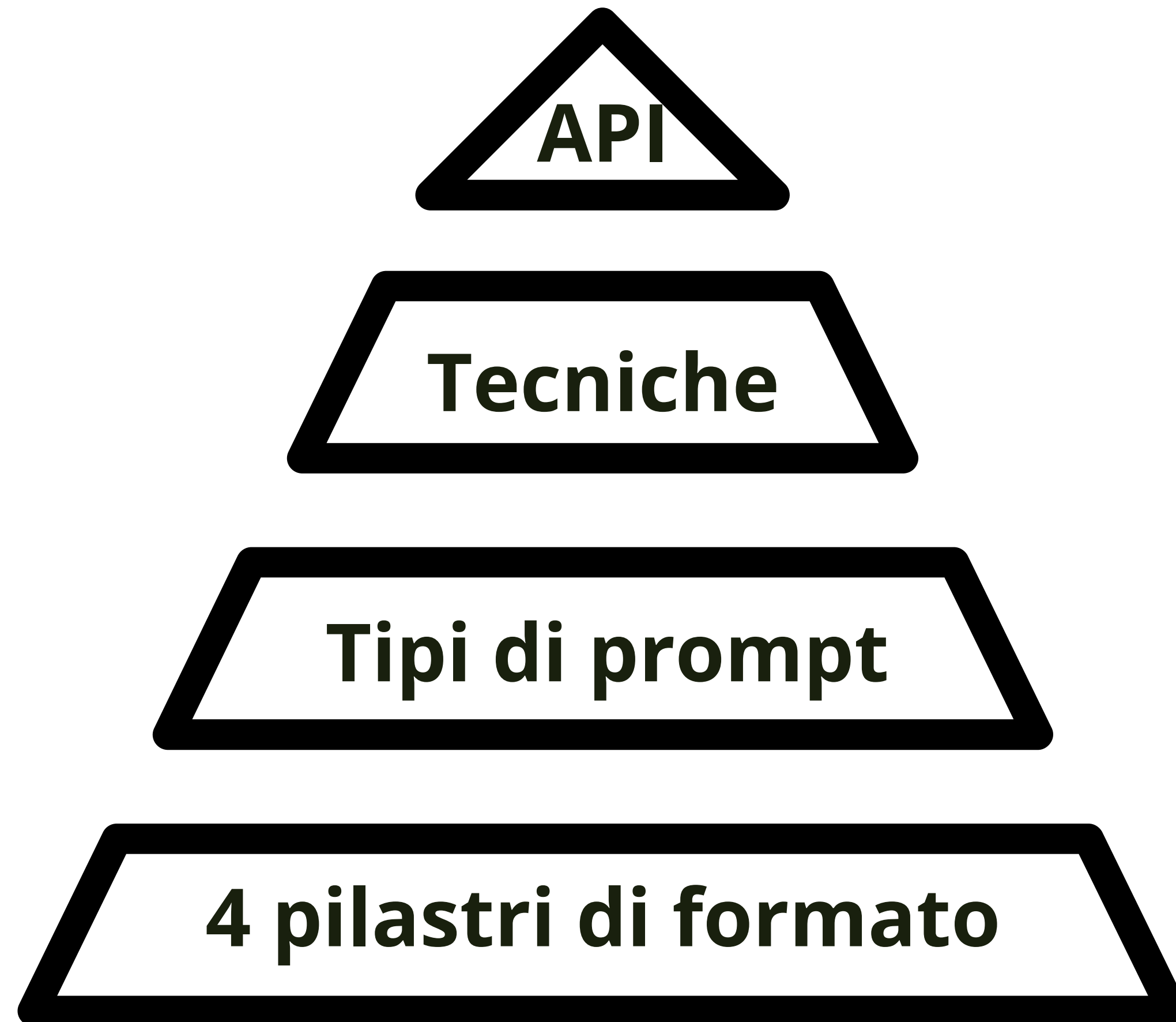


**Gli LLM (Large Language Models) sono modelli di intelligenza artificiale progettati per comprendere e generare linguaggio naturale su larga scala.**

**Si basano su reti neurali profonde, in particolare sull'architettura Transformer, e vengono addestrati su enormi quantità di testi per apprendere pattern linguistici, relazioni semantiche e strutture sintattiche.**

**La loro caratteristica chiave è la generalizzazione: non recuperano risposte predefinite, ma generano output probabilistici coerenti con il contesto fornito, rendendo il modo in cui viene formulato il prompt determinante per il risultato finale.**





## **Tipi di prompt vs Tecniche di prompting**

**Nel contesto del Prompt Engineering è fondamentale distinguere tra tipi di prompt e tecniche di prompting, due concetti spesso utilizzati come sinonimi ma che operano su livelli diversi.**

**I tipi di prompt descrivono che tipo di richiesta viene fatta al modello in base allo scopo dell'output, mentre le tecniche di prompting descrivono come strutturare la richiesta per guidare il comportamento del modello in modo più efficace. Questa distinzione è cruciale per progettare interazioni ripetibili e controllate con l'AI generativa.**



**I tipi di prompt sono quindi legati alla funzione dell'output: ad esempio chiedere una sintesi, una rielaborazione, una generazione creativa o un'estrazione di informazioni.**

## **Tipi di prompt vs Tecniche di prompting**

**Le tecniche di prompting, invece, agiscono sulla costruzione interna del prompt, introducendo elementi come esempi, ragionamento guidato, ruoli o step intermedi.**

**In pratica, due prompt possono appartenere allo stesso tipo (es. “sintesi”), ma utilizzare tecniche completamente diverse, producendo risultati molto differenti in termini di qualità, precisione e affidabilità.**





## In sintesi

### Tipi di prompt

- → Definiscono cosa si chiede al modello
- → Dipendono dall'obiettivo finale (scrivere, riassumere, analizzare, estrarre)

### Tecniche di prompting

- → Definiscono come si guida il modello
- → Dipendono dal livello di controllo, affidabilità e precisione richiesto



- **Zero-Shot Prompting**
- **One-Shot Prompting**
- **Few-Shot Prompting**
- **Instruction Prompting**
- **Role Prompting**
- **Contextual Prompting**
- **Chain-of-Thought Prompting (CoT)**
- **Self-Consistency Prompting**
- **Decomposition Prompting**
- **ReAct Prompting**
- **Tree-of-Thoughts (ToT)**
- **Meta-Prompting**
- **Reflexion Prompting**
- **Iterative Prompting**
- **Constrained Prompting**
- **Output-Format Prompting**
- **Delimiter-Based Prompting**
- **Prompt Chaining**
- **Retrieval-Augmented Prompting (RAG-style)**



## 1. Zero-shot prompting

**Consiste nel porre una richiesta senza fornire esempi. Il modello risponde basandosi esclusivamente sulla conoscenza appresa in fase di addestramento. È rapido ed efficace per compiti semplici, ma meno affidabile in contesti complessi o molto specifici.**

**ESEMPIO:**

**“Spiega cos’è l’AI generativa.”**

**Risultato simulato**

**“L’AI generativa è una tecnologia che consente ai sistemi informatici di creare contenuti come testi e immagini a partire da dati e modelli statistici.”**



- **2. Few-shot prompting**

**Prevede l'inserimento di uno o più esempi all'interno del prompt per mostrare al modello il tipo di output desiderato. Questa tecnica migliora significativamente la qualità e la coerenza delle risposte, soprattutto per task strutturati o ripetitivi.**

**ESEMPIO:**

**Input: Spiega cos'è il cloud computing**

**Output: Il cloud computing è un modello che fornisce risorse IT tramite internet**

**Ora fai lo stesso per l'AI generativa."**

**Risultato simulato**

**"L'AI generativa è un insieme di modelli che permettono di generare contenuti digitali in modo automatico a partire da istruzioni testuali."**



**Le tecniche di prompting rappresentano l'evoluzione metodologica del rapporto tra utente e modello linguistico, permettendo di ottenere risposte più accurate, controllate e coerenti rispetto agli obiettivi richiesti.**

**Tra le tecniche più diffuse e rilevanti troviamo il Chain-of-Thought, che induce il modello a esplicitare il ragionamento passo per passo, migliorando la trasparenza e la correttezza logica dei risultati; consiste nel chiedere esplicitamente al modello di spiegare ogni logica applicata passo dopo passo prima di arrivare alla risposta finale.**

**È particolarmente utile per problemi complessi o concettuali.**



**Esempio: “Spiega cos'è la tokenizzazione illustrando prima il problema, poi il processo e infine l'utilità”.**

**Con il Role Prompting, si assegna al modello un ruolo operativo (es. “sei un analista dati”, “sei un valutatore di rischi”), creando configurazioni più stabili e orientate allo scopo, in questa tecnica il ruolo specificato influenza tono, profondità e prospettiva della risposta.**

**È efficace quando dobbiamo adattare l’output a un pubblico o a un obiettivo preciso.**

**Esempio: “Agisci come un consulente AI per un’azienda e spiega perché la tokenizzazione influisce sui costi”.**



## Limiti delle tecniche

- **Dipendenza dalla qualità del prompt (piccole ambiguità ⇒ output errato)**
- **Scarsa riproducibilità dei risultati**
- **Sensibilità al contesto e all'ordine delle istruzioni**
- **Allucinazioni non completamente eliminabili**
- **Difficoltà nel controllo del ragionamento interno**
- **Costi computazionali più elevati per prompt complessi**
- **Scalabilità limitata in workflow complessi**
- **Manutenzione difficile dei prompt nel tempo**
- **Overfitting al prompt (risposte valide solo in quel contesto)**
- **Scarsa robustezza a input rumorosi o imprevisti**
- **Dipendenza dalle capacità del modello sottostante**

