Gen AI Prompt Engineering

By 

Components of prompt

1. Role (Who the AI should be?)
Defines the persona, expertise, or perspective the AI should assume while responding.
This helps control tone, depth, and style of the output.
Example:
“You are a subject-matter expert in artificial intelligence explaining concepts to beginners.”

2. Task (What needs to be done?)
The overall objective or purpose of the prompt.
Clearly states what the AI is expected to accomplish.
Example:
“Generate a summary of an article.”

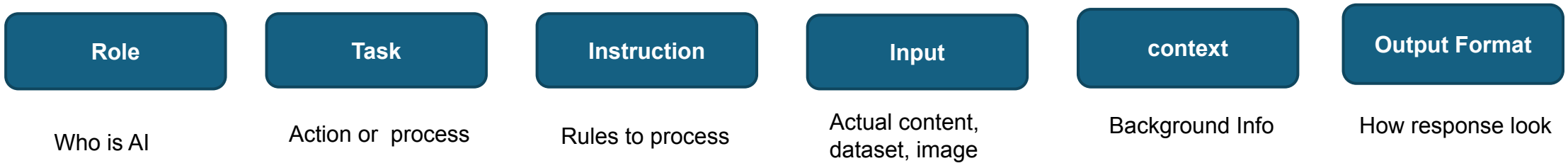
3. Instruction (How should the task be performed?)
Specifies how the task should be executed, including rules, limits, or style.
Example:
“Summarize the article in no more than 100 words, keeping the key points intact.”

Components of prompt

4. Input (What data is provided to perform the task?)
The actual content or information on which the AI operates.
Example:
“The article text:
‘AI is transforming industries by automating tasks…’”

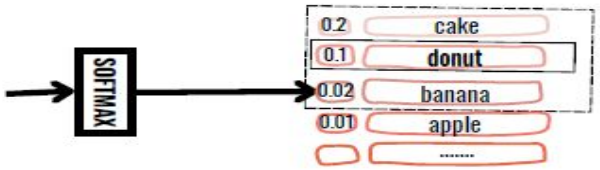
5. Context (Additional details that help improve accuracy)
Background information, constraints, target audience, or usage scenario that refines the response.
Example:
“This summary is intended for a general audience with no technical background.”

6. Output Format (How should the response look?)
Defines the structure, layout, or presentation style of the final output.
Example:
“Present the summary as a single paragraph using simple language.”
“Get the change for 1000 in 100 rupees notes”



Prompt parameters

Top N.



Top N : The word/token is selected using random-weighted strategy but only from amongst the Top 'N' words/tokens

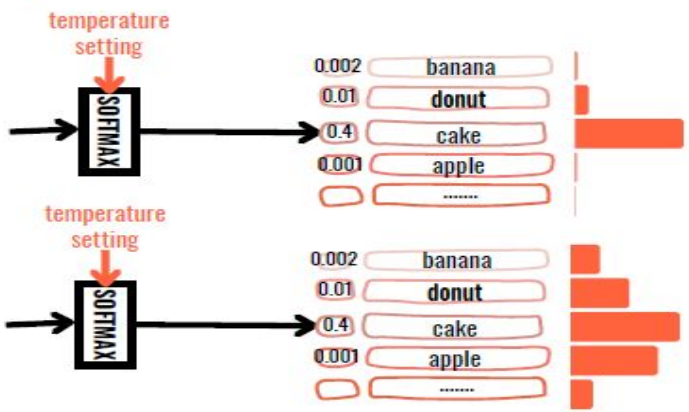
Here for N=3, one of cake, donut or banana will be selected randomly but apple will never be selected

Top P.

Top P : The word/token is selected using random-weighted strategy but only from amongst the top words totalling to probability <=P

Here for P=0.33, one of cake or donut will be selected randomly but apple or banana will never be selected

Temperature



Cooler Temperature (lesser value) :
The distribution is strongly peaked

Warmer Temperature (higher value) :
Flatter probability distribution

Terminology

Term	Meaning
Prompt	Instructions + context; what you send to the model
Generated output	Tokens the model produces
Context window / Max sequence length	Total tokens model can handle at once = prompt + output
max_tokens / max_output_tokens	Limits how long the model output can be; must fit inside context window
Sampling Parameters	temperature, top-p, max_output_tokens

Prompt Engineering

Prompt Engineering is the art of designing inputs for LLMs (like ChatGPT) to obtain desired responses.

Prompt Engineering Techniques :

1. Zero-Shot Prompting
2. Few-Shot Prompting
3. Chain-of-Thought (CoT) Prompting
4. ReAct (Reasoning + Acting)
5. Self-Consistency
6. Instruction-Based Prompting
7. Role-Based Prompting

Zero Shot Prompting

Concept
Zero-shot prompting relies on the model's pre-trained knowledge to answer queries without providing examples.

Example
Prompt:
"What is the capital of France?"

ChatGPT Response:
"The capital of France is Paris."

Use Case:
Simple fact-based queries
General knowledge retrieval

Limitations:
Struggles with tasks requiring context or reasoning

Few Shot Prompting

Concept
Providing a few examples to guide the model on the desired format and reasoning.

Example
Prompt:
"Classify the following texts as Positive or Negative sentiment:
'I love this product!' → Positive
'This is the worst experience ever!' → Negative
'I am very happy with my purchase.' → ?"*

ChatGPT Response:
"Positive.“

Use Case:
Sentiment analysis
Text classification
Pattern learning

Limitations:
Requires carefully selected examples

Chain of Thought (CoT)

Chain-of-Thought (CoT) Prompting

Concept
Encouraging the model to break down reasoning into intermediate steps before answering.

Example
Prompt:
"A farmer has 3 baskets. Each basket contains 5 apples. If he gives away 4 apples, how many apples remain? Let's think step by step.“

ChatGPT Response:
"Step 1: The farmer starts with 3 baskets × 5 apples = 15 apples.
Step 2: He gives away 4 apples.
Step 3: 15 - 4 = 11 apples remain.
Final Answer: 11 apples.“

Use Case:
Complex math problems
Logical reasoning tasks

Limitations:
Requires structured reasoning tasks

ReAct

ReAct (Reasoning + Acting) Reason → Action → Observation → Response

Concept
Combining reasoning with action by making the model think, retrieve, and act iteratively.

Example
Prompt:
"Who won the 2023 FIFA World Cup? Think step by step and search if needed.“

ChatGPT Response (simulating ReAct):
"Step 1: The last FIFA World Cup was held in 2022, and Argentina won.
Step 2: Let me check the 2023 winner (simulated search).
Step 3: The 2023 FIFA Women's World Cup was won by Spain.
Final Answer: Spain won the 2023 FIFA Women's World Cup.“

Use Case:
Dynamic decision-making
Research-based reasoning

Limitations:
Needs external search for live data

Self-Consistency

Run the same prompt multiple times

Example
Prompt:
"If a train travels 100 miles in 2 hours, what is its speed?
or
"If a train travels 100 miles in 2 hours, what is its speed? Solve using two different methods.“ <not exactly a self consistency)

ChatGPT Response:
Method 1:
"Speed = Distance ÷ Time = 100 miles ÷ 2 hours = 50 mph."
Method 2:
"If the train covers 100 miles in 2 hours, in 1 hour it covers 50 miles. So, the speed is 50 mph."
"Final Answer: 50 mph.“

Use Case:
Reducing hallucinations in reasoning tasks
Improving mathematical accuracy

Limitations:
Can be computationally expensive

Other useful Techniques

Instruction-Based Prompting
Instead of a vague query, clearly instruct the model.

Example
❌ "Summarize this article."
✅ "Summarize this article in three bullet points, each under 20 words.“

Role-Based Prompting
Assigning a persona to the model to align responses.

Example
"You are a cybersecurity expert. Explain the importance of multi-factor authentication in banking security.“

By using these techniques, you can significantly improve LLM’s responses for different use cases like NLP tasks, logical reasoning, and automation

Thank you