

# Prompt Engineering

Name : Preetham

## Report

### Introduction to Large Language Models (LLMs)

LLMs are very complex systems, each uniquely designed to execute diverse functions involving NLP and machine learning. The LLMs have been trained on the largest datasets, which greatly enable them to provide human-like texts and various forms of tasks.

Constructing LLMs, such as ChatGPT, comprises integrating several millions of parameters, sometimes surpassing 175 million, that make them understand and respond to complex inputs provided by the users.

### Examples of LLM

1. ChatGPT → A chatbot by OpenAI, designed to generate text-based answers to user inputs.
2. Bard → LLM provided by Google
3. MidJourney → An LLM model capable of generating images based on text descriptions.
4. **Others:** There are many more models, each with unique capabilities.

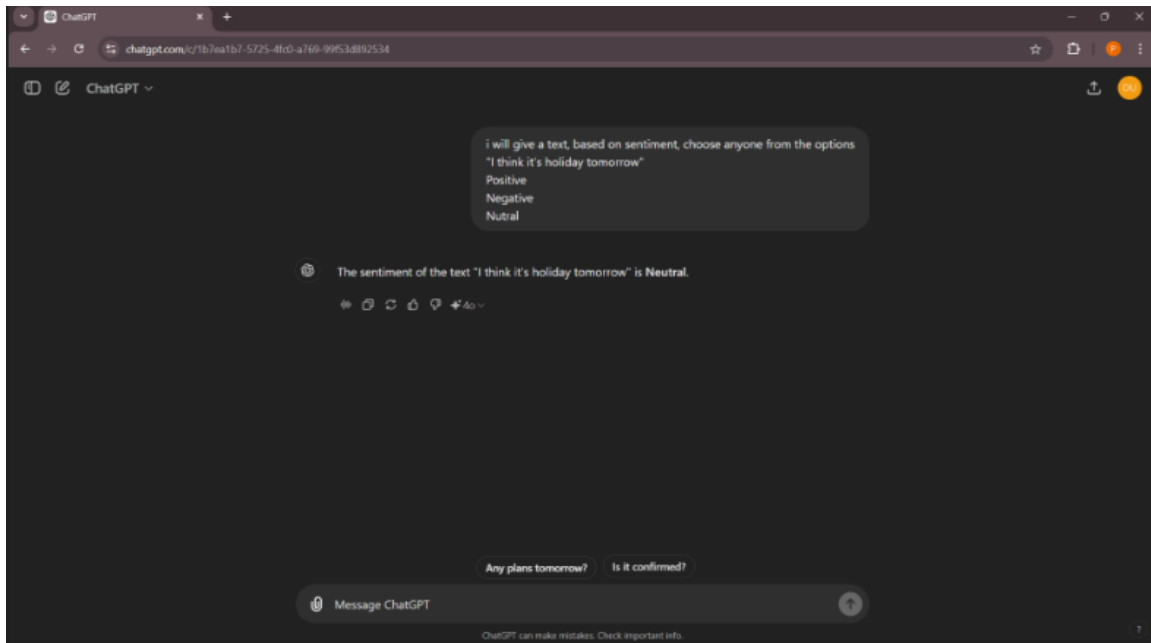
### What is Prompting

Prompting is the ability to instruct an LLM, using user input, to perform a specific task or give targeted output. Well-constructed and effective prompts are crucial in the successful derivation of accurate and relevant responses from the model.

### Types of Prompting techniques

## 1. Zero Shot Prompting

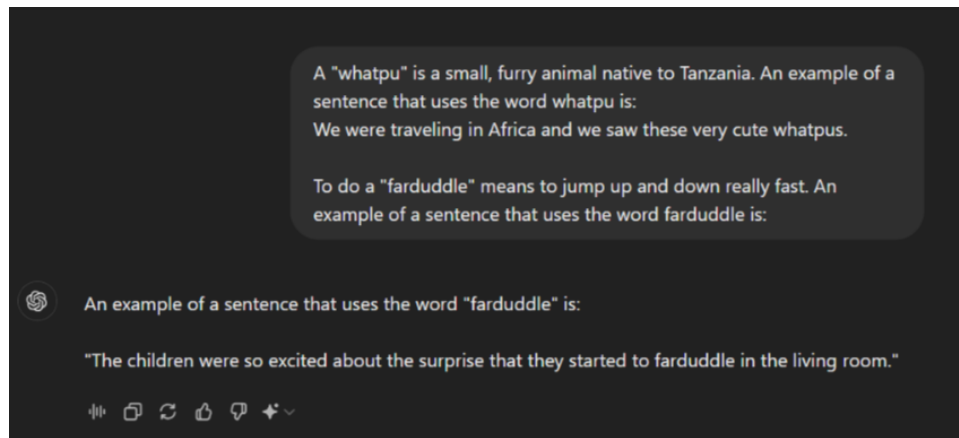
1. **Definition:** A technique where no examples or demonstrations are given to the model before asking a question.
2. **Outcome:** The model generates a response based solely on the information it has been trained on, without any specific guidance from the user.



3. We didn't give any prior example or pattern to the model but still it managed to give us the output.

## 2. Few Shot Prompting

1. **Definition:** Involves providing the model with a few examples or demonstrations before asking the question.
2. **Outcome:** The model uses these examples to understand the desired output pattern and respond accordingly. Few-shot prompting is more effective than zero-shot in delivering context-specific results.
3. This will help model understand about the problem more clearly than zero prompt.



4. In the above context we have demonstrated an example to the model to make it understand clearly. Based on the example model gave output accordingly.

### 3. Chain-Of-Thoughts (CoT)

1. **Definition:** This technique involves breaking down complex tasks into logical steps, making it easier for the LLM to follow and generate coherent outputs.
2. **Usage:** Commonly used for tasks that require logic, calculation, or decision-making.
3. **Example:** Adding phrases like "Explain your reasoning step by step" encourages the model to provide a detailed breakdown of how it arrived at a particular answer.

(a) Few-shot

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?  
A: The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?  
A:

(Output) The answer is 8. ✗

(b) Few-shot-CoT

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?  
A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?  
A:

(Output) The juggler can juggle 16 balls. Half of the balls are golf balls. So there are  $16 / 2 = 8$  golf balls. Half of the golf balls are blue. So there are  $8 / 2 = 4$  blue golf balls. The answer is 4. ✓

(c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?  
A: The answer (arabic numerals) is

(Output) 8 ✗

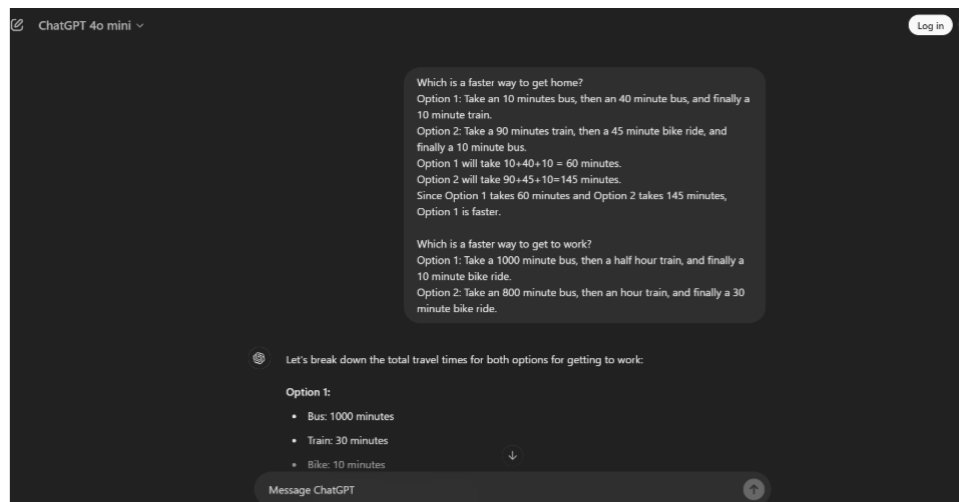
(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?  
A: Let's think step by step.

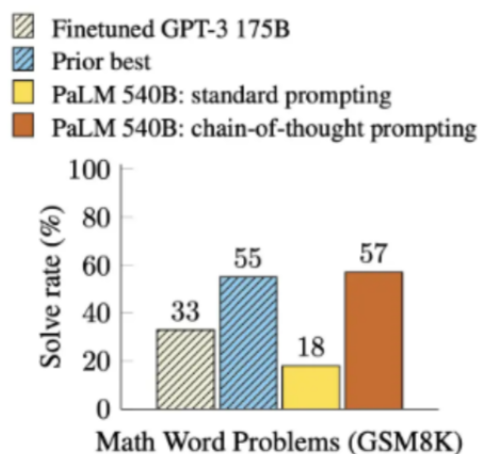
(Output) There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls. ✓

4. You can find the key difference between Few-shot, Zero-Shot with Few-shot-CoT and Zero-shot-CoT in the above image.

**Note :- To construct a chain-of-thought prompt, a user typically appends an instruction such as "Describe your reasoning in steps" or "Explain your answer step by step" to the end of their query to a large language model (LLM). In essence, this prompting technique asks the LLM to not only generate an end result, but also detail the series of intermediate steps that led to that answer.**



5. In this case, examples are given to the model to generate the output in the given pattern.
6. Here more examples (or) patterns is directly proportional to good prompt and accurate answer.



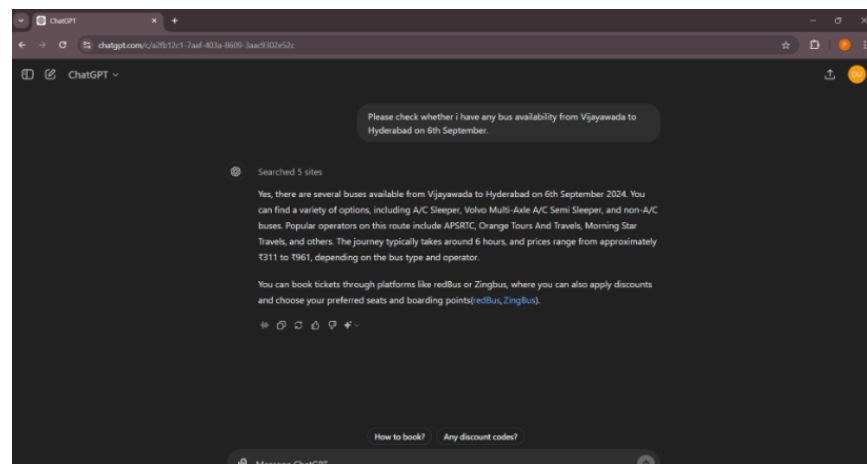
## Drawbacks of CoT

While effective, CoT prompting has some limitations:

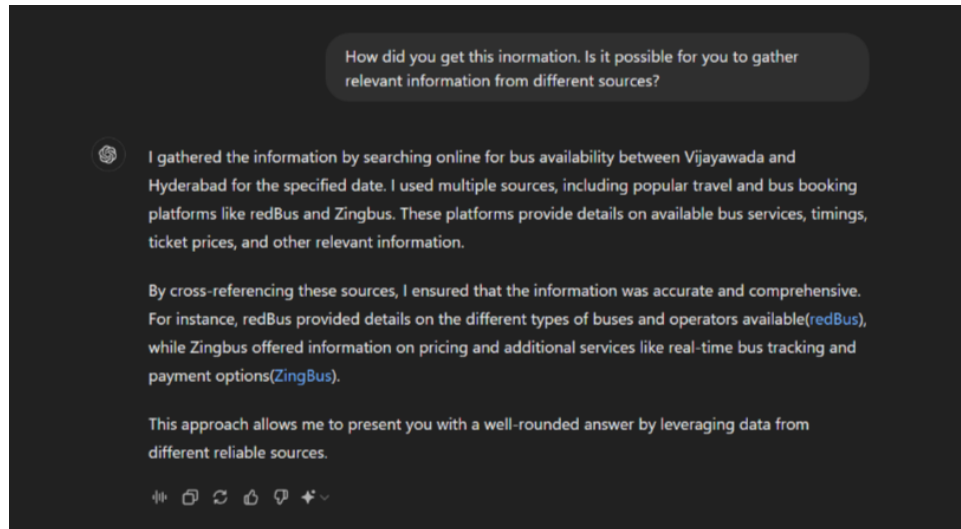
- **Manual Effort:** Requires more manual input and guidance from the user.
- **Model Size:** Larger models are needed to handle this type of prompting effectively.
- **Prompt Bias:** The model may become biased towards specific answers based on the provided examples.

## 4. ReAct (Reasoning and Action)

1. **Definition:** ReAct is a technique where the LLM uses external sources to support its responses based on the user's input.
2. **Example:** A prompt can guide the model to use external information and then reason through it to generate the desired output.



3. It gives the reasoning for the external useage as well



# Google's AI Studio

## Understanding the platform.

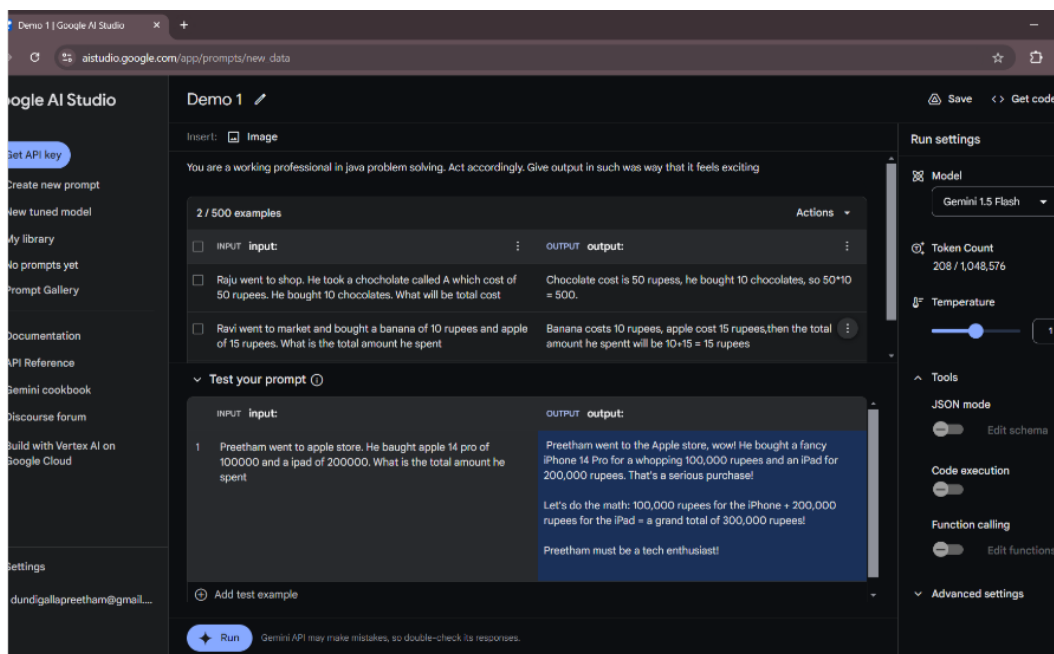
Google's AI Studio provides a flexible environment for working with LLMs. Key sections include:

- **API Key Access**
- **Prompt Gallery**
- **Create New Prompt**
- **Tuning Models**

## Create new prompt

1. Here we have 3 roles.
  - a. **Role 1 (System Tone):** Defines the behavior of the model (e.g., software developer, sportsperson).
  - b. **Role 2 (User Input):** The user provides the prompt.
  - c. **Role 3 (Model Response):** The model generates a response based on the input and defined roles.

2. Here we can give videos, images as an prompt as well. Based on the input and the tone of the system we will get output accordingly.
3. Inside this we have another prompting technique, that is Structured prompt
4. **Structured Prompt** : Structured prompting involves giving both the input and the expected output as examples. This method helps the model understand how to respond to similar inputs in the future. The more examples provided, the more accurate the model's responses become.



5. As you can see, we have given examples input and example output to that. By thus model will understand how to respond to the given problem.