

LLM: Prompts and Applications

Prompt Patterns & Techniques - 4

Diego Pavan Soler
03/2025

Insper

A decorative graphic on the right side of the slide. It consists of three overlapping squares: a black square in the background, a red square in the foreground, and a square with diagonal grey and white stripes behind the black one. The red square contains the text 'Insper.edu.br' in white.

Insper.edu.br

Review of Prompt 1

- **Basic Patterns & Techniques**

- **The Persona Pattern:** Instruct the model to assume specific roles.
- **The Audience Persona Pattern:** Adjusting prompts for different audiences.
- **The Emotion Prompting Pattern:** Add emotion on your prompts to make AI response more thoughtful and helpful.
- **The Tail Generation Pattern:** Ask the AI to add supporting text to the end of prompts.

- **Prompt Refinement Patterns**

- **Ask AI to Refine your Prompt**
- **The Question Refinement Pattern:** Ask the AI to improve and use your prompts.

- **Context Expansion and Understanding**

- **The Flipped Interaction Pattern:** Ask the model to generate YOU question to deepen discussions.
- **The Ask for Input Pattern:** Prompting users for additional input to refine AI responses.
- **Dealing With Long Form Content:** Split and summarize the chunk.

Review of Prompt 2

- **Prompt Improvement and Interactions**

- **The Template Pattern:** Ensure that its responses follow a specific template AI.
- **Few-Shot Prompting:** Provide the AI with a small set of examples to illustrate the desired format or behavior.
- **Game Play Pattern:** Ask the LLM to play games with you, useful for fun or work.

- **Verification Techniques**

- **Re-reading (RE2):** Prompt the model to re-read the question or initial input
- **Self-Refine:** Use the AI to provide feedback and improve it own work.
- **Self-Consistency:** Great for classification tasks, as the AI the same prompt multiple times and average the results.
- **LLM Self-Evaluation:** Similar to Self-Refine, but in these case we use a different model to give feedback or limit the work of another.

- **Self-Improvement Techniques**

- **Generated Knowledge:** Prompts LLMs to generate relevant information before producing a final response, enhancing accuracy.
- **Code Prompting:** Transforms natural language tasks into structured code representations with conditional logic.
- **Plan-and-Solve Prompting:** Encourages the model to first outline a step-by-step plan before solving a problem.
- **Cognitive Verifier Pattern:** Split the question into an easier sub-question and then respond then

Review of Prompt 3

- **Prompt Improvement and Interactions**

- **The Recipe Pattern:** Structures the AI responses as clear and predictable instructions.
- **The Alternative Approaches:** Prompt the AI to generate different ways to address a query.
- **The Outline Technique:** Prompt the AI to generate new or expand outlines.
- **The Fact Check List:** Prompt the ai to generate facts that support or validade response.

- **Automation and Workflow Efficiency**

- **Dataset Creation:** Prompt the AI to generate a dataset, or column options.
- **The Menu Actions:** This pattern involves designing prompts that function as shortcuts to complex actions.
- **The Meta Language:** Craft or use unique terminologies or symbols within prompts.
- **Prompt Debiasing:** Prompt debiasing is the process of crafting prompts to remove biases from AI outputs.
- **Semantic Filter:** Prompt the AI to filter only relevant information from text.
- **Self-Generated In-Context Learning:** Improve classification task by asking the AI to generate annotated examples.

Prompt Tools

OpenAI's “Create Your GPT”

Create Your GPT is a feature provided by OpenAI that allows users to design and deploy custom versions of GPT tailored to specific tasks, industries, or creative projects. It empowers developers and non-technical users to fine-tune GPT models for unique applications.

- Use it when you need a GPT model tailored to a specific domain or problem.
- “Explore GPTs” has many GPTs created by the community.
- You can use it to with to automate a Root Prompt we learned in this Course.
- You can upload documents and give API access for further context improvements.

Example:

- On the create tab of create new GPT prompt the assistant
 - Hello, help me create a GPT that will **slightly** improve my e-mails.
 - You can give current time access to by adding API calls. [\[Link\]](#)
 - The assistant could also use DALL-E-3 to generate a custom emoji signature to emails.
- Apply ROOT prompts to GPTs to facilitate workflow

Notebook LLM

What is NotebookLM?

- NotebookLM is an experimental AI-powered research tool developed by Google.
- It allows users to upload up to 50 documents, such as PDFs, Google Docs, or text files, and train a specialized AI based on those documents.
- Think of NotebookLM as an AI assistant that reads the documents with you, making research and learning more dynamic and hands-on.

Notebook LLM

How to Use NotebookLM?

- **Upload Your Documents:** NotebookLM supports various file formats and sources, including PDFs, Google Docs, website links, and copied and pasted text.
- **Ask Questions:** Ask natural language questions about your documents and get concise answers with specific citations.
- **Generate Ideas:** Ask NotebookLM to generate ideas, such as summaries, talking points, or even haikus.
- **Organize Your Notes:** Use NotebookLM to structure your notes in formats like FAQs, study guides, or timelines.
- **Experiment with Different Prompts:** Explore various ways to interact with NotebookLM to discover creative insights.
- **Create Personalized Podcasts:** Generate podcasts based on your documents, featuring two AI voices discussing the key points.

Remember that NotebookLM is still an experimental product, so you may encounter issues while using it. However, NotebookLM has the potential to transform how we interact with content, making research easier and more intuitive.

Notebook LLM

The previous slides, were created by Notebook LLM, here's how it was done:

- Go to [Google NotebookLM](#)
- Print as PDF the following pages:
 - [Datacamp - NotebookLLM](#)
 - [NoteBook LM is amazing : r/notebooklm](#)
 - [Unlock Research Power with Google's Notebook LLM: The AI Assistant You've Been Waiting For! | by Teendifferent | Medium](#)
 - [Googles Notebook LLM](#)

Notebook LLM - Task 1

- Create a podcast about a subjects in one of the classes your currently taking - expect for these one.

ChatGPT's "Canvas"

Canvas (Lousa) is a feature that allows collaborative work of long-form documents or code files. It enables real-time updates, such as, create, edit and review in a dedicated space, while keeping the main chat open for discussion.

- Enhanced Collaboration: Share, edit, and refine ideas collaboratively.
- Improved Clarity: Visualize complex concepts clearly and effectively.
- Efficient Workflow: Integrate writing, editing, and reviewing in one seamless environment.

Tought

Mentimeter



What is a thought?

How do you think?

**Is it currently possible for a
machine to think? Why?**

AI vs AGI

Artificial Intelligence (AI)

- Narrow or specialized intelligence
- Designed for specific tasks (e.g., image recognition, chatbots, self-driving cars)
- Lacks adaptability beyond programmed functions
- Example: Siri, ChatGPT, Google Translate

Artificial General Intelligence (AGI)

- Human-like intelligence
- Capable of reasoning, problem-solving, and learning across various domains
- Adapts to new tasks without reprogramming
- Hypothetical – not yet achieved

AI excels at specific tasks, while AGI would possess human-like understanding and adaptability across many fields.

Chain of Thought

Chain of Thought (CoT) involves breaking down complex tasks into a series of intermediate steps, enabling the model to reason through problems step by step. This approach is particularly effective for tasks that require logical reasoning, mathematical calculations, or contextual deductions.

Zero-Shot

- "{question}. Let's think step by step."
- "{problem}. Let's solve this step by step."
- "I will provide you with a problem or question. Your goal is to solve it step by step, showing your work and reasoning clearly at each stage."

Chain of Thought

Standard Prompting

Model Input

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: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

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: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅

Chain of Thought - Zero-Shot

<p>Math Word Problems (free response)</p> <p>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?</p> <p>A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.</p>	<p>Math Word Problems (multiple choice)</p> <p>Q: How many keystrokes are needed to type the numbers from 1 to 500?</p> <p>Answer Choices: (a) 1156 (b) 1392 (c) 1480 (d) 1562 (e) 1788</p> <p>A: There are 9 one-digit numbers from 1 to 9. There are 90 two-digit numbers from 10 to 99. There are 401 three-digit numbers from 100 to 500. $9 + 90(2) + 401(3) = 1392$. The answer is (b).</p>	<p>CSQA (commonsense)</p> <p>Q: Sammy wanted to go to where the people were. Where might he go?</p> <p>Options: (a) race track (b) populated areas (c) desert (d) apartment (e) roadblock</p> <p>A: The answer must be a place with a lot of people. Race tracks, desert, apartments, and roadblocks don't have a lot of people, but populated areas do. So the answer is (b).</p>
<p>StrategyQA</p> <p>Q: Yes or no: Would a pear sink in water?</p> <p>A: The density of a pear is about 0.6 g/cm^3, which is less than water. Thus, a pear would float. So the answer is no.</p>	<p>Date Understanding</p> <p>Q: The concert was scheduled to be on 06/01/1943, but was delayed by one day to today. What is the date 10 days ago in MM/DD/YYYY?</p> <p>A: One day after 06/01/1943 is 06/02/1943, so today is 06/02/1943. 10 days before today is 05/23/1943. So the answer is 05/23/1943.</p>	<p>Sports Understanding</p> <p>Q: Is the following sentence plausible? "Joao Moutinho caught the screen pass in the NFC championship."</p> <p>A: Joao Moutinho is a soccer player. The NFC championship is part of American football, not soccer. So the answer is no.</p>
<p>SayCan (Instructing a robot)</p> <p>Human: How would you bring me something that isn't a fruit?</p> <p>Explanation: the user wants something to eat that isn't a fruit. An energy bar is not a fruit, so I will bring the user an energy bar.</p> <p>Plan: 1. find(energy bar) 2. pick(energy bar) 3. find(user) 4. put(energy bar) 5. done().</p>	<p>Last Letter Concatenation</p> <p>Q: Take the last letters of the words in "Lady Gaga" and concatenate them.</p> <p>A: The last letter of "Lady" is "y". The last letter of "Gaga" is "a". Concatenating them is "ya". So the answer is ya.</p>	<p>Coin Flip (state tracking)</p> <p>Q: A coin is heads up. Maybelle flips the coin. Shalonda does not flip the coin. Is the coin still heads up?</p> <p>A: The coin was flipped by Maybelle. So the coin was flipped 1 time, which is an odd number. The coin started heads up, so after an odd number of flips, it will be tails up. So the answer is no.</p>

Figure 3: Examples of \langle input, chain of thought, output \rangle triples for arithmetic, commonsense, and symbolic reasoning benchmarks. Chains of thought are highlighted. Full prompts in Appendix G.

Chain of Thought - Few-Shot

Examples:

- The sum of the digits is greater than 15: 234, 118, 45, 90.
- A: Adding all the digits ($2+3+4 + 1+1+8 + 4+5 + 9+0$) gives 37. The answer is True.
- The sum of the digits is greater than 15: 101, 211, 13, 5.
- A: Adding all the digits ($1+0+1 + 2+1+1 + 1+3 + 5$) gives 15. The answer is False.
- The sum of the digits is greater than 15: 12, 56, 78, 45.
- A: Adding all the digits ($1+2 + 5+6 + 7+8 + 4+5$) gives 38. The answer is True.
- The sum of the digits is greater than 15: 3, 4, 5, 6.
- A: Adding all the digits ($3 + 4 + 5 + 6$) gives 18. The answer is True.
- The sum of the digits is greater than 15: 9, 8, 6, 2.

Chain of Thought - Task 2

Use the Few-Shot CoT technique to teach the LLM to find the subject of a phrase.

ReAct Prompting

ReAct (Reasoning + Acting) combines reasoning with actionable steps to solve problems. It guides the model to think about what is needed, act on the requirements, and then solve the problem.

- Helps on decision-making processes
- Reduces hallucinations
- Problem-solving scenarios with interdependent steps
- Task planning

ReAct Prompting

- Task: Calculate when I need to leave for the Eiffel Tower tour to ensure I arrive on time for my 11:00 AM slot.
- Think: I need to find out what time the Eiffel Tower opens to ensure I plan my arrival properly. I can use a web search of the official Eiffel Tower website to get the information.
- Action: SEARCH <https://www.toureiffel.paris>
- Result: The Eiffel Tower opens at 9:30 AM.
- Think: I need to find out how long it takes to get from my hotel to the Eiffel Tower. I can use a mapping service like Google Maps to estimate the travel time.
- Action: SEARCH Google Maps - Hotel to Eiffel Tower
- Result: It takes 20 minutes by car or 30 minutes by metro to get to the Eiffel Tower from my hotel.
- Think: I need to allow time for security checks and potential queues at the Eiffel Tower. I will search for an estimate of how long security and entry take.
- Action: SEARCH Eiffel Tower security queue time
- Result: Security checks and entry typically take 30 minutes.
- Think: If travel takes 20 minutes by car and security checks take 30 minutes, I should plan to leave my hotel by 10:10 AM to ensure I arrive 20 minutes early at 10:40 AM for my 11:00 AM slot.

ReAct Prompting - Task 3

Using the ReAct Prompting approach, simulate an example of an IT professional assisting a professor with an issue related to their Lenovo Yoga notebook.

Scenario:

The professor reports that when he presses the ON button, the notebook does not turn on. However, the light near the charging port blinks three times in orange.

- Ensure the troubleshooting follows a logical thought process, with the IT professional gathering information and making decisions.
- Avoid solving the issue too quickly—show an iterative diagnostic process.

Tasks

Project 2 - Survey Paper on CoT Improvements

Part 1: Writing a Survey Paper

In this assignment, you will write a **survey paper** based on the following research papers, analyzing:

- The functioning of the presented techniques.
- The evolution of these techniques over the Chain-of Thought
- Application examples different from those provided in the papers.
- A comparison between the techniques using tables and arguments.

Mandatory Articles:

1. **Chain-of-Thought** - [Link](#)
2. **Plan-and-Solve Prompting** - [Link](#)
3. **Self-Consistency** - [Link](#)
4. **Self-Verification** - [Link](#)
5. **Tree of Thoughts (ToT)** - [Link](#)
6. **Perspective-Taking** - [Link](#)
7. **PAL: Program-aided Language Models** - [Link](#)

Additionally, select **one more paper** that presents a technique **different** from those listed above or discussed in class, but expands on AI thought mechanics.

Project 2 - Survey Paper on CoT Improvements

Additionally, select **one more paper** that presents a technique **different** from those listed above or discussed in class, but expands on AI thought mechanics.

Part 2: Preparation of Presentation Slides

- Prepare a presentation explaining how the technique from your extra selected paper works.
- Include examples of how this technique is applied.
- Develop a practical exercise for the class.

Note: Due to time constraints, **you do not need to present the slides**. They will be compiled and shared with the entire class.

References

- <https://www.vanderbilt.edu/generative-ai/prompt-patterns/#h2-audience-persona-pattern>
- <https://www.coursera.org/learn/prompt-engineering/home/module/1>
- <https://www.promptingguide.ai/>
- <https://learnprompting.org/docs/introduction>

Inspire

