

Spring 2024

Introduction to Artificial Intelligence

Homework 5: Let's Play LLM

Due Date: 6/10(Mon) 23:59

Introduction

As one of the most powerful advancements in artificial intelligence, LLM offers excellent capabilities in understanding and generating human-like text. Not only in research but also for aiding communication, enhancing productivity, efficient information retrieval, and so on, LLM has become a regular part of our daily life.

In this assignment, we focus on a pivotal aspect of LLM's functionality: the prompt. Crafting a well-constructed prompt is critical to inspire the full potential of LLM. Your task is to engage with LLM, experiment with different prompts, and uncover intriguing content related to this remarkable model. So, let's dive in and discover the fascinating world of LLM together!

Part 1: Designing a Task for LLM and Explore the Capability of LLM (70%)

In this part, your goal is to create a task that an LLM can solve. Try to solve this task using an LLM and list any challenge you encounter during the process, as well as the methods you use to improve the results.

Our evaluation criteria are listed below. Your report should include them, and the more comprehensive, the better.

- **Task Description (10%)**
 - Describe the task you want to solve
 - Describe the format of the input prompt
 - Describe the ideal output
 - ... (Any other things you want to discuss)
- **Motivation (10%)**
 - Describe the reason why you choose this task
 - Describe the capability of LLM you want to explore
 - ... (Any other things you want to discuss)

- **Describe your Initial Attempt (15%)**
 - Input
 - Output
 - Analysis
- **Describe Improvements (35%)** (Describe the methods you've tried to improve the result and compare them)
 - Input
 - Output
 - Analysis

We've provided two examples using ChatGPT3.5 for your reference, please check out **Appendix 1**.

Part 2: Questions Answering (30%)

In this part, you need to answer the following two questions.

Question 1

From this assignment, it is evident that the quality of the prompt greatly influences the quality and accuracy of LLM's responses. The following is an example illustrating the impact of good and bad prompts.

Without prompt design, Gemini cannot solve the problem correctly.



How many pencils might be left when a box of pencils is divided among 5 people?

A. 3,

B. 4,

C. 5,

D. 6

Select all the possible options. The answer may be multiple.



That's right, there are several possibilities (A, B, C, D) for the number of pencils left over when a box is divided among 5 people.

Here's why:

- We don't know how many pencils were in the box to begin with.
- If the box originally had 23 pencils, then dividing among 5 people would leave 3 pencils leftover.
- If it had 24 pencils, there would be no pencils leftover (each person would get 4 and there would be none remaining).
- Any number of pencils between 23 and infinity could have been in the box, resulting in a remainder between 0 and 4 pencils leftover.

X

So, all of the answer choices (3, 4, 5, and 6) are possible depending on the original number of pencils in the box.

With prompt design, Gemini outputs correct answers with explanations.



How many pencils might be left when a box of pencils is divided among 5 people?

- A. 3,
- B. 4,
- C. 5,
- D. 6

Please carefully observe the question and calculate the answer step by step according to the concept of the Division Theorem. Also, ensure that reasonable remainders are chosen according to the question.

Select all the possible options. The answer may be multiple.



You're absolutely right. By applying the Division Theorem, we can determine the possible remainders more precisely.

The Division Theorem states that when a dividend (d) is divided by a divisor (b), there exists a quotient (q) and a remainder (r) such that:

$$d = b * q + r$$

where $0 \leq r < b$

In this case:

- Divisor (b) = 5 (number of people)
- We want to find possible remainders (r) for the number of pencils leftover.

According to the theorem, the remainder must be less than the divisor (5). So, the possible remainders are:

- $r = 0$ (no pencils leftover, everyone gets an equal share)
- $r = 1$ (one pencil leftover)
- $r = 2$ (two pencils leftover)
- $r = 3$ (three pencils leftover)
- $r = 4$ (four pencils leftover)

Therefore, based on the Division Theorem and the context of a reasonable number of pencils in a box, the possible remainders are:

- A. 3
- B. 4
- C. (not possible, as 5 pencils leftover would be more than the divisor 5)
- D. (not possible, as 6 pencils leftover would be more than the divisor 5)

So the possible answers are **A and B only.**



Please answer the questions:

1. Besides providing a more detailed description of the question in the prompt, what other methods can be employed to improve the accuracy of the responses?
2. Please provide an example demonstrating the impact of implementing the aforementioned methods on a prompt.

Question 2

While using LLM, you probably find that it cannot accurately answer your queries. This indicates that current large language models (LLMs) still face numerous challenges. Please search online (papers, blogs, etc.) for the existing challenges of LLMs and briefly explain the motivation, difficulties, and current methods available to optimize or solve these issues.

Below is an example for your reference; feel free to use your own way to answer the question!

- Question : Update outdated knowledge that existed in LLMs
- Motivation : Real-world knowledge changes every day, but re-training LLM is expensive.
- Challenges :
 - How to update outdated knowledge efficiently?
 - Will the rephrased knowledge also be updated?
 - Will unrelated knowledge be unchanged?
- Solution :
 - Train a hyper-network with constrained optimization to modify a fact without affecting the rest of the knowledge.
- Reference
 - [Editing Factual Knowledge in Language Models - ACL Anthology](#)

QA Page

If you have any questions about this homework, please ask them on the following Notion page. We will answer them as soon as possible. Additionally, we encourage you to answer other students' questions if you can.


<https://lopsided-soursop-bec.notion.site/HW5-QA-Sheet-e07d9829ea3849079c276c6c5fbf55a6?pvs=4>

Submission

Due Date: 6/10(Mon) 23:59

You only need to submit the report pdf file **to the New E3 System**.

The file structure should look like:

 {student_id}_hw5.pdf
--

Wrong submission format leads to -10 point

Late submission leads to -20 points per day

Appendix 1: Tasks Example

Example 1

Task introduction

- **Motivation:** ChatGPT's comprehension of scenes and decision-making abilities.
- **Task description:** In this scenario, a robot is placed on a tabletop amidst various objects. The robot has several actions it can execute. We will assign specific tasks for the robot to accomplish, and ChatGPT must guide it to complete the tasks using the provided action options whenever feasible.
- **Input:** The input prompt consists of three parts, the first part specifies the available actions for the robot, the second part describes the objects on the table, and the third part lists several tasks for the LLM to solve.
- **Ideal output:**
 - Questions 1 to 3 have correct answers, ChatGPT should answer them correctly.
 - Questions 4 and 5 cannot be accomplished, ChatGPT should recognize them and provide reasons.

Simple prompt

Input: here we only specify what actions can the robot perform and what objects are there in the scene

You are a robot able to perform the following actions:

- * pick
- * place
- * open
- * close
- * hang
- * stack

In the kitchen, there are many objects:

- * a microwave on the table
- * a cup on the table
- * a bowl on the table
- * three cans with colors red, green, and blue on the table
- * a hook on the wall
- * a rack on the table

Describe how to achieve the following tasks with the combination of the provided actions step-by-step:

1. Stack three cans onto the rack in the order of blue green red
2. Hang the cup on the wall
3. Put the bowl into the microwave
4. Hang the rack on the blue can

5. Put the microwave into the cup

Output of ChatGPT

Sure, here are step-by-step instructions for each task:

1. ****Stack three cans onto the rack in the order of blue green red****:
 - Pick up the blue can.
 - Place the blue can on the rack.
 - Pick up the green can.
 - Place the green can on top of the blue can on the rack.
 - Pick up the red can.
 - Place the red can on top of the green can on the rack.
2. ****Hang the cup on the wall****:
 - Pick up the cup.
 - Open the hook on the wall.
 - Hang the cup on the hook.
 - Close the hook on the wall.
3. ****Put the bowl into the microwave****:
 - Pick up the bowl.
 - Open the microwave door.
 - Place the bowl into the microwave.
 - Close the microwave door.
4. ****Hang the rack on the blue can****:
 - Pick up the rack.
 - Hang the rack on the blue can.
5. ****Put the microwave into the cup****:
 - Pick up the microwave.
 - Place the microwave into the cup.

These steps should help you accomplish each task efficiently! Let me know if you need further clarification.

Since ChatGPT will generate different responses even with the same prompt, we've observed that sometimes it can generate correct answers. However, generating unreasonable answers like the above either shows the limitation of ChatGPT or the prompt is not precise enough.

Analysis

- Question 2: For a general hook, we don't need to open or close it, we can directly hang objects on it
- Question 4: The task cannot be completed since we cannot hang a rack onto the can
- Question 5: The task cannot be completed since the microwave is bigger than the cup

By analyzing these responses, we can observe that ChatGPT has a limited understanding of the actions it can perform as well as the objects' properties or functionality. Therefore, we propose providing a more comprehensive description of these actions and the attributes of the objects. This detailed information may facilitate a deeper understanding of the scene, enabling ChatGPT to complete the task.

Prompt with more detailed description and constraints

Input: specify the details of the actions can the robot perform and the description of some objects are there in the scene

- Describe how those actions are executed
- Describe the property and the functionality of those objects
- Add constraints that the task should be reasonable

You are a robot able to perform the following actions:

- * pick: grasp the object with the gripper
- * place: release the object with the gripper
- * open: grasp the handle to open something
- * close: grasp the handle to close something
- * hang: hang an object to another object
- * stack: put an object onto another object

In the kitchen, there are many objects:

- * a microwave on the table: a microwave with a door, you can open or close it
- * a cup on the table: the cup has a handle on it
- * a bowl on the table
- * three cans with colors red, green, and blue on the table
- * a hook on the wall: the hook is stuck on the wall with a part that can be hung

Describe how to achieve the following tasks with the combination of the provided actions step-by-step, if the task cannot be completed, please also describe the reason:

1. Stack three cans onto the rack in the order of blue green red
2. Hang the cup on the wall
3. Put the bowl into the microwave
4. Hang the rack on the blue can
5. Put the microwave into the cup

Output of ChatGPT

```
1. **Stack three cans onto the rack in the order of blue, green, red:**
  - **Pick:** Pick up the blue can.
  - **Stack:** Stack the blue can onto the rack.
  - **Pick:** Pick up the green can.
  - **Stack:** Stack the green can onto the blue can.
  - **Pick:** Pick up the red can.
  - **Stack:** Stack the red can onto the green can.

2. **Hang the cup on the wall:**
  - **Pick:** Grasp the cup by its handle.
  - **Hang:** Hang the cup onto the hook on the wall.

3. **Put the bowl into the microwave:**
  - **Pick:** Grasp the bowl.
  - **Open:** Open the microwave door.
  - **Place:** Place the bowl inside the microwave.
  - **Close:** Close the microwave door.

4. **Hang the rack on the blue can:**
  - This task cannot be completed as described because the rack is not
  mentioned in the initial setup, and there's no action to attach the rack
  to the blue can.

5. **Put the microwave into the cup:**
  - This task cannot be completed because the cup is smaller than the
  microwave, and there's no action to resize objects or place one object
  inside another.
```

Analysis

- Question 2: ChatGPT outputs the correct answer given a more detailed description of actions and object property
 - Question 4: ChatGPT points out why the task cannot be completed based on the object property and the given constraints
 - Question 5: ChatGPT points out why the task cannot be completed based on the object property and the given constraints
-

Example 2

Task introduction

- **Motivation:** GPT knowledge editing

- **Task description:** In this example, We will create a table setting with specific tools and tasks for GPT to complete, simulating its role as a robot. Subsequently, We will remove some critical tools and observe whether GPT can autonomously adjust its strategy.
- **Input:**
 1. Scene illustration, including provided tools and task descriptions.
 2. Prompts for modifying the tool list or providing alternative task descriptions.
- **Ideal output:** To make GPT has the ability to select the appropriate tools from the **existing** tool list, even when encountering unseen objects in the prompt.

Initial prompt

Input : Set the scenario and give the tools that can complete the task intuitively.

Imagine yourself as a robotic entity with grippers. Perform the assigned task sequentially utilizing the designated tools, indicating the tools employed at each stage.

Tools Provided: Spoon, knife, clip, spatula, chopstick.

Task:

Slice the loaf of bread and place it on the plate.

Output of GPT :

1. Tool: Knife

- **Action:** Grip the knife with the grippers.
- **Task:** Slice the loaf of bread into slices.

2. Tool: Spoon

- **Action:** Release the knife and grip the spoon.
- **Task:** Pick up the sliced bread pieces one by one and transfer them onto the plate.

3. Tool: Clip

- **Action:** Release the spoon and grip the clip.
- **Task:** Use the clip to hold the bread slices in place on the plate if needed.

Remove knife by prompt **remove knife from the tools provided** and do this task again:

1. **Tool:** Clip

- **Action:** Grip the clip with the grippers.
- **Task:** Use the clip to hold the loaf of bread in place.

2. **Tool:** Spatula

- **Action:** Release the clip and grip the spatula.
- **Task:** Slide the spatula under the loaf of bread to hold it steady.

3. **Tool:** Chopstick

- **Action:** Release the spatula and grip the chopstick.
- **Task:** Use the chopstick to slice the loaf of bread into slices.

4. **Tool:** Spoon

- **Action:** Release the chopstick and grip the spoon.
- **Task:** Pick up the sliced bread pieces one by one and transfer them onto the plate.

Based on the results, it seems that GPT comprehends the context and selects appropriate tools for the task. Moreover, it can readily update its tool list with a simple prompt. Now, I'll let this GPT perform another task under the same experimental conditions.



你

cut tomato



ChatGPT

To cut the tomato, we'll need to reintroduce the knife since it's the most suitable tool for slicing through its skin and flesh. Here's how we'll do it:

1. **Knife:** I'll grip the knife firmly and make a steady, controlled slicing motion through the tomato, cutting it into slices or wedges as desired.

From this outcome, I note that the tool "knife," which shouldn't appear on the list was chosen. This suggests that the GPT's knowledge wasn't properly modified by the preceding prompt. The tool selection is still influenced by the verb "cut."

Enhance the prompt with a more detailed description of the scenario

Input : For accurately remove "knife" from the list of tools, I employ the following prompt.



你

Set the scene once more, but this time, exclude the knife from the list of tools needed to cut the tomato.

output of GPT:

Task: Cut the tomato.

1. **Tool:** Clip

- **Action:** Grip the clip with the grippers.
- **Task:** Use the clip to hold the tomato in place.

2. **Tool:** Chopstick

- **Action:** Release the clip and grip the chopstick.
- **Task:** Use the chopstick to pierce and hold the tomato steady.

3. **Tool:** Spoon

- **Action:** Release the chopstick and grip the spoon.
- **Task:** Use the edge of the spoon to carefully slice the tomato into pieces.

4. **Tool:** Spatula

- **Action:** Release the spoon and grip the spatula.
- **Task:** Use the spatula to lift the sliced tomato pieces onto a plate or into a bowl.

Analysis :

After utilizing this prompt, we notice that the tool list has been appropriately updated. Additionally, even when presented **with different tasks involving unseen items**, the modifications made in the previous prompt are retained.



你

cut watermelon



ChatGPT

Task: Cut the watermelon.

1. Tool: Clip

- **Action:** Grip the clip with the grippers.
- **Task:** Use the clip to hold the watermelon in place on the cutting board.

2. Tool: Spatula

- **Action:** Release the clip and grip the spatula.
- **Task:** Slide the spatula under the watermelon to stabilize it.

3. Tool: Chopstick

- **Action:** Release the spatula and grip the chopstick.
- **Task:** Use the chopstick to mark a cutting line on the watermelon's surface.

4. Tool: Spoon

- **Action:** Release the chopstick and grip the spoon.
- **Task:** Insert the spoon along the marked cutting line and gently pry the watermelon apart.

<https://chat.openai.com/share/3b1bd7e2-44e0-4a83-8fb0-adf8b83f4e82>