AI 3603 Artificial Intelligence: Principles and Techniques Homework 1

Adhere to the Code of Academic Integrity. You may discuss background issues and general strategies with others and seek help from course staff, but the implementations that you submit must be your own. In particular, you may discuss general ideas with others but you may not work out the detailed solutions with others. It is never OK for you to see or hear another student's code and it is never OK to copy code from published/Internet sources. Moss (Measure Of Software Similarity) will be used for determining the similarity of programs to detect plagiarism in the class (`https://theory.stanford.edu/~aiken/moss/`) If you encounter some difficulties or feel that you cannot complete the assignment on you own, discuss with your classmates in Discussion forum on Canvas, or seek help from the course staff.

You are required to complete this homework ***individually***. Please submit your assignment following the instructions summarized in Section 6.

1 Task 1: Prompt Engineering

The attachment provides the test papers and answer analysis for three subjects from Shanghai's 2024 college entrance examination. You are required to select one question from each subject (three questions in total), and think about how to design the prompt so that GPT can generate accurate answers to these questions. Please document your trials in a report. The report should include, but is not limited to, the following:

- Your train of thoughts when designing the prompts.

- The input to ChatGPT and its output.

- Your attempts to adjust the prompts according to ChatGPT's answers.

- Evaluate ChatGPT's performance on different questions and analyze which subject it may be better at.

  Guide to the Prompt Engineering: https://www.promptingguide.ai/

2 Task 2: LLM Evaluation

In this task, you are required to choose some (at least two) LLMs, and evaluate their performance utilizing same questions designed by yourself. You are required to evaluate the performance on multiple dimensions, such as:

- Logical performance: evaluate whether the LLM can tackle logical problems, such as founding that there is a paradox in the question you proposed.

- Coding performance: whether the LLM solves the problem you proposed by generating correct codes.

- Mathematical problem-solving: whether the LLM can solve math problems correctly.

  Please show your questions, LLM outputs, and corresponding scores in the report. Tell us which is the best LLM in your leaderboard.

3 Task 3: LLM Failures

ChatGPT tends to perform poorly in answering questions concerning logic, self-awareness, inference and so on. In this task, you are required to come up with a case where LLM fails, and think about how to guide LLM to generate the right answers by modifying prompts.

4   Online Resources

1. ChatGLM (https://chatglm.cn/): official website for ChatGLM.

2. ChatGPT (https://chat.openai.com/): official website for ChatGPT.

3. Doubao (https://www.doubao.com/): official website for Doubao.

4. Yiyan (https://yiyan.baidu.com/): official website for WenXinYiYan.

5. QWen (https://tongyi.aliyun.com/): official website for TongYiQianWen.

Other LLMs are also acceptable for this homework, such as Bard, Xinghuo, Claude, etc.

5   Discussion and Question

You are encouraged to discuss your ideas, ask and answer questions about this homework. If you encounter any difficulty with the assignment, try to post your problem **on Canvas** for help. The classmates and the course staff will try to reply.

6   Submission Instructions

1. Zip all your files to a file named as **HW1_ID_name.zip**.

2. Upload the file to the homework 1 page on the Canvas.

3. **Deadline**: 2024.09.30 22:00