HW1 – Prompt Engineering

[TA] 曹立武

Submission deadline: 2025/3/17 23:55

Objectives

- In this homework, you will learn to use the prompt engineering techniques with Free LLM API (Gemini / Groq) on part of the MMLU dataset.

- Try as many prompt engineering techniques as you can.

- The sample data is provided for you to verify the correct format or to use it for the prompt design.

Data format in MMLU dataset

- Input: 學科問題
- A/B/C/D : 四個選項
- Task : 學科
- Target: 正確答案(A or B or C or D)

- Sample file: mmlu_sample.csv (Input, A/B/C/D, Task, Target)
- Benchmark file: mmlu_submit.csv (Input, A/B/C/D, Task)
  - Predict / Generate: Target
  - Submission format: submit_format.csv (Question ID, Target)

Grading Policy

- Baseline: 80
- Public: 10 (Top 25%), 5 (Top 25~50%)
- Private: 10 (Top 25%), 5 (Top 25~50%)

Some tips

- Different LLMs might have slight variance to control the output format
- Parse the output
- Get the best use of different LLM output to check the answer if needed

Rules

- Use your student ID as the team name on Kaggle.
  - You are not allowed to use multiple IDs, otherwise we will not adopt your score for ranking.
- A maximum of 3 submissions per day is allowed on Kaggle.
- Write your own code.
  - You won’t get any score if your code is too similar to others. (Plagiarism check)
- Design your own prompt strategy.
- You need to upload the code to E3 that generates your answer.
- You can only use the Free API from Gemini and Groq. (The baseline can be achieved by these)

Submissions

- Submit your results to Kaggle:
  - https://www.kaggle.com/t/7ec7362b76c346959b1ad79a291d34a8
- Submit your zipped source code {student_id}.zip to E3. The zip file should contain
  - {student_id}/main.py
  - {student_id}/prompt.txt
    - Prompt you use: ... (Design by yourself)
    - API you use: ... (For example, Gemini-2.0-Flash)
    - Your prompt strategy: ... (For example, Few-Shot? Chain-of-Thought?)
  - {student_id}/requirements.txt (If you need to download some libraries)

Homework information

- Deadline: 2025/3/17 23:55
- You can send an e-mail through E3 if you have any problems.
- [TA]曹立武

- If you are interested in solving this kind of problem, there are more challenges for your reference.
  - Multimodal dataset: https://huggingface.co/datasets/MMMU/MMMU
  - Traditional Chinese dataset: https://arxiv.org/html/2403.01858v3