Foundations of Natural Language Processing

Peking University, 2025

Assignment

Due: Sunday, March 23, 2025, 11:59 PM

Task Description

While large language models (LLMs) have achieved remarkable success in various tasks and domains, they are not omnipotent. In this assignment, you will explore the limitations of current LLMs by designing **at least two** challenging questions that they **fail to answer correctly**.

Specifically, you should

1. **Design Challenging Questions**
    a. The questions should be **clear, well-defined, and solvable**.
    b. The topic of the questions can be anything—related to your major, a personal interest, or any domain where LLMs might struggle. As a (not-that-exciting) example, you could pose a particularly difficult math problem (e.g., one from the International Mathematical Olympiad).
    c. The question must be **text-based**—no images, videos, or audio.
    d. The questions can be written in either Chinese or English.
2. **Provide a Reference Answer**
    a. For each question, include a well-reasoned reference answer that demonstrates the correct solution.
3. **Evaluate Powerful LLMs**
    a. Submit responses from **two different powerful LLMs** (e.g., ChatGPT, DeepSeek, Claude, Qwen).
    b. Analyze their responses and explain why they are incorrect.

Submission Format

- Please submit your assignment in a **JSON file** following this format:

```
[
   {
      "id": 1,
      "question": "Your question here",
      "reference_answer": "Your reference answer here",
      "model_responses": [
         {
            "model": "DeepSeek-R1",
            "output": "The model's response",
            "remark": "Explanation of why the response is
incorrect"
         },
         {
            "model": "GPT-4o",
            "output": "The model's response",
            "remark": "Explanation of why the response is
incorrect"
         }
      ]
   },
   {
      "id": 2,
       ...
   }
]
```

- Name your file as “ID_NAME.json", such as “2300011111_张三.json”.
- Please ensure that your file is properly formatted and follows the expected structure. **You can use the provided Python script (check_format.py) to validate your JSON file before submission.**

Important Notes

1. **Deadline:** Sunday, March 23, 2025, 11:59 PM
2. **Grading Criterion:** Your grades are based on whether the submitted questions meet all the requirements described above (on-time submission, correct JSON

format, at least two clearly-stated questions, reasonable reference answers, incorrect responses from two LLMs, etc.).

3. **Late Submission Policy:**
    a. Late submissions will incur a 5% deduction per day from your final grade.
    b. **No submission will be accepted after March 30**, one week after the due date.
    c. Your submission time is determined by the timestamp of your last upload to course.pku.edu.cn.
    d. Please be aware that all late submissions, regardless of the reason, will be penalized according to the late submission policy. While we understand that unexpected events can occur, it is your responsibility to plan accordingly and submit your homework on time.
4. **Contact:** If you have any questions, feel free to contact the TA zhangch@pku.edu.cn (张晨).