Assignment 4

Please submit the code and report (in pdf format) on Blackboard system before 23:59 Jan. 4. Report can be written either in English or Chinese. Name your report as `studentID_Name.pdf`

Q1

[3 points] Model interpretability is a key research area in understanding LLM behaviors. Read the paper Emergent Introspective Awareness in Large Language Models and use an LLM to help answer the **10 classic research paper reading questions** suggested in this blog.

- Use an LLM (e.g., ChatGPT or similar) to generate answers to each question based on the paper.
- Select three QA pairs among the LLM-generated contents to add your own **commentary** (agreement/disagreement, new insights, questions, etc).

Q2

[3 points] Large language models often face challenges in **information fusion tasks**, including data omission, hallucinations, and inconsistency. This task explores these limitations through case analysis.

> Task Instructions:

1. **Scenario Selection**: Choose a real-world application scenario (e.g., news, healthcare). Focus on common LLM failures during information integration.
2. **Data Preparation**: Collect **3 articles (or news reports)** related to the same event (around 600 words, in Chinese or English, you can search them online and ensure these are different documents), published within the past **six months**. We add this filter to target at those information that is not trained during pretraining.
3. **Model Test**: Write prompt to use GPT-5 or another proprietary LLM to **fuse the contents** of the three articles into one unified article.
4. **Issue Analysis**: Evaluate and analyze potential issues (e.g., data loss, factual errors, hallucinations) in the output. You can write your report and commentary following the json format below.

> Output Format Example:

```
[
  {
    "text1": "Content of input document 1",
    "text2": "Content of input document 2",
    "text3": "Content of input document 3",
    "fusion_text": "LLM-generated fused summary",
    "evaluation_problems": [
      # should clearly identify paragraph numbers (e.g., `"Paragraph 3"`).
      "Omission: Paragraph 3: Missing key information from input document 2",
      "Incorrect data: Paragraph 5: Mistaken event date from 2023 to 2024",
      "Hallucination: Paragraphs 7-8: Fused summary contains facts not in any input"
    ]
  }
]
```

Q3

[4 points] Use the DeepResearch framework to explore **LLM-based deep scientific reasoning**.

> Task Instructions

1. Access the **Bailian online service** linked in the repository under:
   "👋 Welcome to try Tongyi DeepResearch via our Modelscope online demo or 🤗 Huggingface online demo or bailian service!"
2. Choose a graduate-level **mathematical modeling problem** and let the LLM solve it via the DeepResearch interface.

In your report, include:

- The problem you posed. You can refer to the problems in the MCM/ICM contest
- The LLM's response.
- Your **critical evaluation** of the response:
  - What was correct or helpful?
  - What errors or limitations did the LLM show?
  - Discuss **strengths and weaknesses** of using LLMs for deep research.