大模型竞赛考试程序

2025年1月31日

注意事项

- 考试日期：2025年2月15日全天
- 考试通过在服务器API调用进行（无法连接外网）
- 参加考试的每组将被分配一个唯一的时间段和GPU
- 每组考试时间30分钟，含调试时间。过时不候
- 每组只有一次作答机会
- 作答方式：每组在指定时间内开启openAI相容的API端口。考官通过调用程序向端口提问并记录答案。答题中间步骤不计入分数。

考前程序调试

- 在服务器上启动openAI API服务。以vllm为例：

创建vllm环境并安装之

```
conda create -n vllm python=3.11
pip install vllm
```

```
(vllm) ➜  vllm vllm serve /data/konformal/nn/model/Qwen2.5-14B --port 3030
```

你参与考试的模型存放位置

为防止端口冲突，请指定一个不常见端口

命令行查看API端点运行情况

模型名字

```
(base) ➜  LLM_competition curl -X GET http://127.0.0.1:3030/v1/models
{"object":"list","data":[{"id":"/data/konformal/nn/model/Qwen2.5-14B","object":"model","created":173
8335826,"owned_by":"vllm","root":"/data/konformal/nn/model/Qwen2.5-14B","parent":null,"max_model_len
":131072,"permission":[{"id":"modelperm-3aa2205c8d714aec99d400571b7087a9","object":"model_permission
","created":1738335826,"allow_create_engine":false,"allow_sampling":true,"allow_logprobs":true,"allo
w_search_indices":false,"allow_view":true,"allow_fine_tuning":false,"organization":"*","group":null,
```

考前程序调试

- 建立一个目录并将考试程序和测试题样例复制到此

```
(base) ➜  LLM_competition ls
llm_solve_promblem.py  test_problem.json
```

考前程序调试

- 修改llm_solve_problem.py中的对应内容

- 考试前将修改后的llm_solve_problem.py发给监考老师。具体由哪位老师与你对接将在考前通知

```python
from openai import OpenAI
import json

api_base = "http://127.0.0.1:3030/v1"
api_key = "dummy_kep"
model_name = "data/konformal/nn/model/Qwen2.5-14B"
problem_set = "test_problem.json"
group_num = "001"

client = OpenAI(
    base_url=api_base,
    api_key=api_key,
)

def predict(message):
    response = client.chat.completions.create(
        model= model_name,
        messages=[
            {'role': 'user', 'content': message}],
        max_tokens=8196,
        temperature=0.5,
        stream=False
    )
    return response.choices[0].message.content


if __name__=="__main__":
    with open(problem_set, 'r', encoding='utf-8') as file:
        data = json.load(file)

    for i in range(len(data)):
        problem = data[i]["problem"]
        print(f"Working on problem {i}.")
        answer = predict(problem)
        print(f"problem {i} done!")
        data[i]["answer"] = answer

    with open(f'Answer_{group_num}.json', 'w', encoding='utf-8') as file:
        json.dump(data, file, ensure_ascii=False, indent=4)
```

改成你的端口！

改成你的模型名称！

改成你的组号！

调用参数可酌情修改

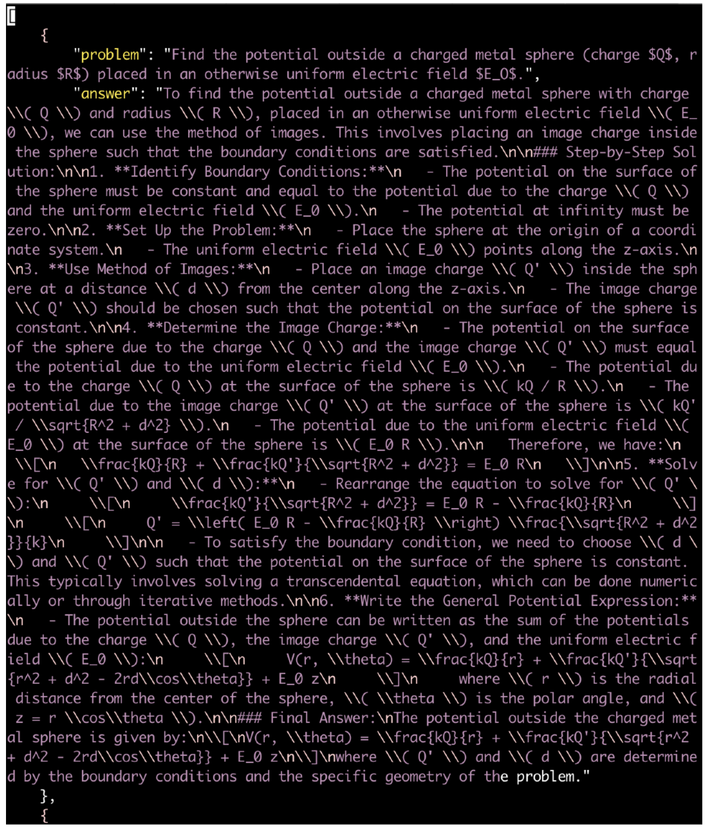
考前程序调试

- 运行程序测试作答情况

```
(vllm) ➜  LLM_competition python llm_solve_promblem.py
Working on problem 0
problem 0 done!
Working on problem 1
problem 1 done!
Working on problem 2
problem 2 done!
Working on problem 3
problem 3 done!
Working on problem 4
problem 4 done!
Working on problem 5
problem 5 done!
Working on problem 6
problem 6 done!
Working on problem 7
problem 7 done!
Working on problem 8
problem 8 done!
```

- 打开作答结果检查是否有答案输出

vim Answer_001.json



考试愉快！