1. 准备数据

实验数据	链接	说明	分工
deita	https://hf-mirror.com/datasets/hkust-nlp/deita-6k-v0?row=0https://hf-mirror.com/datasets/hkust-nlp/deita-6k-v0?row=0	多轮对话数据，去掉最后一轮assistant的response。以之前的对话历史为输入，生成assistant的回复，判别其质量	骜桀下载数据然后提供一个服务器的路径到群里面，这样明楷也可以使用这份数据完成后面的内容

2. 采用lmdeploy工具（https://lmdeploy.readthedocs.io/en/latest/llm/pipeline.html）实现模型的部署和推理：

测试模型	链接	分工
Internlm2.5 (7B/20B)	https://hf-mirror.com/internlm/internlm2_5-20b-chat https://hf-mirror.com/internlm/internlm2_5-7b-chat	骜桀
Llama 3.1 (8B)	https://hf-mirror.com/meta-llama/Meta-Llama-3.1-8B-Instruct	骜桀
Qwen2 (1.5B 7B)	https://hf-mirror.com/Qwen/Qwen2-1.5B https://hf-mirror.com/Qwen/Qwen2-7B-Instruct	明楷
GLM4 (9B)	https://hf-mirror.com/THUDM/glm-4-9b-chat	明楷

   a. 需要完成两种不同的推理方式
      i. <Q,R>: 直接让LLM根据对话上下文，生成新的回复
      ii. <Q,C,R>: 首先LLM根据对话上下文，补充一个query问题，让LLM生成一组evaluation criteria，然后根据生成得到的evaluation criteria，生成一个response要求response满足evaluation criteria的要求。如下面例子所示，其中多轮对话历史最后一轮为“为什么天空是蓝色的”。然后，补充一个空的response然后开启新的对话内容，即让LLM生成高质量的criteria，同时可以提供几个criteria的case作为示例（我随便加了一个

accuracy的case，你们可以酌情补充3-4个，并提示语言模型请根据user的问题来定制evaluation criteria,不同的user query对应的evaluation criteria不尽相同）。最后，LLM生成完criteria之后让其生成满足criteria要求的高质量回复

{"role": "user", "content": "为什么天空是蓝色的"}
{"role": "assistant", "content": "<PLACEHOLDER>"}
{"role": "user", "content": "Please generate a list of the detailed evaluation criteria, and a higih-quality response should satisfy all these criterias. For example, ## accuracy \nDescription: All contents provided or mentioned in the response should be accurate and correct. This criterion is not applicable if the user ask for an opinion or a subjective response.\n\n…"}
{"role": "assistant", "content": "# Criteria List\n## …"}
{"role": "user", "content": "Now, please generate your answer, which perfectly answer the previous user query and satisfy the proposed evaluation criteria."}

3. 推理完成之后，采用reward model来对生成的回复的质量进行打分（即给一个完整的对话上下文历史，和模型生成的最后一轮assistant的回复整体一起用reward model来评估打分，分数越高代表response质量越好）
   a. 采用Internlm2-20B-reward 来实现自动打分，具体教程请见如下链接：https://huggingface.co/internlm/internlm2-20b-reward