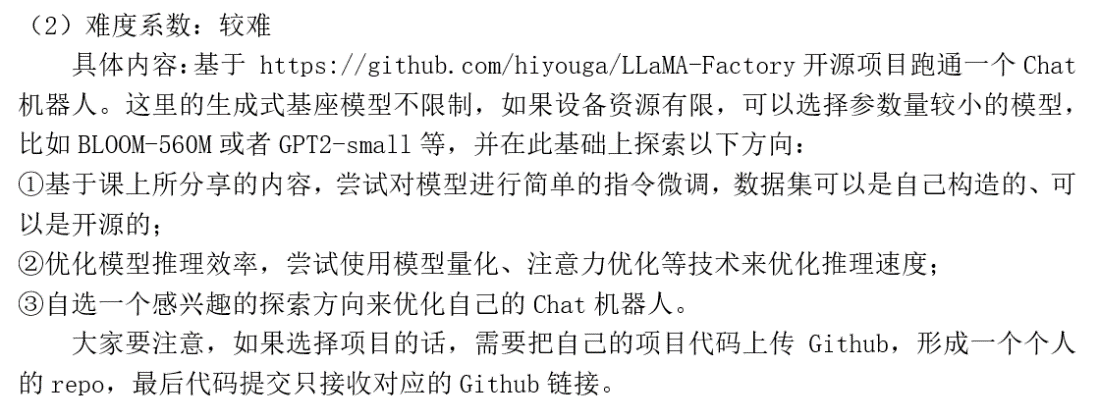
**自然语言处理大作业项目报告**

**一．选题：**

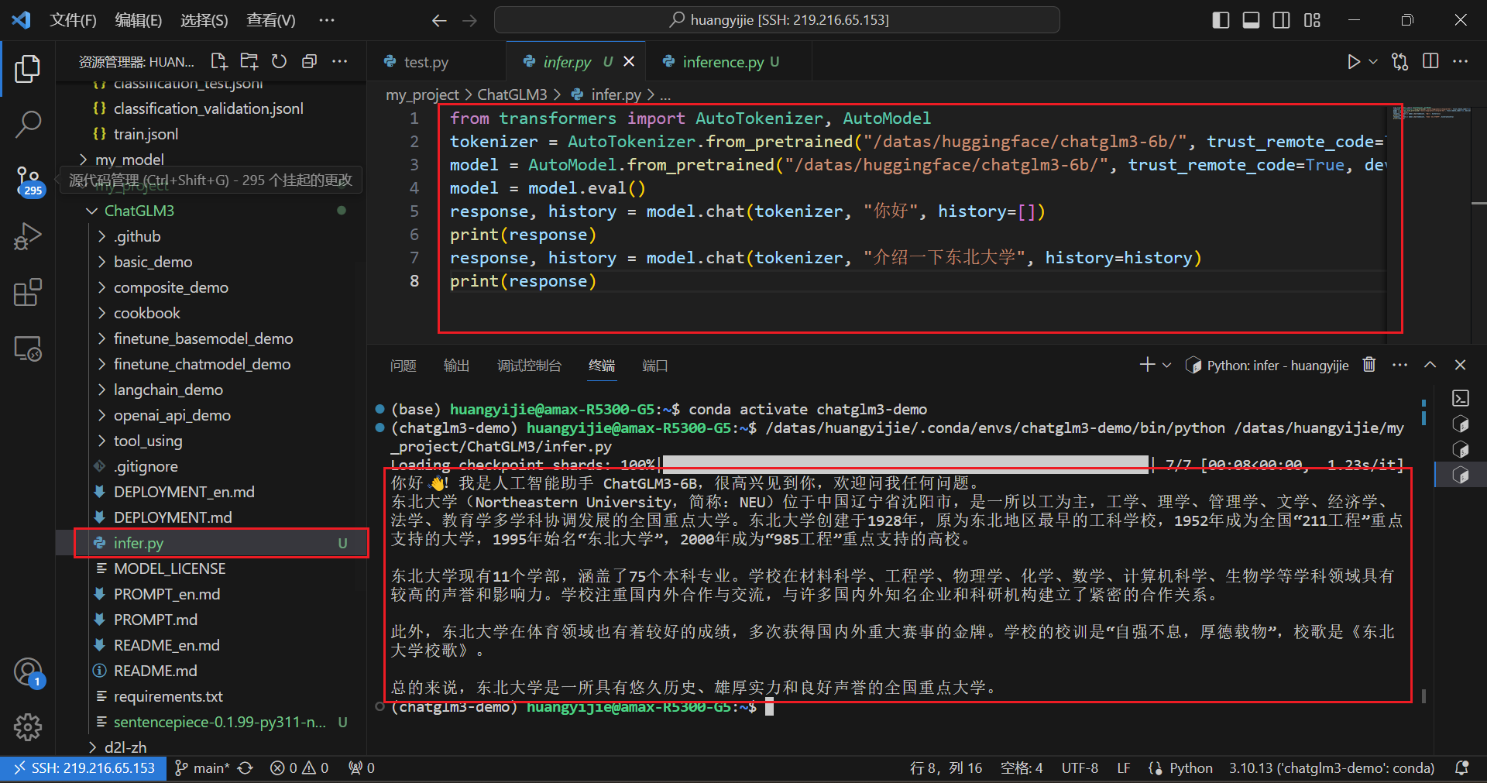
****

**二．准备工作：**

1. 基座模型：ChatGLM3-6B（服务器中已经部署了ChatGLM3-6B基座模型）

2. 实验环境：课题组A6000服务器

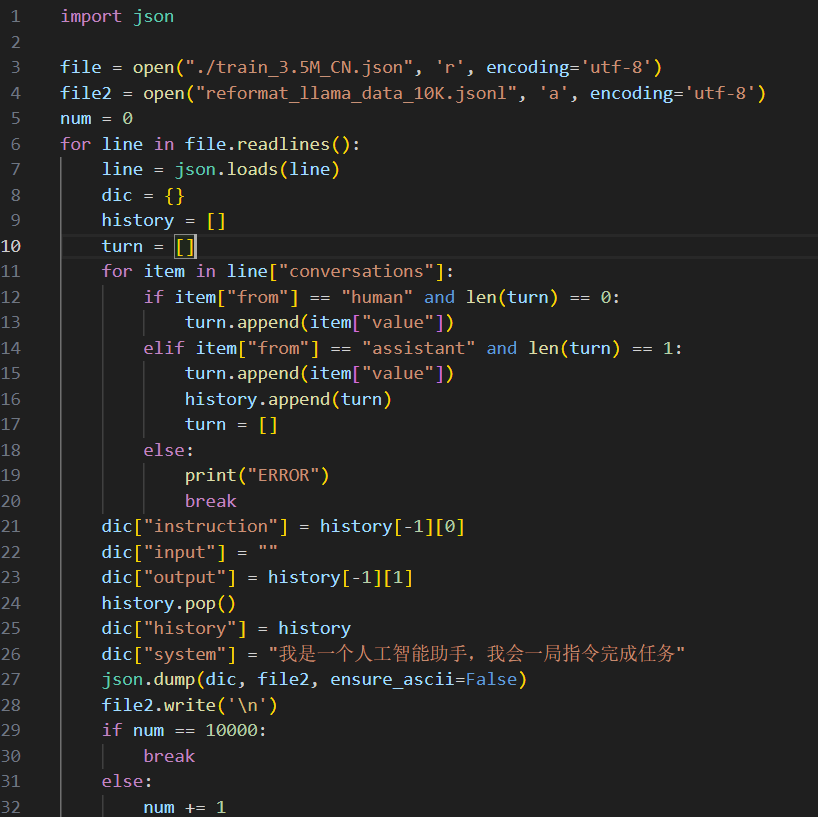
**三．完成内容：**

1. 按照ChatGLM3官方说明文档跑通了代码，可以调用 ChatGLM3 模型来生成对话：

2. 选用belle数据集进行修改格式，并进行指令微调：

（1）将belle数据集改为alpacha数据集格式。

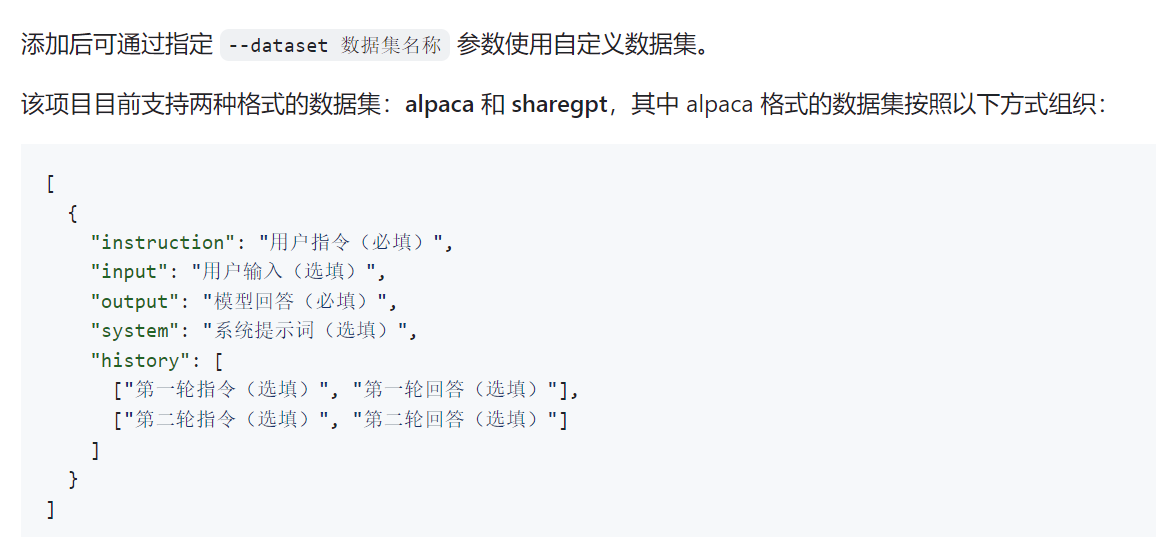
修改数据集格式代码：

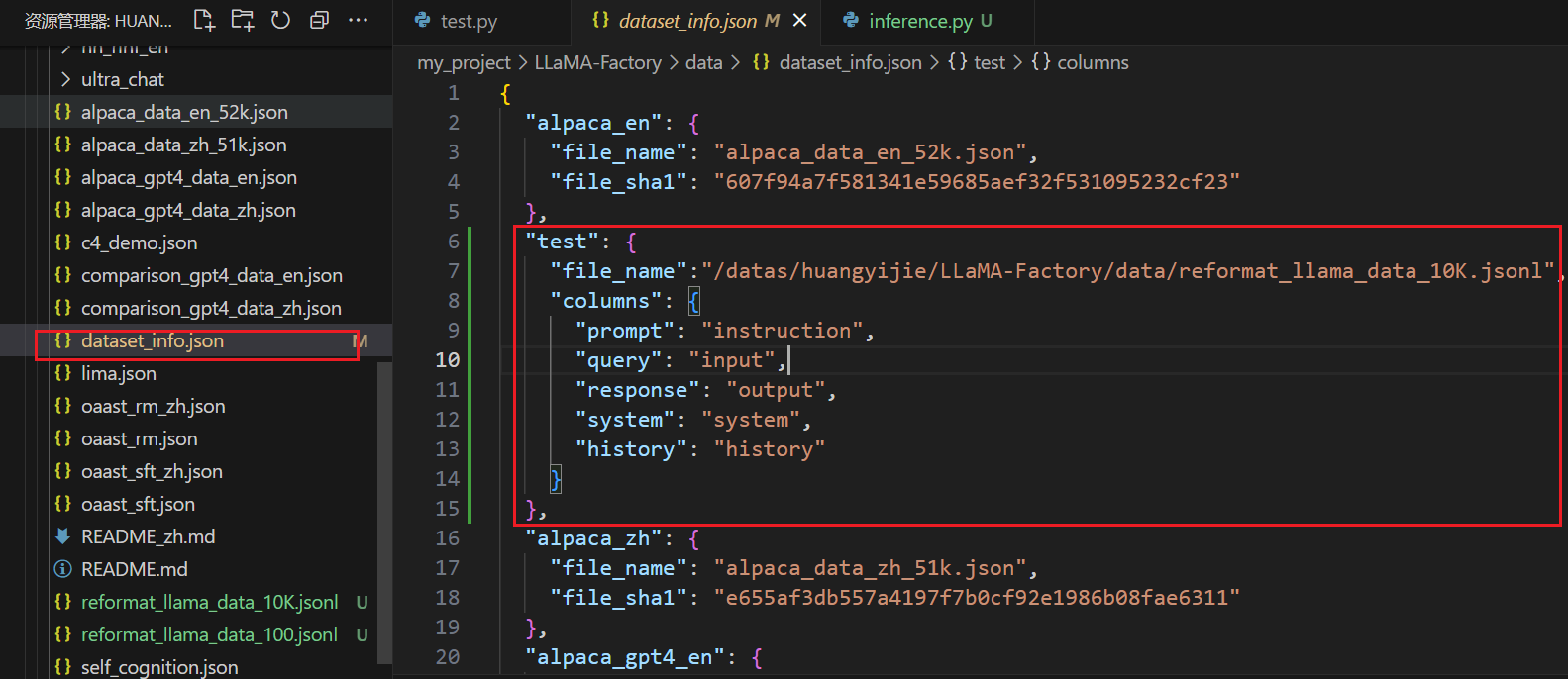


修改之后的数据集格式：



（2）根据LLaMA-Factory官方说明文档（https://github.com/hiyouga/LLaMA-Factory/blob/main/data/README\_zh.md），先在dataset\_info.json文件中加入我自己定义的数据集（alpaca数据集格式）：

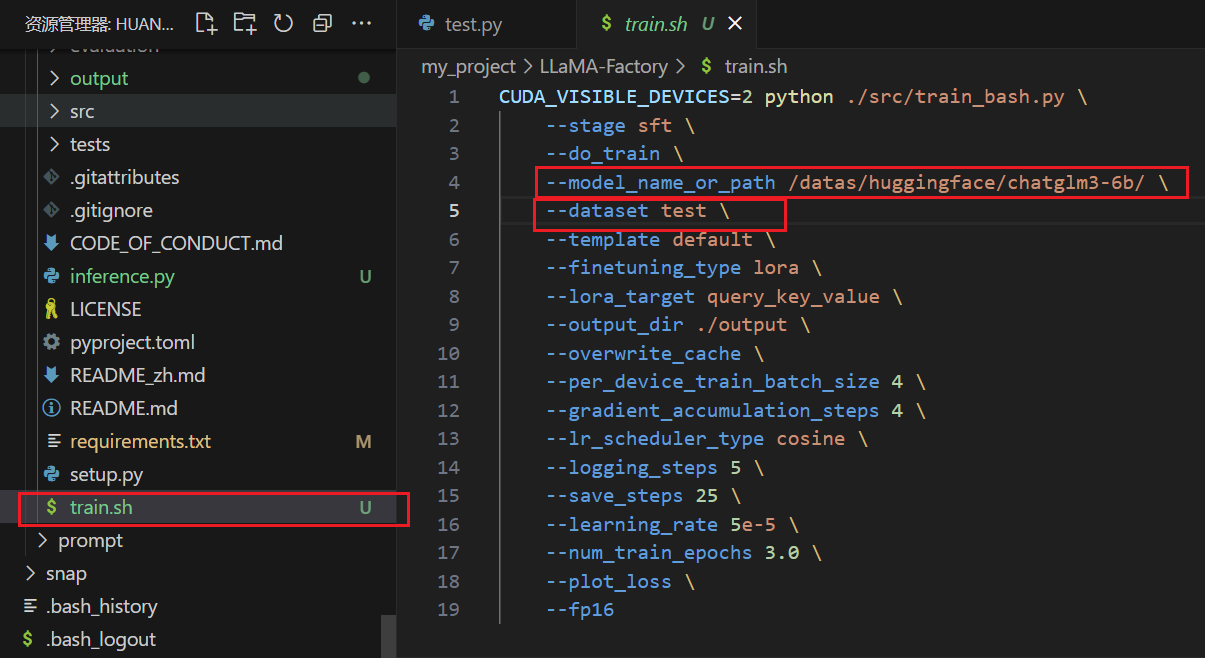




（修改完的dataset\_info.json文件）

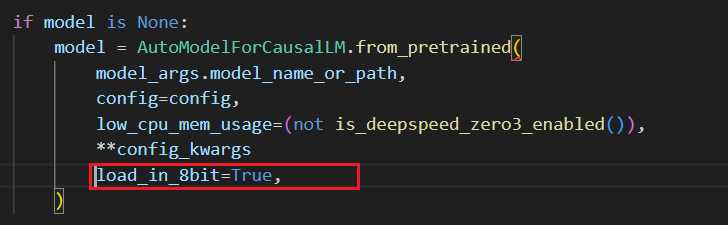
（3）进行指令微调：

修改train.sh文件中的模型路径和数据集名称



运行CUDA\_VISIBLE\_DEVICES=0 python train\_bash.py命令进行指令微调

3. 优化模型推理效率



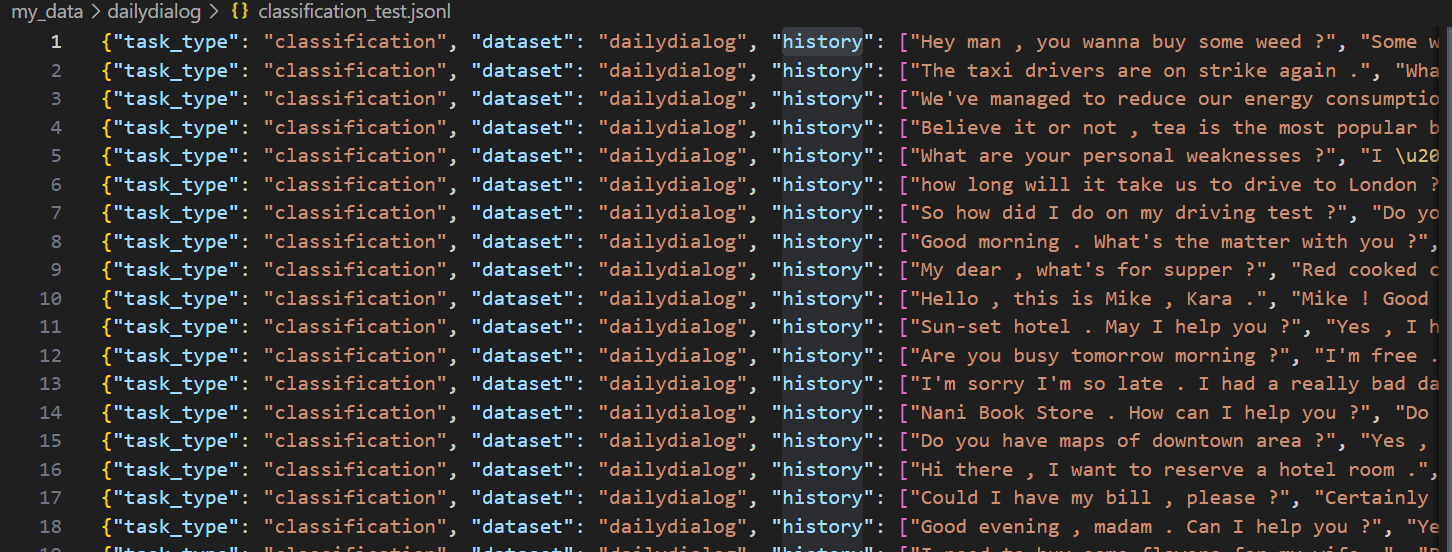
4.因为我的研究方向是通用情感分析，所以我尝试将这个模型应用在在EIC（对话中的情感推理）任务上：

（1）选择dailydialog数据集（https://huggingface.co/datasets/NEUDM/dailydialog）并进行格式修改：

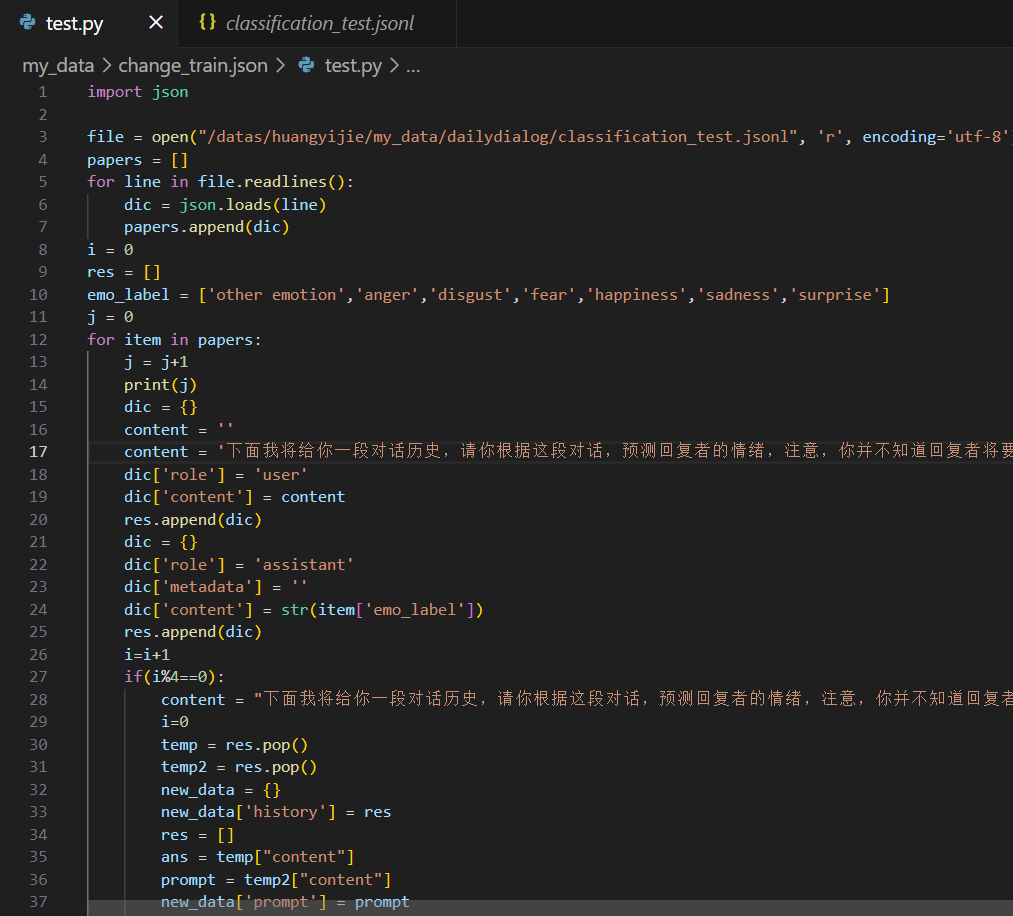
Dailydialog数据集可以用于生成任务以及分类任务，我选择分类任务。将原数据集中的对话历史——history每三个进行拼接，形成ChatGLM3 history的格式：

[{'role': 'user', 'content': '你好'}, {'role': 'assistant', 'metadata': '', 'content': '你好👋！我是人工智能助手 ChatGLM3-6B，很高兴见到你，欢迎问我任何问题。'}]

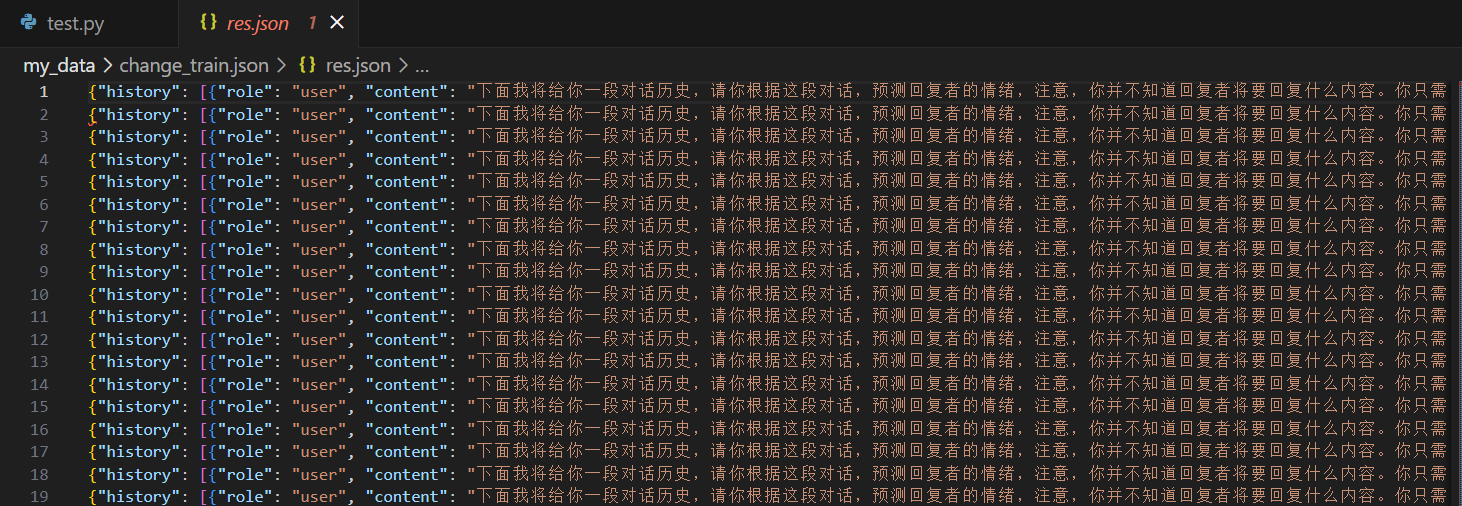
并将下一条history作为要输入给模型的prompt：



（原数据集格式）

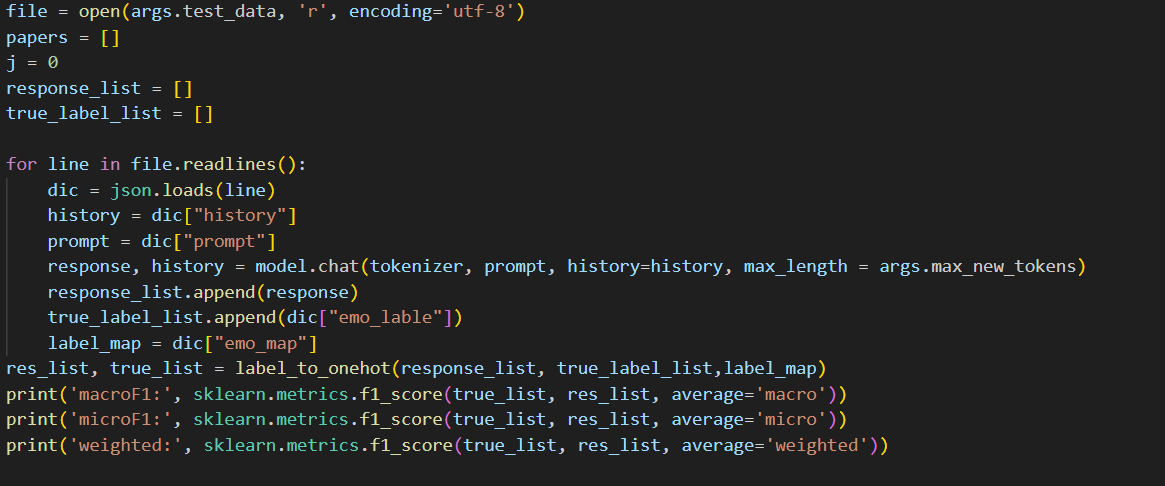


（修改数据集格式代码）



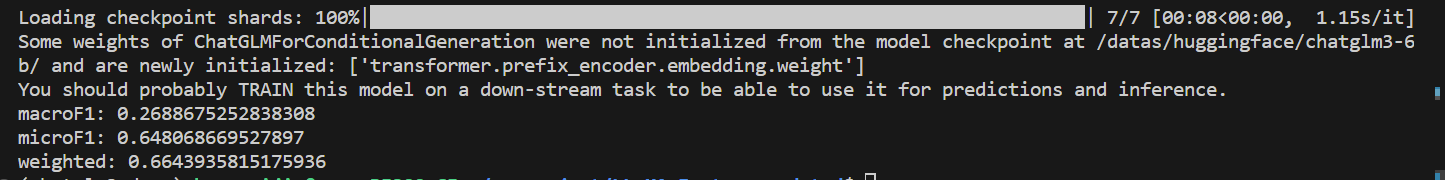
（修改后的数据集）

运行CUDA\_VISIBLE\_DEVICES=0 python inference.py命令进行推理。接收模型的response并与正确结果进行对比，调用sklearn.metrics库中的方法计算macroF1等分数：



（inference.py部分代码）

模型结果：



5.今后将继续在情感分析的多种子任务上进行相关数据集的改写继续进行实验。