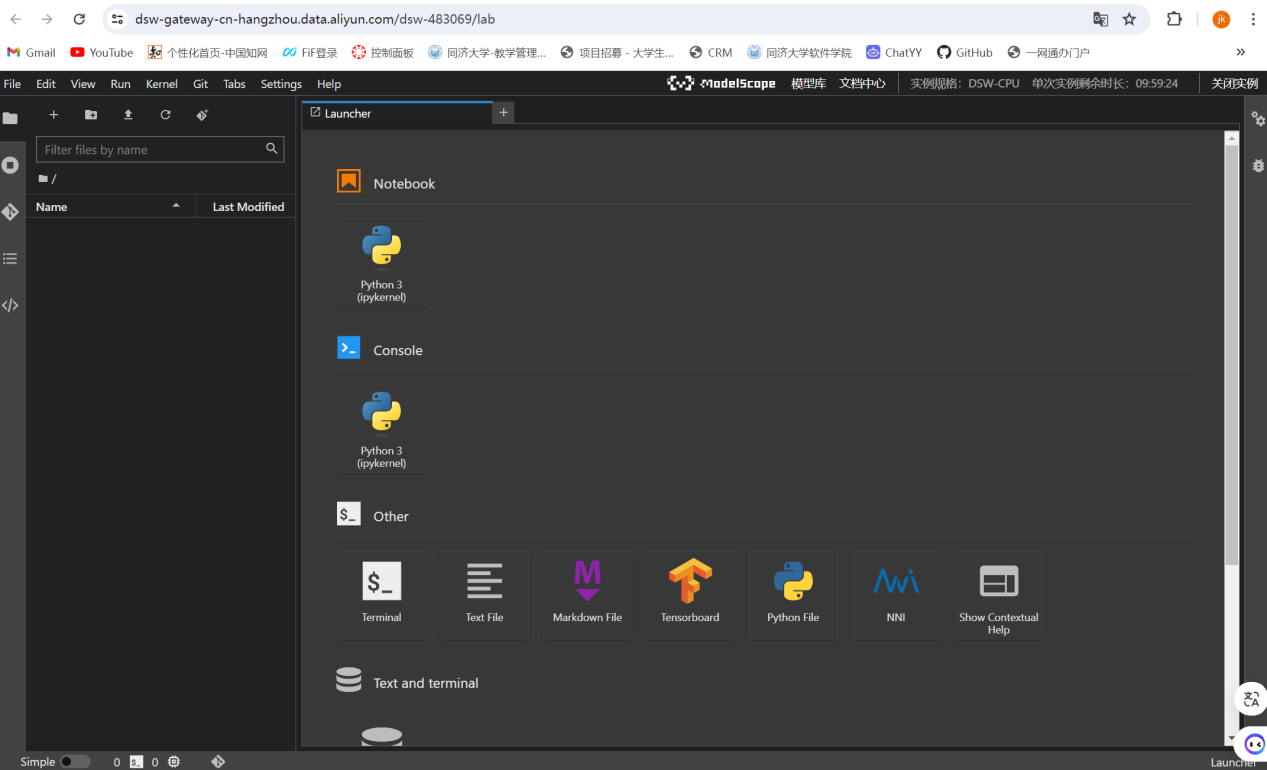
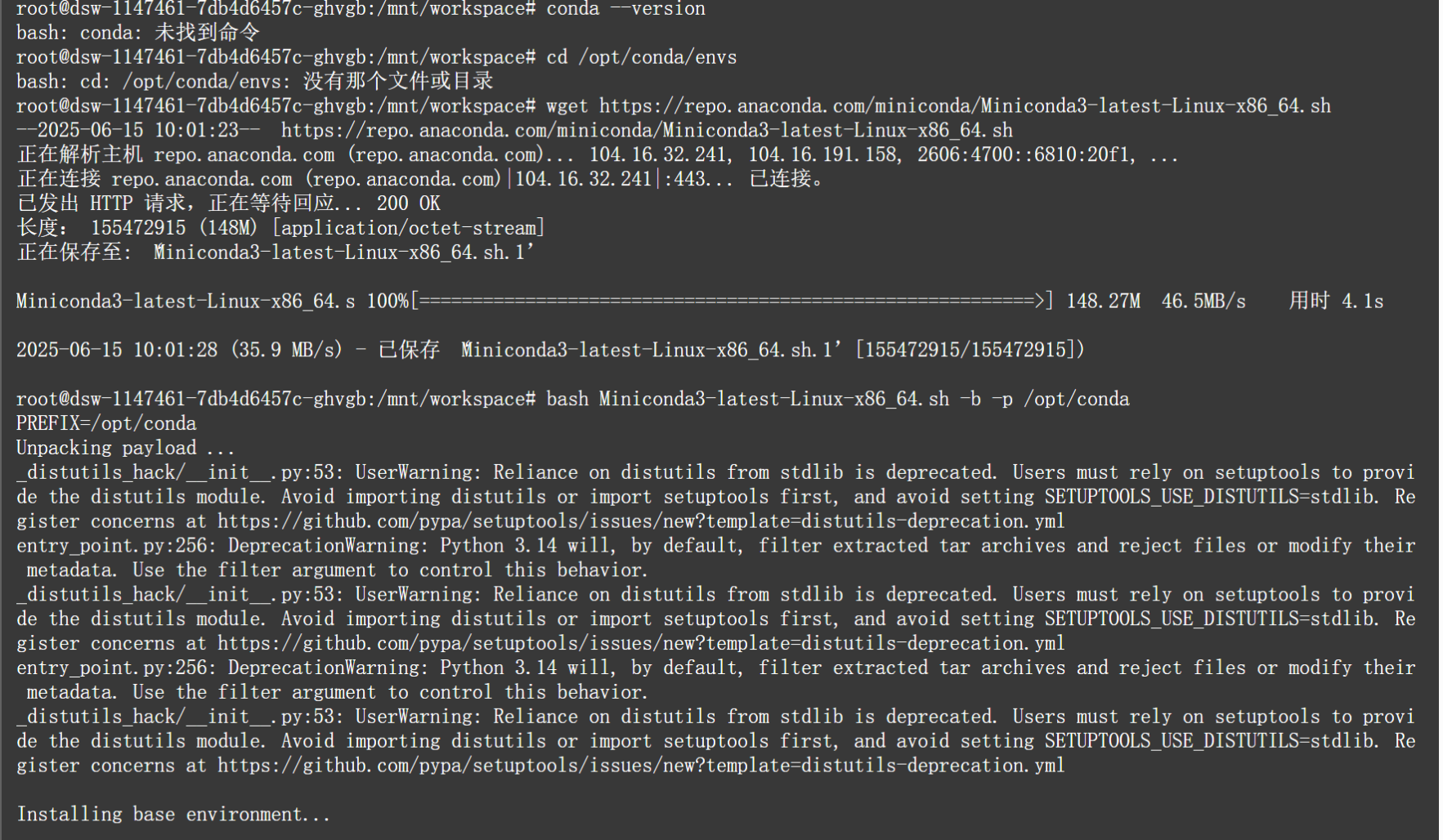
**大模型部署报告**

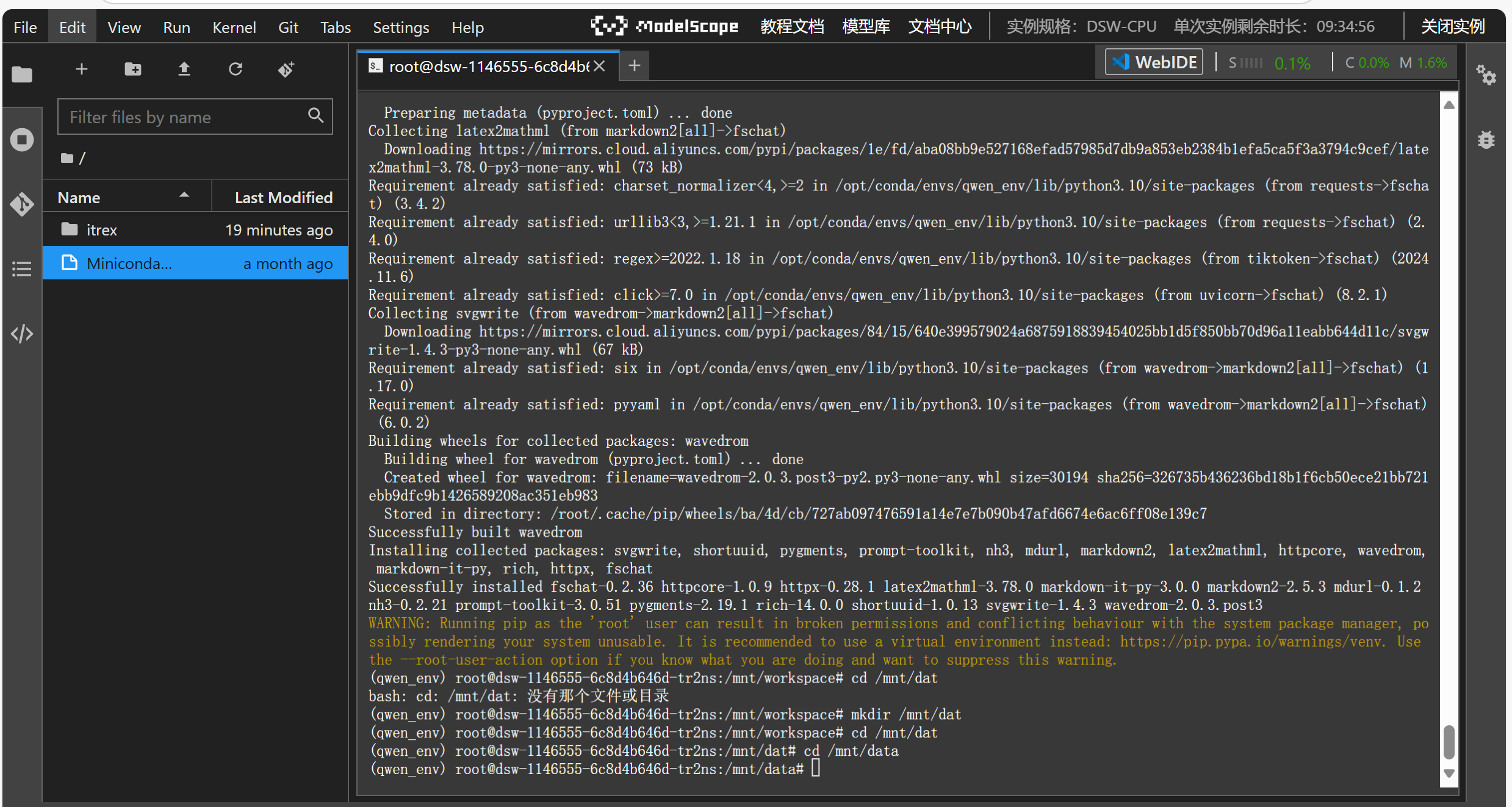
**1.配置和部署(此部分对应作业要求：完成git clone相关git的截图或部署完成的相关截图)**

注册完成后打开NoteBook



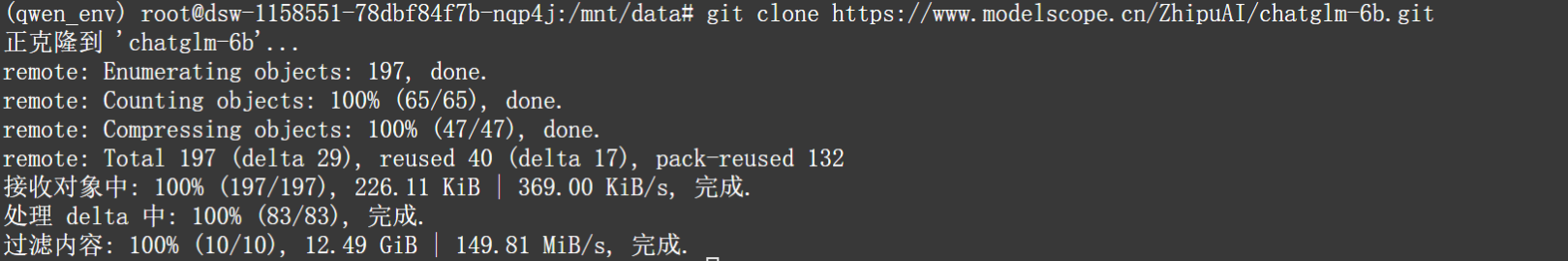
环境搭建





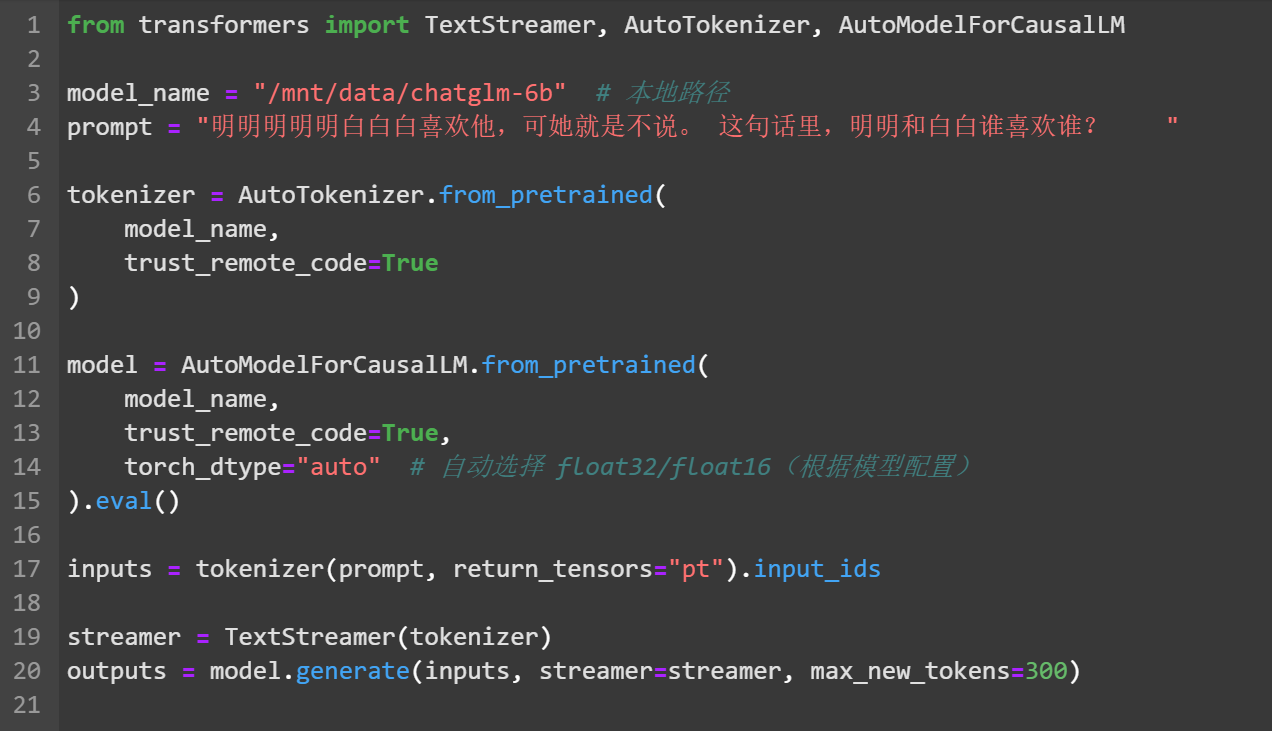
根据实验需要下载相对应的中文大模型至本地(完成git clone相关git的截图或部署完成的相关截图)





**2.对话与分析(此部分对应作业要求：问答测试结果的相关截图和语言模型之间的横向对比分析)**

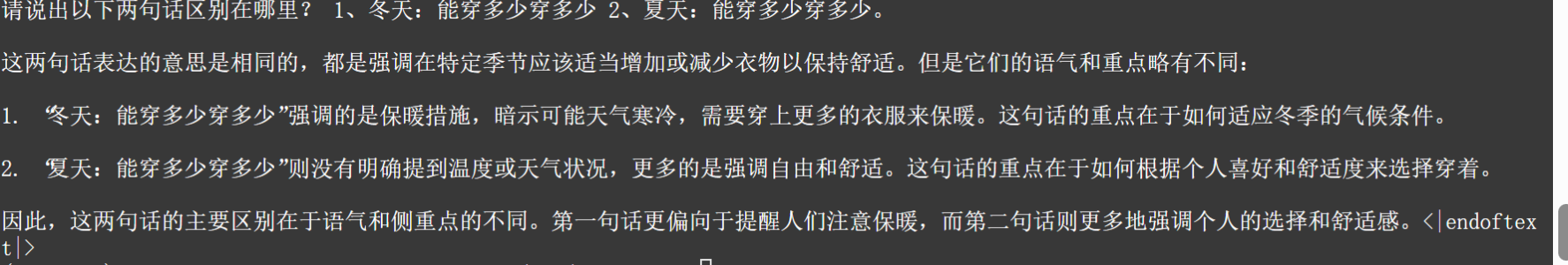
2.1编写Python文件运行ai，更换问题只需更改prompt，更换模型需要更改路径。



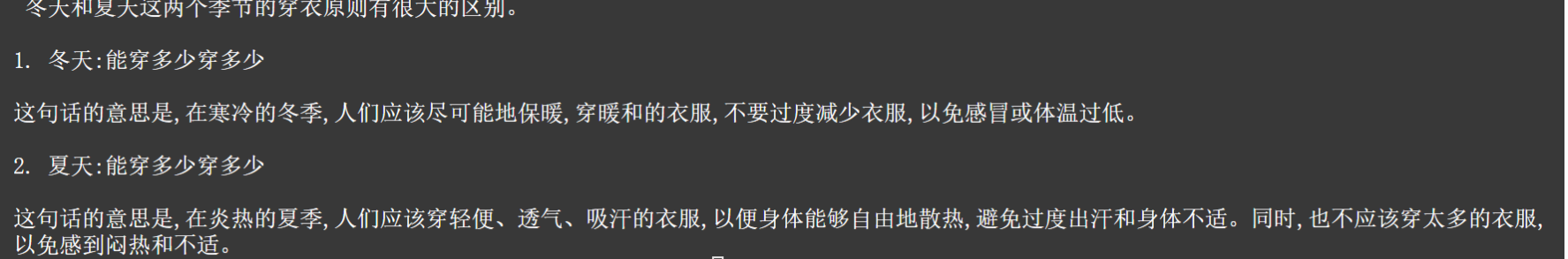
2.2.对话测试和横向分析比较

问题一.请说出以下两句话区别在哪里?1.冬天：能穿多少穿多少；2.夏天：能穿多少穿多少；

通义千问Qwen-7B-Chat回答：



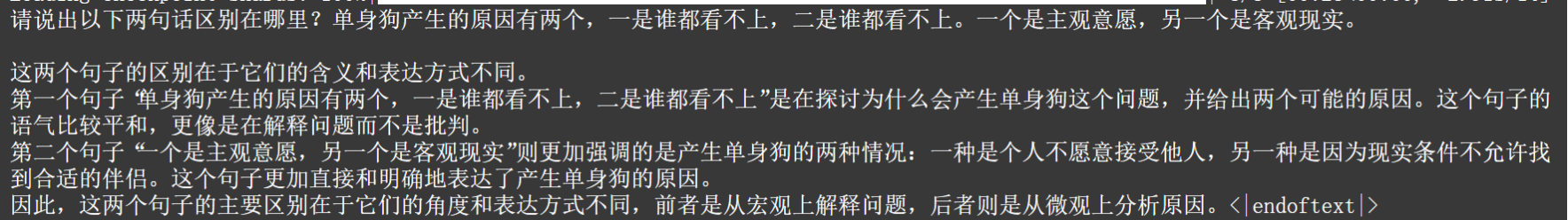
智谱ChatGLM-6B回答



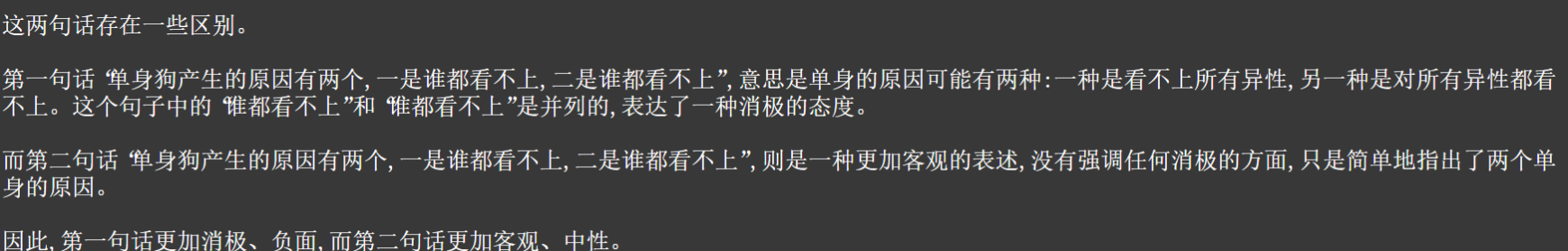
对于这个问题，正常理解是冬天尽量多穿，夏天尽量少穿。通义千问Qwen-7B-Chat的回答基本理解了两句话的差别，明白第一句是多穿，但没理解第二句应该少穿。智谱ChatGLM-6B完全理解该问题的意思，知道了冬天应多穿，夏天少穿。综合而言，智谱ChatGLM-6B在这个问题上表现更好，更能理解问题含义，与正常理解相近。

问题二.请说出以下两句话区别在哪里？单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上

通义千问Qwen-7B-Chat回答：



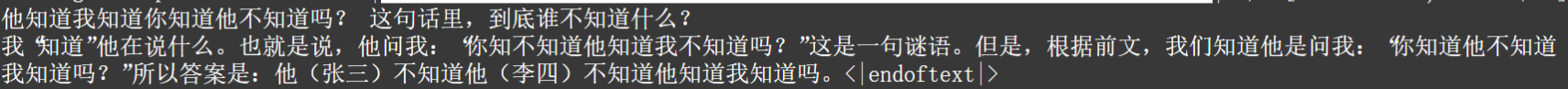
智谱ChatGLM-6B回答



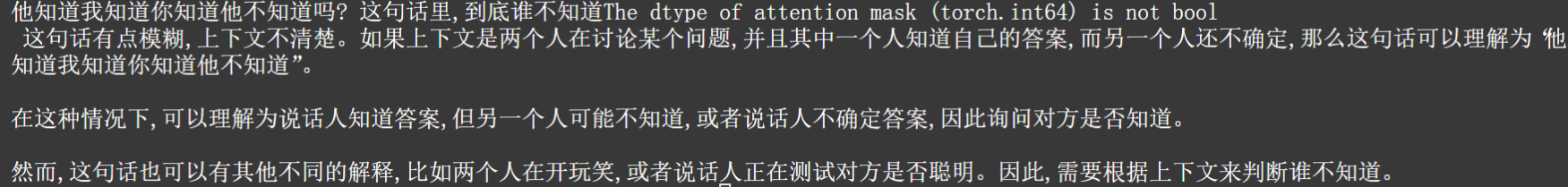
第二个问题，正常理解是单身的两种情况是自身条件太差导致没有人看上和眼光太高导致都不满意。通义千问Qwen-7B-Chat回答点出了主观意愿与客观现实的区别，但没有解释背后的原因。智谱ChatGLM-6B没有找到是哪两句，但还是理解了两者的区别。对于这个问题，智谱ChatGLM-6B回答的更好，具体解释了这句话的内涵。

问题三.他知道我知道你知道他不知道吗？这句话里，到底谁不知道；

通义千问Qwen-7B-Chat回答：



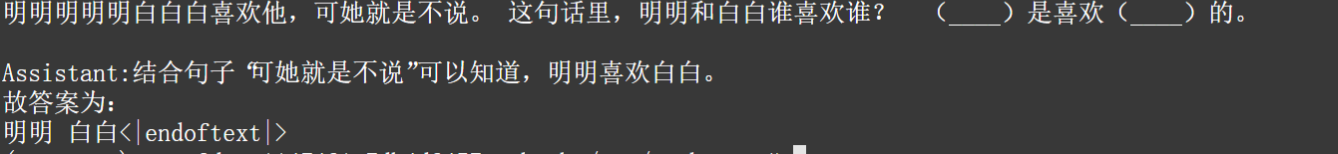
智谱ChatGLM-6B回答：



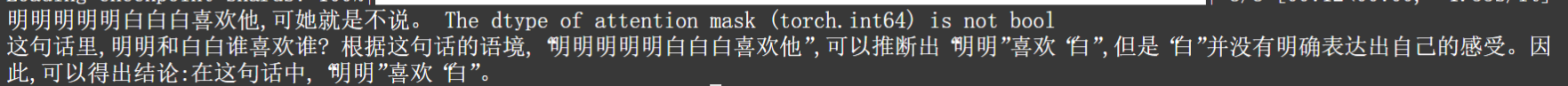
对于第三个问题，我们人类的正常理解应该是“他”不知道，这句话大致意思是，“你”知道“他不知道”这一事实，我又知道“你知道他不知道”这一事实，“我”想问“他”到底知不知道“我知道你知道他不知道”这一事实。通义千问Qwen-7B-Chat给出了接近正确理解的答案，也回答了问题。 智谱ChatGLM-6B并没有理解题目意思，只是重复题目意思以及进行无意义地猜测。总之，这道问题通义千问Qwen-7B-Chat回答地更好，给出了较为合理的答案。

问题四.明明明明明白白白喜欢他，可她就是不说。这句话里，明明和白白谁喜欢谁?

通义千问Qwen-7B-Chat回答：



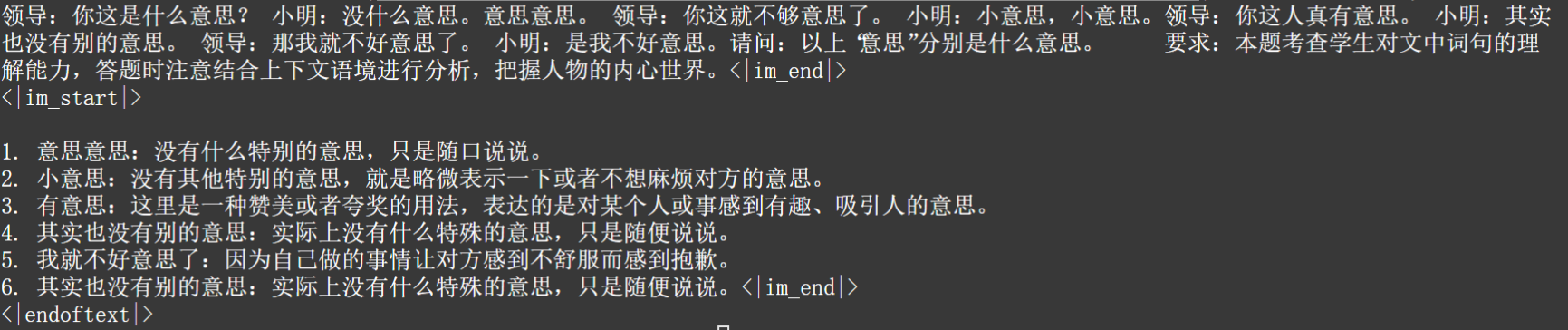
智谱ChatGLM-6B回答：



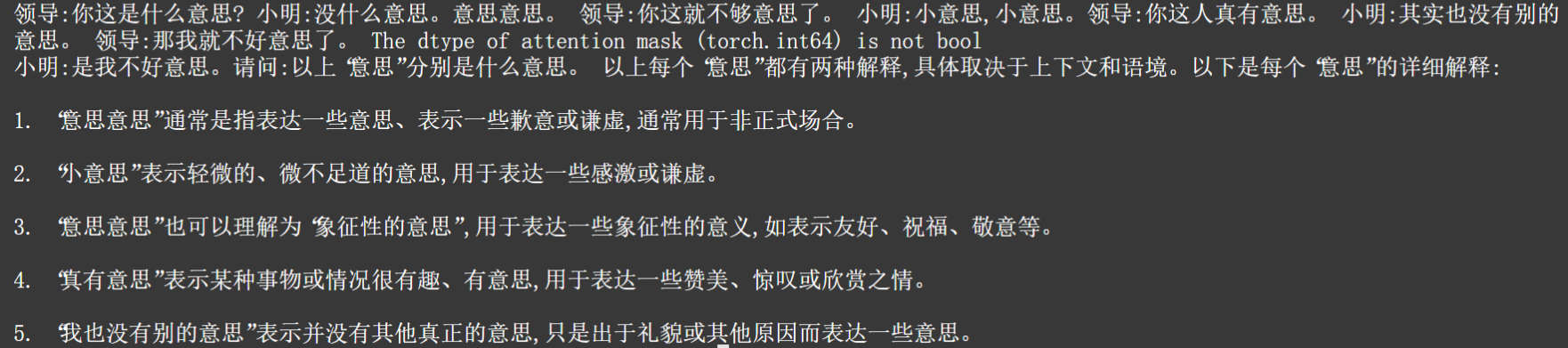
这个问题正常理解是明明知道白白喜欢他，但白白却没有说，并且明明是男生，白白是女生，他指明明，她指白白。通义千问Qwen-7B-Chat没有搞清她指代的是谁，所以误认为明明喜欢白白。智谱ChatGLM-6B也只根据第一句且没有正确断句，也认为明明喜欢白。综上所述，两者对这道题的回答均不理想，都只凭借单一语句判断，没有结合所有语句。

问题五.领导：你这是什么意思?小明：没什么意思。意思意思。领导：你这就不够意思了。小明：小意思，小意思。领导：你这人真有意思。小明：其实也没有别的意思。领导：那我就不好意思了。小明：是我不好意思。请问：以上“意思”分别是什么意思。

通义千问Qwen-7B-Chat回答：



智谱ChatGLM-6B回答：



对于第五个问题，一般理解：领导一开始问小明行为的意图(第一个意思)，小明回答没啥重要的(第二个意思)，只是想表达一下对领导的感激(第三个意思意思)，领导回答小明这行为可能不太好(第四个意思)，小明回答这只是小问题(比如送小礼物，第五个意思)，领导回复小明这人不错，知道送礼物 (第六个意思)，小明回答其实也只是想感谢一下领导，没有想叫领导帮忙照顾啥的(第七个意思)，领导回复那他这样白白收礼但不帮助小明他不太好意思，同时也表示接收礼物(第八个意思)，小明谦称说领导不要不好意思是我自己送礼不好意思。通义千问Qwen-7B-Chat基本上解释了各个意思的含义，也基本上正确，但是没有找到所有的“意思”。智谱ChatGLM-6B也基本上解释了各个“意思”的意义，也解释的比较正确。但两者都只是解释了意思的含义，没有结合语境进行分析。总而言之，这个问题两个模型回答的都很好。

**3.总结分析(此部分对应作业要求3：语言模型之间的横向对比分析)**

在本次项目中，我们对阿里云推出的通义千问-7B（Qwen-7B）和智谱AI与清华大学KEG实验室联合发布的ChatGLM-6B两款大语言模型进行了深入的对比测试。经过分析五个不同难度层次的问题，我们发现二者各有优势与待提升之处。

对于较为简单的问题，例如问题一和问题二，两款模型都能给出较为出色的回答。不过，ChatGLM3-6B在题目理解与回答能力上更胜一筹。在中等难度的问题，如问题三和问题四中，通义千问Qwen-7B-Chat虽然能够理解问题的意图，但给出的答案存在错误；而ChatGLM3-6B则似乎未能精准把握问题的关键，仅给出了诸如“无法做出判断”等不确定性的回答。至于难度较高的问题，以问题五为例，通义千问Qwen-7B-Chat能够给出较为准确的答案，但回答速度有待提升；ChatGLM3-6B的回答则显得较为笼统，未能充分结合实际语境，可能是由于它未能深入理解问题的核心所在。

综合本次项目测试结果，我们认为通义千问Qwen-7B-Chat在提升回答的准确性方面还有进一步改进的空间；而ChatGLM3-6B则需要在深化对问题的理解能力上加强训练。二者均为当前优秀的大语言模型，通过此次实验，我们从个人开发的角度提出了上述建议，以期为后续模型的优化提供参考。