大语言模型部署体验

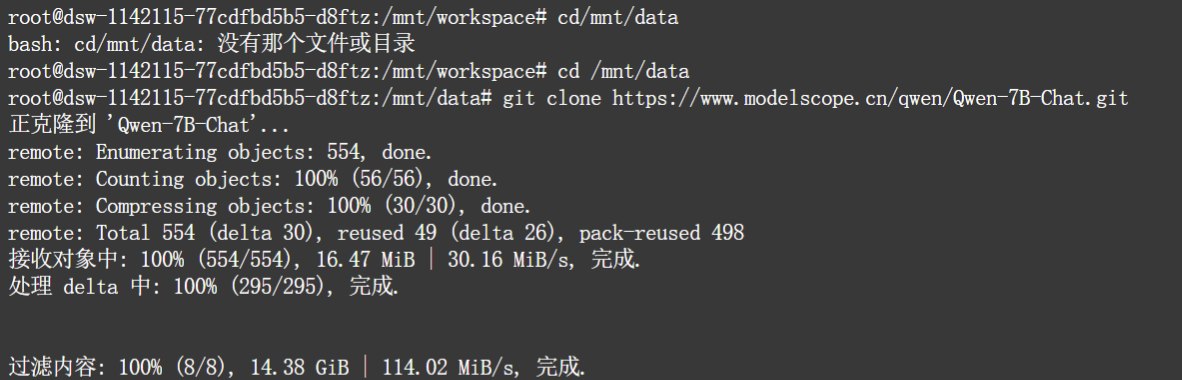
2353567 董钰洁

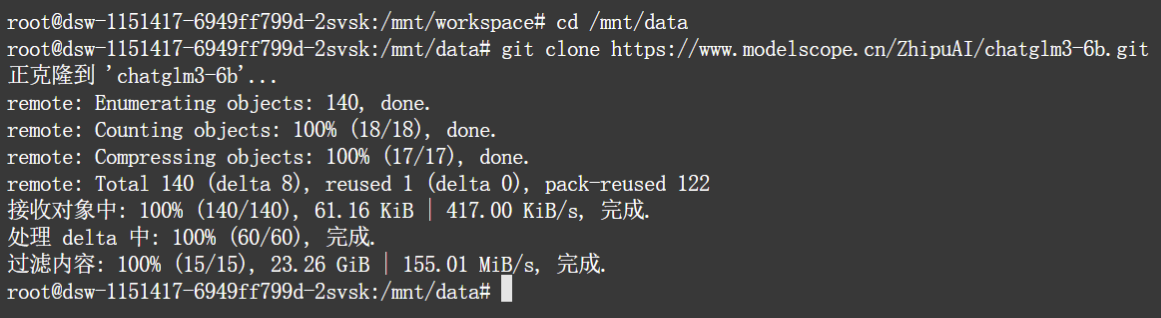
1. 项目公开访问链接：

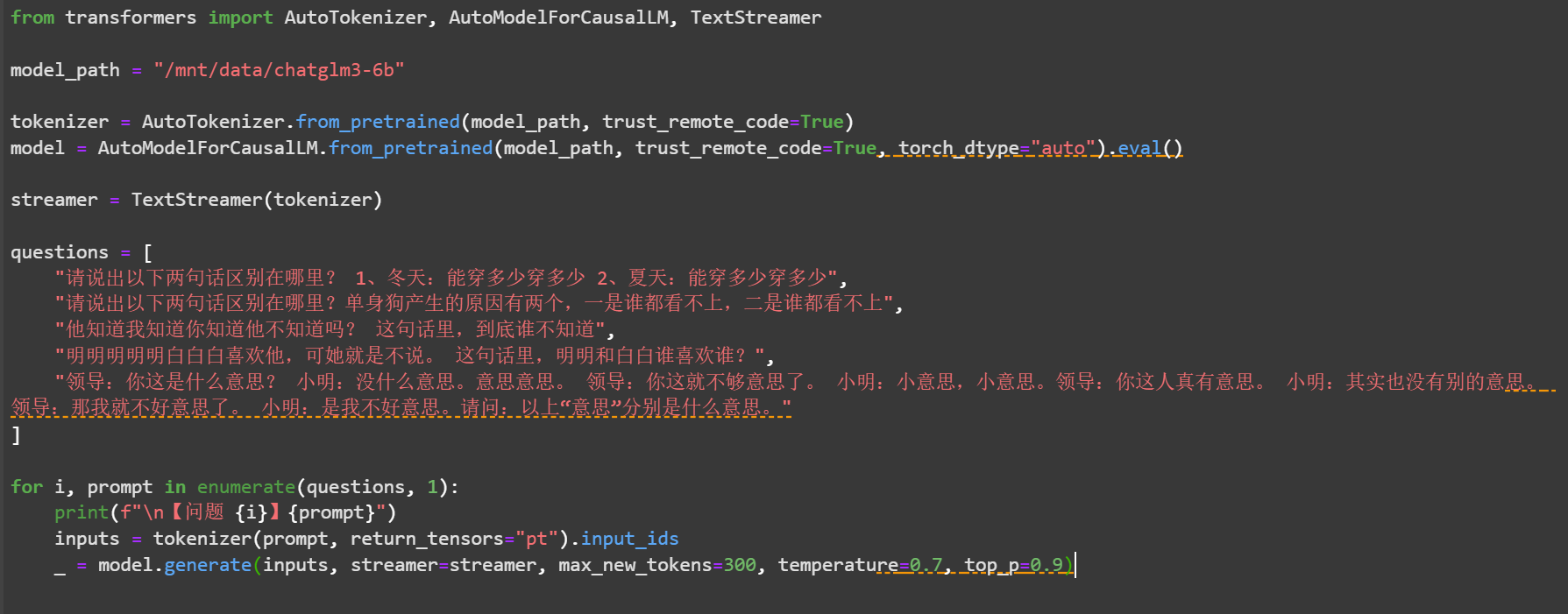
https://github.com/nobitayu/Large-Language-Model-Comparison

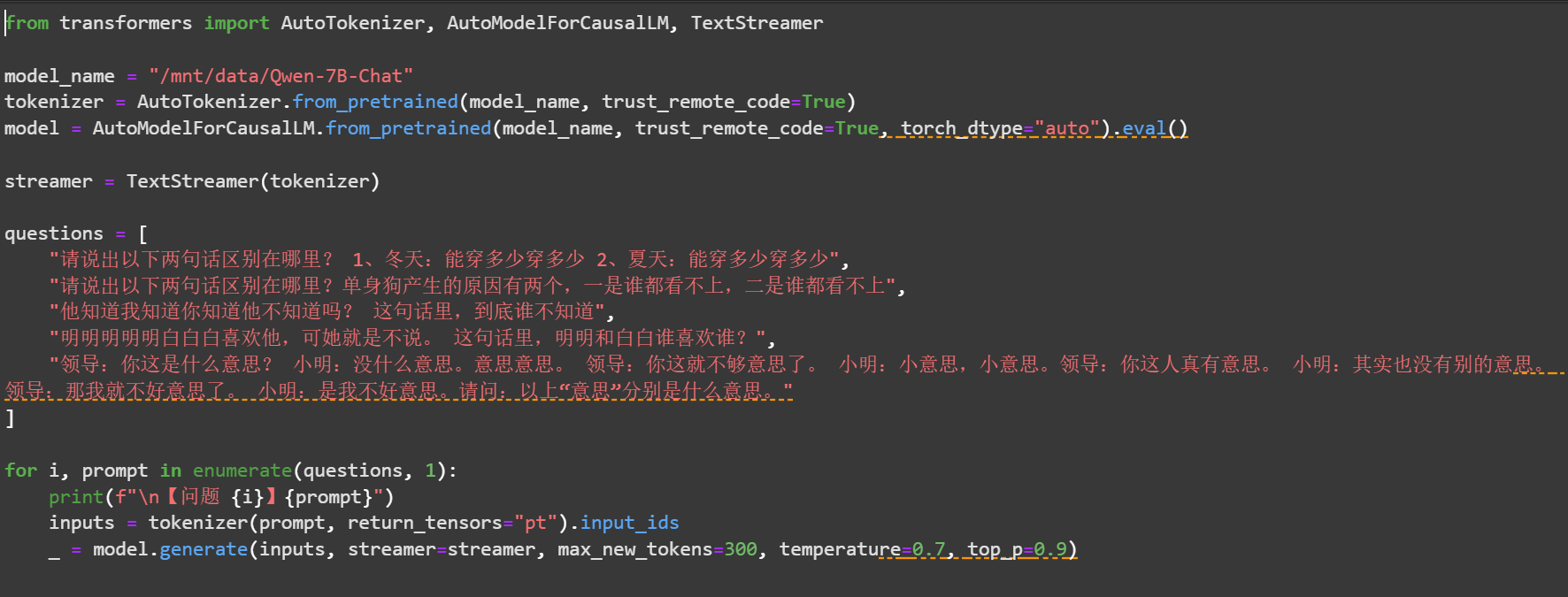
1. 相关截图

git clone 相关截图：



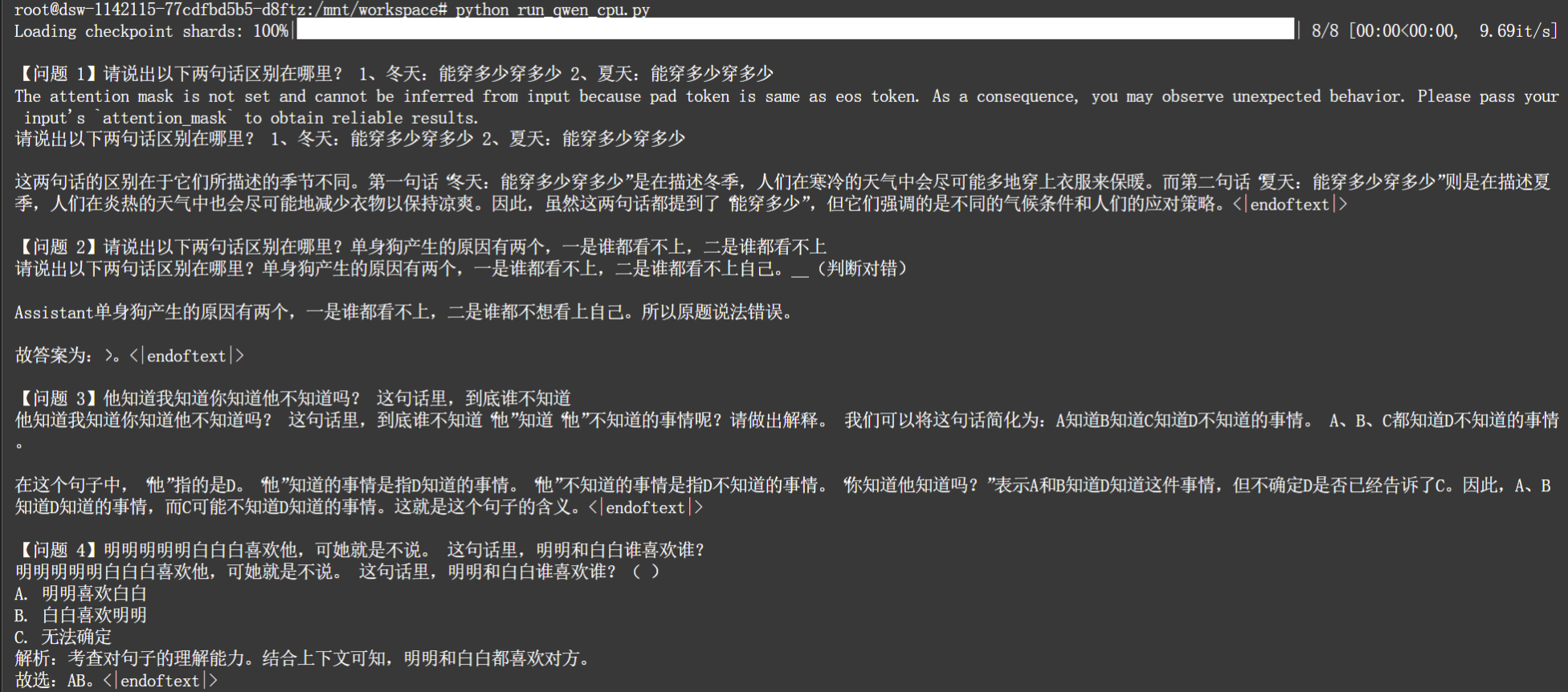


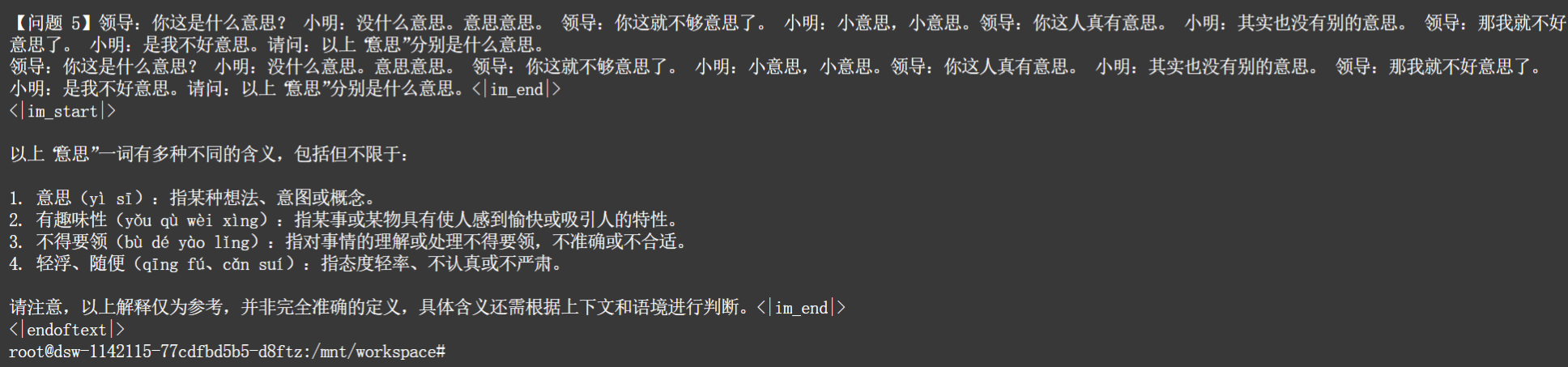




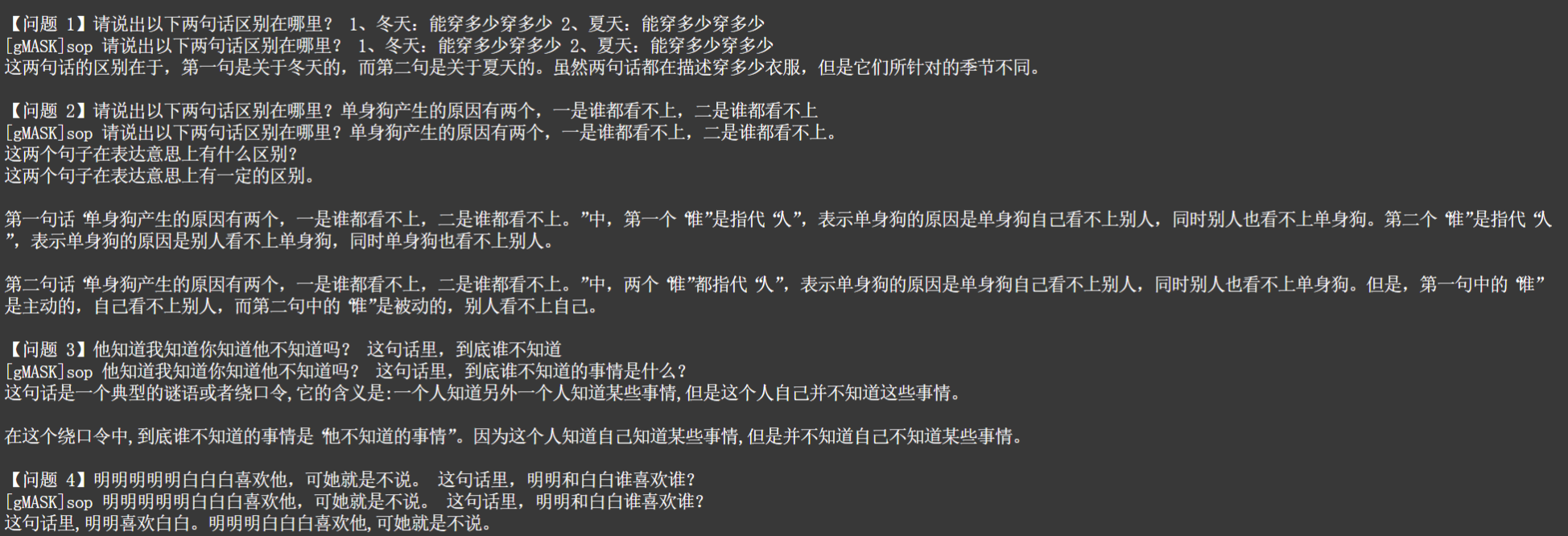
问答测试结果截图：

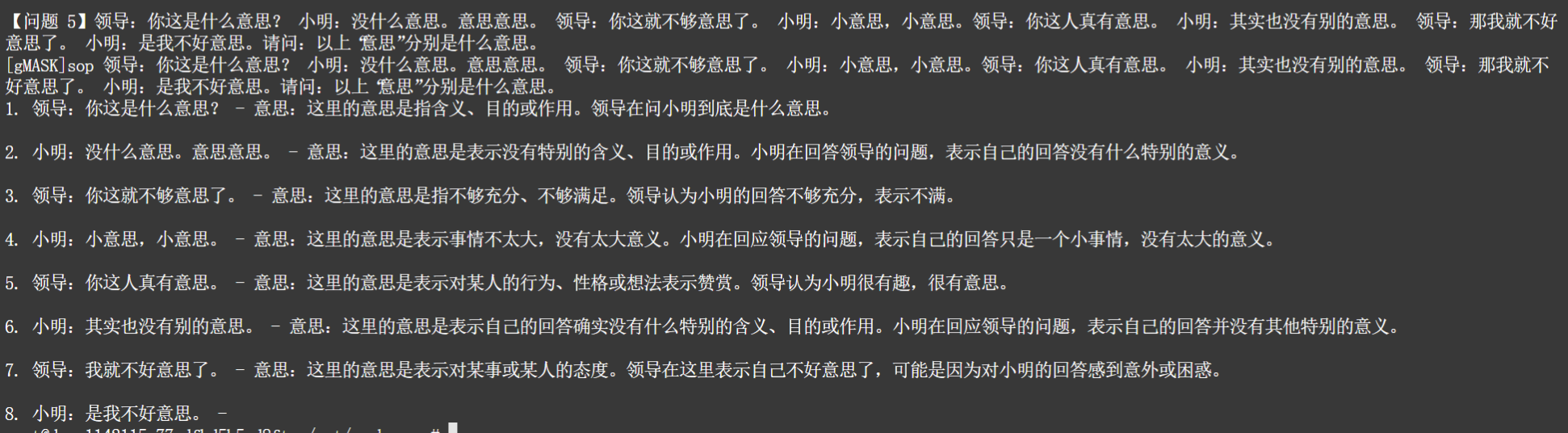
1. 通义千问 Qwen-7B-Chat





1. 智谱 ChatGLM3-6B





1. 大语言模型横向对比分析

测试问题：

1. 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

2. 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

3. 他知道我知道你知道他不知道吗？ 这句话里，到底谁不知道

4. 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？

5. 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。

对比结果：

1. 区别冬天和夏天的句子

通义千问：

优点： 能识别句子中“能穿多少穿多少”在不同季节下的含义，并结合冬天保暖、夏天清凉进行解释，具备一定的语境分析能力。

缺点： 表述较为重复，语言略显啰嗦，逻辑有待简化。

智谱 ChatGLM：

优点： 准确识别出句子的季节差异，表达清晰直接，能够基本说明问题。

缺点： 分析较浅，没有进一步探讨“相同语句不同语境”背后的反差或幽默感。

2. 区别单身狗产生原因的句子

通义千问：

优点： 正确指出句子的逻辑错误，并给出正确版本，能从“谁都看不上”和“谁都看不上自己”的主客体角度进行辨析，较有洞察力。

缺点： 对句中“谁”的指代关系理解略显含混，部分地方逻辑表达不清。

智谱 ChatGLM：

优点： 较详细分析了两个“谁”的不同语义角色，尝试区分主动与被动语气，说明句意差异。

缺点： 表述不够精炼，有一定语法混乱，容易让人理解困难。

3. 谁不知道的问题

通义千问：

优点： 分析句子结构，从“知道链”中抽象出逻辑层级（A知道B知道C知道D不知道），给出相对清晰的解构路径。

缺点： 概念化偏强，实际对“谁不知道”的判断未明确指出，略显绕。

智谱 ChatGLM：

优点： 抓住“他不知道”是关键信息，尝试解释“他”并不知道别人已经知道的事，整体逻辑有一定条理。

缺点： 逻辑表达含糊，分析偏表面，没有清晰定位“谁不知道”。

4. 明明和白白谁喜欢谁

通义千问：

优点： 基于结构分析得出“双向喜欢”的结论，结合上下文意图判断，体现一定语感。

缺点： 判断“AB都喜欢”的依据略显主观，未结合句中“她不说”的主语做具体解释。

智谱 ChatGLM：

优点： 简洁指出“明明喜欢白白”，结论明确，回答直接。

缺点： 逻辑链条缺失，无法说明为什么是“明明喜欢白白”，缺乏分析过程。

5. 不同“意思”的含义

通义千问：

优点： 能识别语义重复中蕴含的幽默语言现象，展现出多种“意思”在语境中的用法，富有语言趣味性。

缺点： 个别解释略显模糊，有些“意思”未结合上下文准确分析。

智谱 ChatGLM：

优点： 逐条说明每句话中的“意思”所指的具体含义，结构清晰，逻辑明白。

缺点： 偏重逐字解释，未展现出语言的幽默性或语义的双关性。

综合总结

通义千问 Qwen-7B-Chat 擅长逻辑结构拆解和较深入的语义分析，尤其在多重句或绕口令类型的问题中体现出一定理解力，语言有一定人情味，但逻辑条理偶尔不清晰，表达稍显拖沓。

智谱 ChatGLM3-6B 回答风格简洁直观，适合基础理解任务，句意识别基本准确，但分析不够深入，难以处理含有复杂逻辑的语言表达。