**LLM部署报告**

# 部署的模型

1. [智谱ChatGLM3-6B](https://www.modelscope.cn/models/ZhipuAI/chatglm3-6b/summary)
2. [通义千问Qwen-7B-Chat](https://www.modelscope.cn/models/qwen/Qwen-7B-Chat/summary)

# 部署完成截图

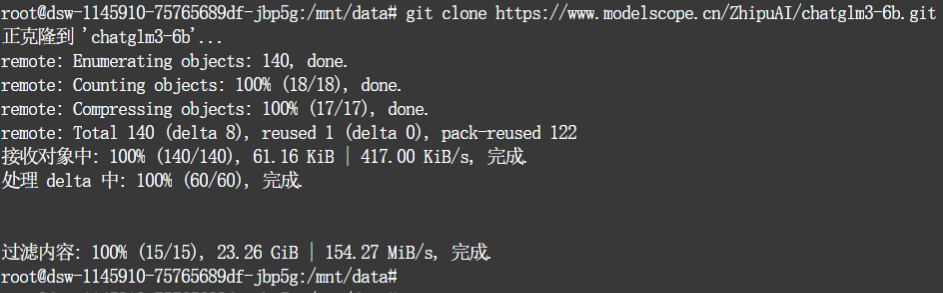


图 1 智谱ChatGLM3-6B git clone截图



图 2 通义千问Qwen-7B-Chat git clone截图

# 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。

# 测试结果

## 智谱ChatGLM3-6B测试结果

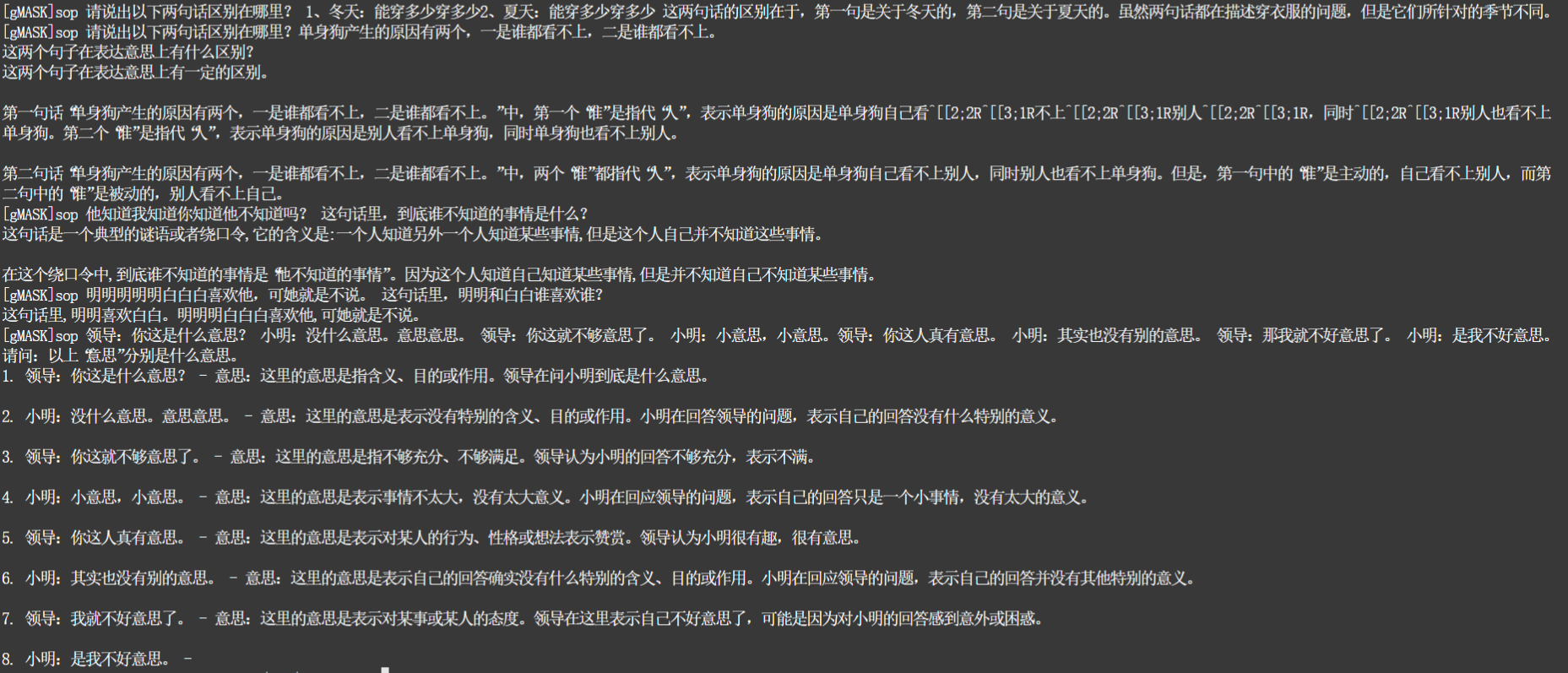


图 3 智谱ChatGLM3-6B问题测试结果

## 通义千问Qwen-7B-Chat测试结果

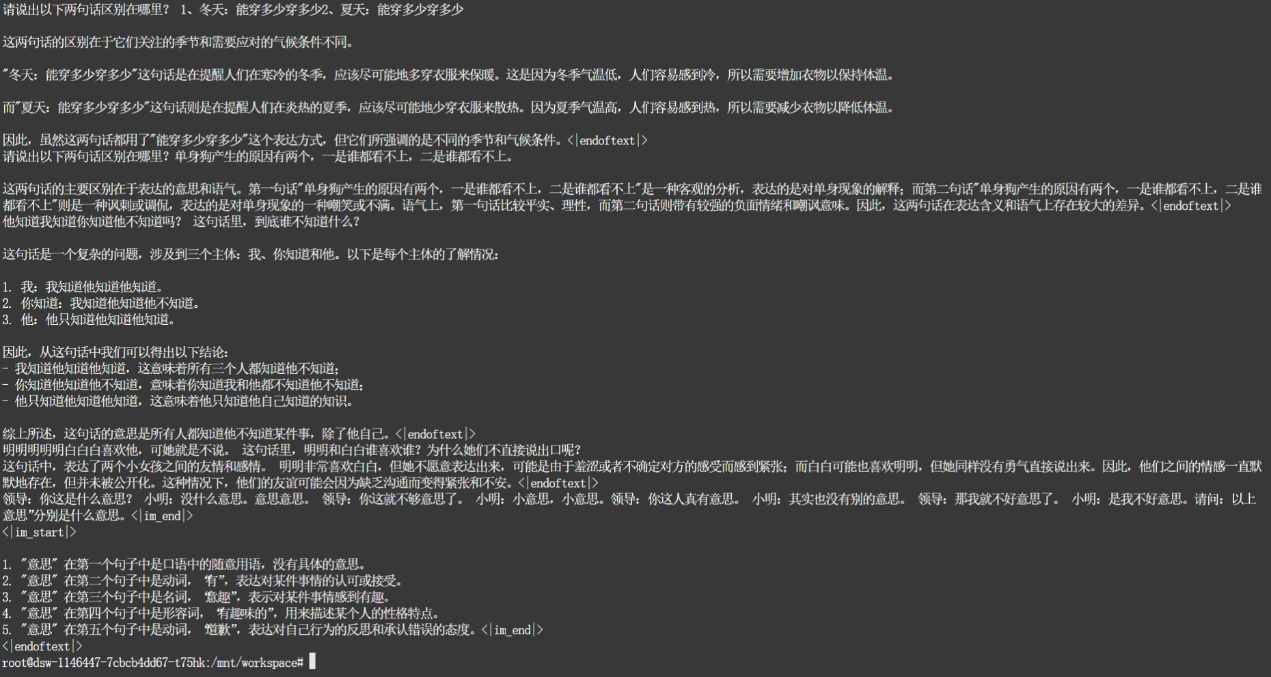


图 4 通义千问Qwen-7B-Chat问题测试结果

# 基于测试结果的LLM横向对比分析

## 冬天和夏天“能穿多少穿多少”的问题

### 智谱ChatGLM3-6B

·优点：能够发现虽然都描述的是穿衣服，但是针对的季节不同的问题

·缺点：缺乏对于两个情境下“能穿多少穿多少”实际语义的详细解释，回答分析不够深入

### 通义千问Qwen-7B-Chat

·优点：能够指出两句话的区别在于气候条件不同，并给出可能的背景解释

·缺点：回答冗长，语言不够精炼，存在重复表达

## 区分单身狗两种“谁都看不上”的问题

### 智谱ChatGLM3-6B

·优点：能够正确指出两种“谁都看不上”在语义上的区别

·缺点：回答的语言组织和表达比较混乱，同时仅仅是就事论事，只从句子语法结构角度展开分析，没有更进一步的分析。

### 通义千问Qwen-7B-Chat

·优点：从语气和表达的态度上指出了两者的不同

·缺点：回答相对冗长，没能简明地指出两种“谁都看不上”的语义区别

## “谁不知道”的问题

### 智谱ChatGLM3-6B

·优点：能够识别出问题的故意混乱化与模糊性

·缺点：没能给出正确的回答，同时也错误识别了实际的问题（是“谁不知道”而非“谁不知道的事情是什么”），语言组织同样比较混乱

### 通义千问Qwen-7B-Chat

·优点：分步整合已知信息并进行分析，进一步得出结论

·缺点：分析过程中对问题的理解出现了较多差错，另外回答冗余

## 明明和白白谁喜欢谁的问题

### 智谱ChatGLM3-6B

·优点：正确识别出问题并开门见山给出回答

·缺点：给出了完全错误的回答，没能理解原句实际表达的逻辑与含义

### 通义千问Qwen-7B-Chat

·优点：对问题及问题描述的现象出现的可能原因进行了分析

·缺点：自设了问题当中并不存在的语境，同时并未给出明确的“谁喜欢谁”的结论，没能完全理解原问题

## 区分不同“意思”含义的问题

### 智谱ChatGLM3-6B

·优点：按点分析了不同“意思”的含义并给出了相应解释

·缺点：存在缺漏和错误，对不同“意思”的理解不完全正确，同时，对同一句中有多个不同含义的“意思”的情况，没能识别出所有不同含义的“意思”（如没什么意思，意思意思中）。另外，解释比较死板。

### 通义千问Qwen-7B-Chat

·优点：按点分析了不同句子中“意思”的词性和含义

·缺点：存在缺漏和错误，仅部分分析正确，另外，没有结合原句具体分析

## 总结

由于token限制在300，各模型的表现可能不够理想。

就目前获得的结果而言，在整体表现方面，智谱 ChatGLM3-6B与通义千问 Qwen-7B-Chat各有优劣。

ChatGLM3-6B 在概念识别上表现中等，通常能够识别出关键词和语境的基本结构，但在语言组织方面逻辑性较弱，经常出现表达混乱或跳跃的问题。它的解释往往停留在表层，偏向于句法和结构分析，对语义深度的挖掘有限。其回答通常较为简洁，适合快速理解任务，但也因此容易产生误解。在推理能力方面，它虽然有一定的尝试和探索，但逻辑链条容易断裂，容易出现结论性错误。

相比之下，Qwen-7B-Chat 在识别概念、理解语境方面相对更强一些。它的语言表达更有逻辑，能够将语义链条较完整地表达出来，虽然有时会显得冗长。它在解释过程中常常尝试从语气、表达意图等语用角度切入，具有一定的分析深度。不过，它的回答有时为了详尽而牺牲了精炼度。推理上，它具备结构化思维能力，能条分缕析地展开分析，但判断准确性仍有提升空间。

在具体使用场景上，两个模型也展现出不同的适用程度。

如果任务涉及语言幽默、绕口令、语义层次解构等方面，推荐使用 Qwen-7B-Chat。它较擅长处理复杂、嵌套的语义关系，并能结合语境进行多层解释。

如果任务主要是句式结构分析、语法逻辑训练或语言表达基础教学，则 ChatGLM3-6B 更为合适。它的表达较为直接，逻辑结构清晰，适合更基础的语言训练。

在需要教学或讲解时，尤其是希望提供背景补充、情境解释、引导式说明的场景下，Qwen-7B-Chat 的拓展能力更占优势。它可以围绕核心内容延伸出更多辅助信息，增强教学效果。

而在需要快速响应、简洁明了的语义任务中，例如初级语言理解或日常问答中，ChatGLM3-6B 的短平快风格更适合这种轻量级应用。

总体来看，对于本次实验测试的这两个模型，Qwen 更注重语义深度与语境完整性，适合分析型场景；而 ChatGLM 更适合结构清晰、目标明确的任务环境。两者各有专长，可按需互补。

# 项目可公开访问链接