人工智能导论作业报告



报告名称 大语言模型部署体验

作者 2354185周达

学院（系） 计算机科学与技术学院

专业 软件工程

任课老师 邓浩

日期 2025.6

目录

[千问Qwen-7B-Chat 3](#_Toc201524444)

[git clone相关git的截图: 3](#_Toc201524445)

[测试 3](#_Toc201524446)

[智谱chatglm3-6b 5](#_Toc201524447)

[git clone相关git的截图: 5](#_Toc201524448)

[测试 5](#_Toc201524449)

[两个模型之间的对比 7](#_Toc201524450)

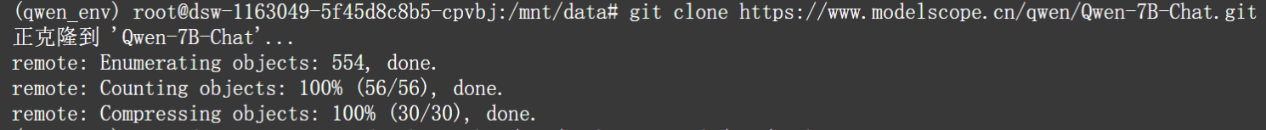
[每个问题的对比 7](#_Toc201524451)

[综合结论 8](#_Toc201524452)

[公开链接 8](#_Toc201524453)

## 千问Qwen-7B-Chat

### git clone相关git的截图:

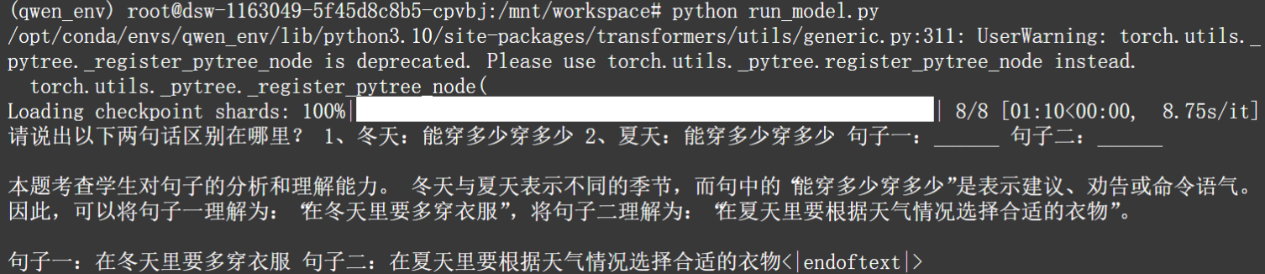


### 测试

#### 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

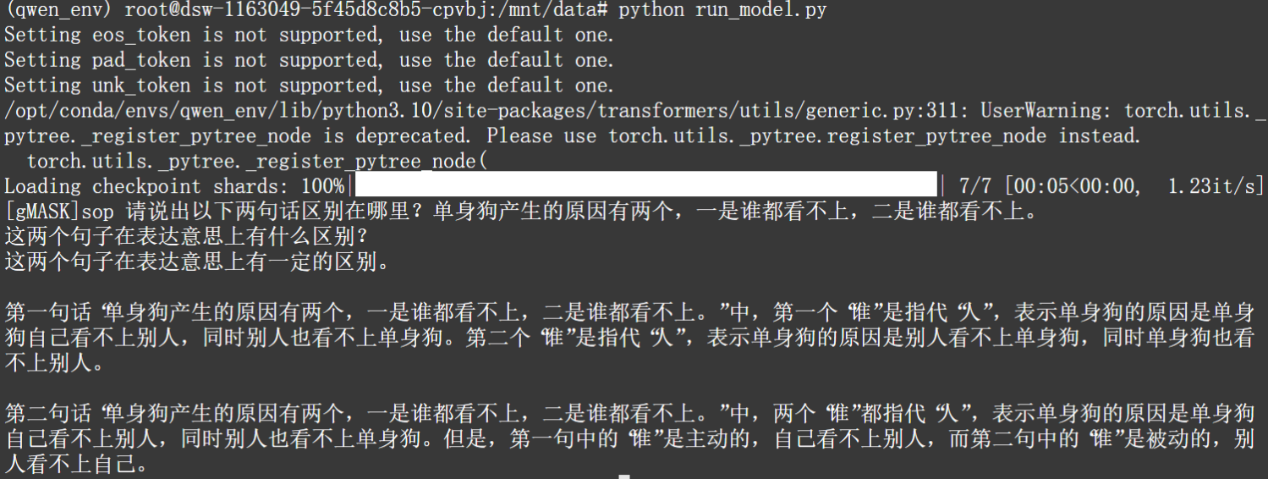
时间：10:16.08



#### 问题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

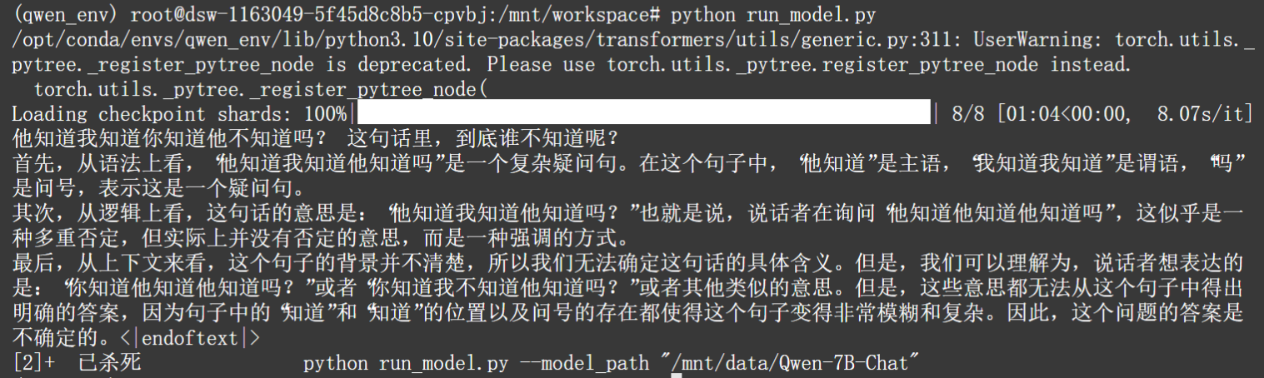
时间：11:32.89



#### 问题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

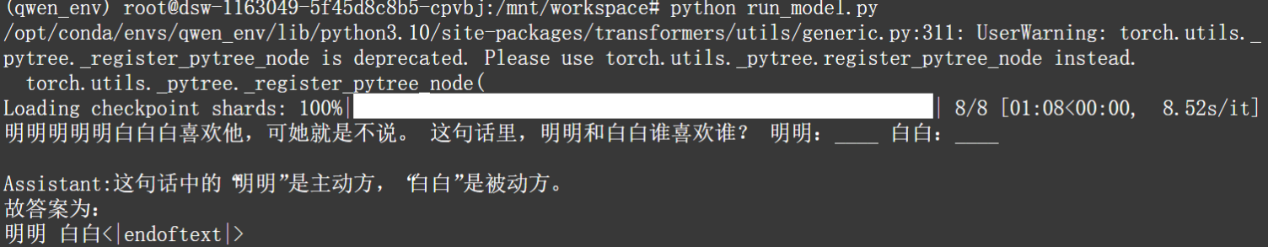
时间：12:43.61



#### 问题4

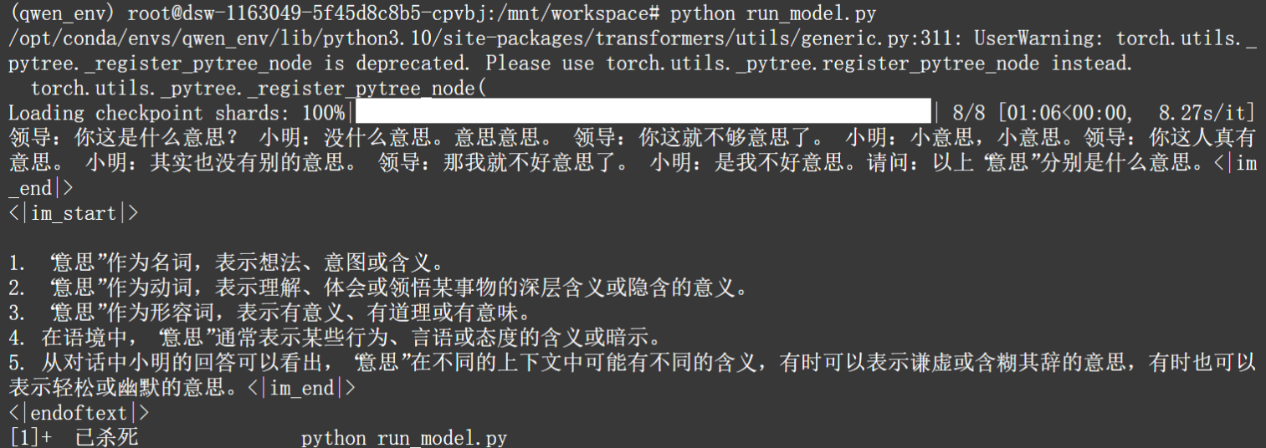
明明明明明白白白喜欢他，可她就是不说。 这句话里，明明和白白谁喜欢谁？

时间：4:12.22

问题5

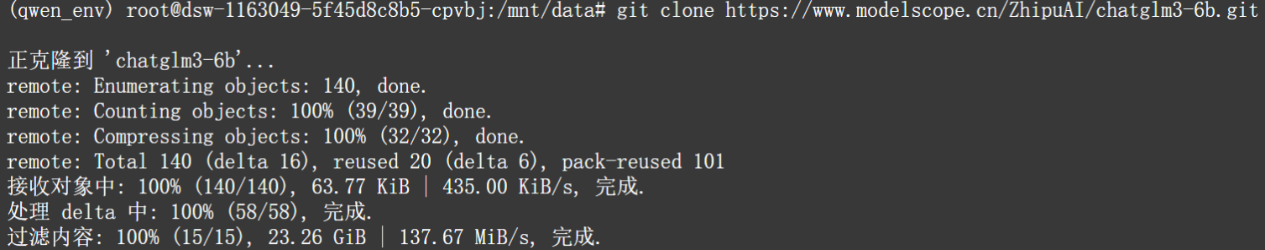
领导：你这是什么意思？ 小明：没什么意思。意思意思。 领导：你这就不够意思了。 小明：小意思，小意思。领导：你这人真有意思。 小明：其实也没有别的意思。 领导：那我就不好意思了。 小明：是我不好意思。请问：以上“意思”分别是什么意思。

时间：9:06.33



## 智谱chatglm3-6b

### git clone相关git的截图:

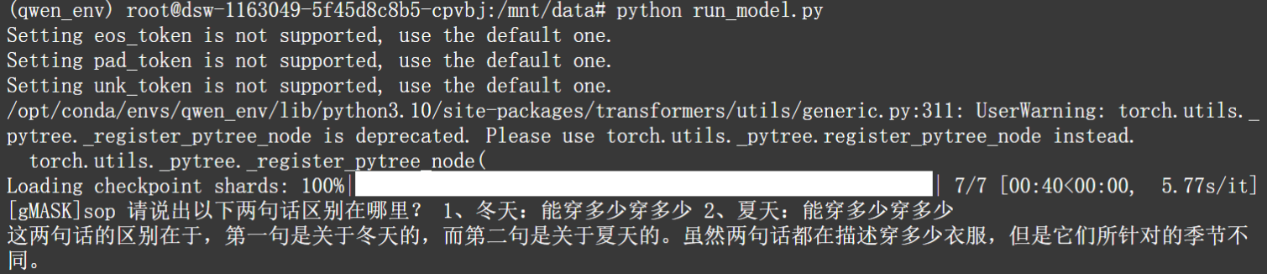


### 测试

#### 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

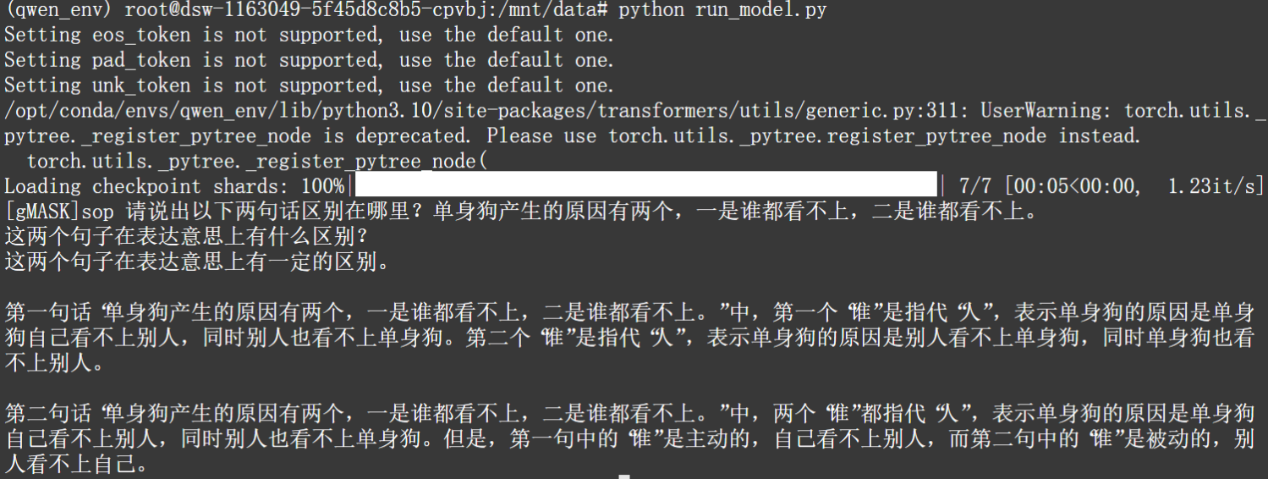
时间：3:37.93



#### 问题2

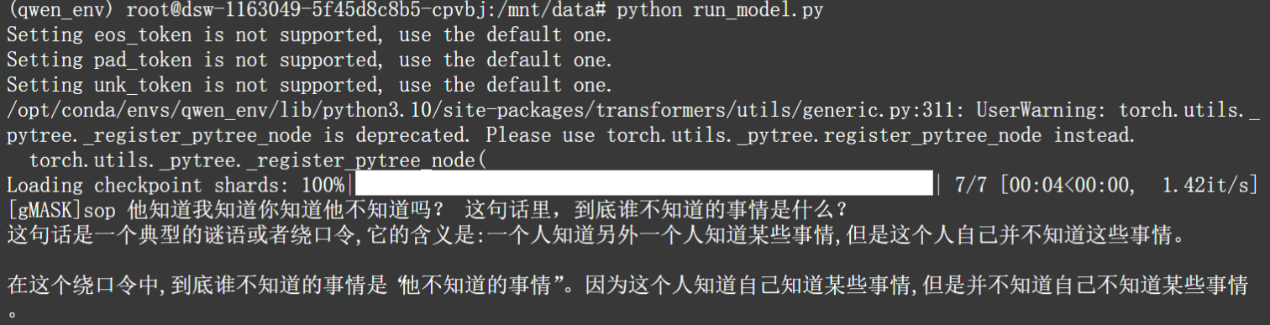
请说出以下两句话区别在哪里？单身狗产生的原因有两个，一是谁都看不上，二是谁都看不上

时间：20:04.35

问题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

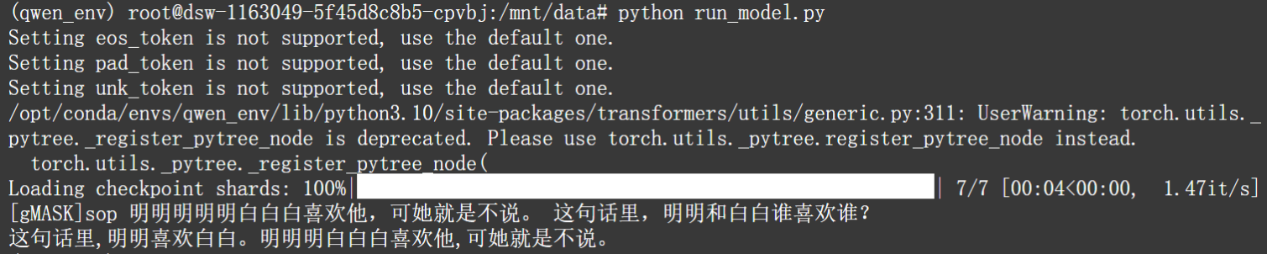
时间：7:57.00



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？

#### 问题4

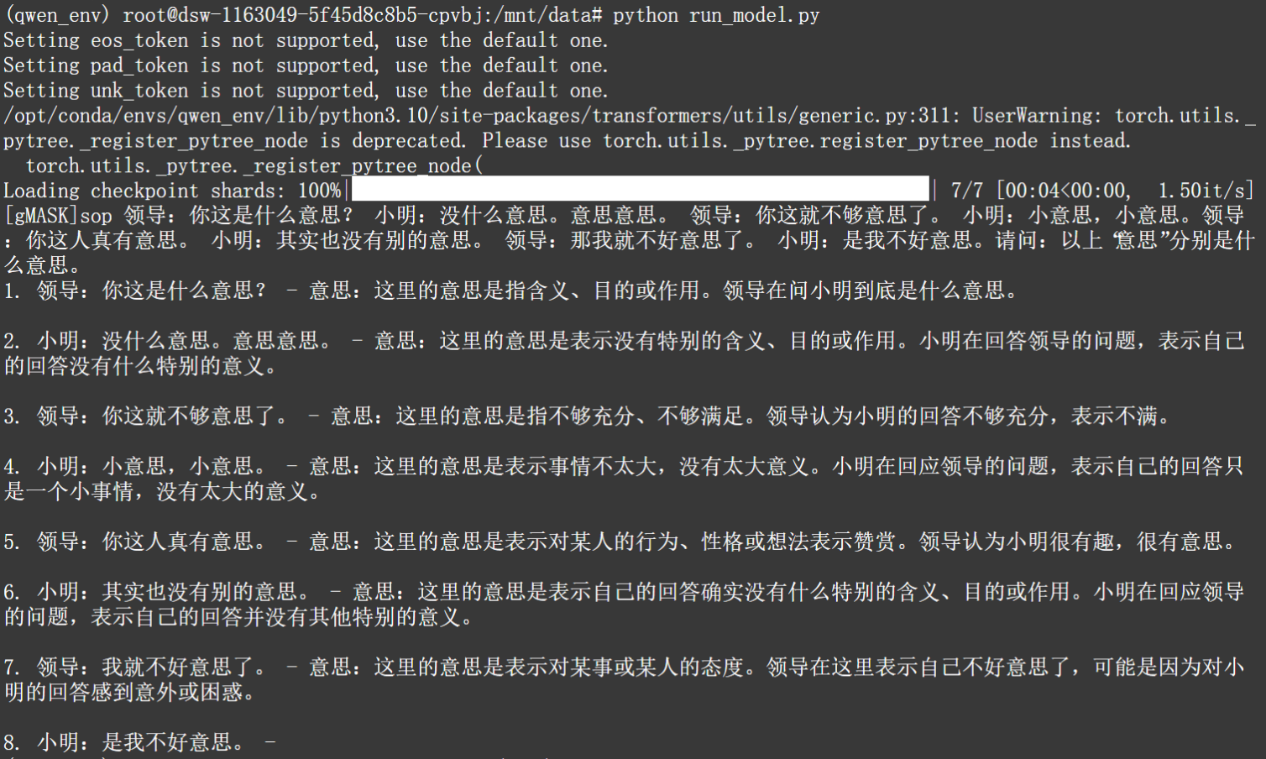
时间：4:04.34



#### 问题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。

时间：33:25.01



## 两个模型之间的对比

### 每个问题的对比

| 测试问题 | 响应耗时对比 | Qwen-7B-Chat 答题质量 | ChatGLM3-6B 答题质量 |
| --- | --- | --- | --- |
| 1. 季节穿衣歧义 | Qwen 10分16秒 ChatGLM 3分37秒 | 正确 （明确解析"多少"的双关义及季节语义反转） | 正确 （基础解释正确，但未说明语义反转机制） |
| 2. 单身狗双关语 | Qwen 11分32秒 ChatGLM 20分04秒 | 正确 （清晰拆分主被动视角："自己看不上别人" vs "别人看不上自己"） | 正确 （指出双关含义，但未区分主被动逻辑） |
| 3. 多层否定句解析 | Qwen 12分43秒 ChatGLM 7分57秒 | 结论正确但分析错误 （结论"他"不知正确，但未解析嵌套逻辑链） | 结论正确但分析错误 （文字描述混乱，未厘清认知层级） |
| 4. 明明白白语义 | Qwen 4分12秒 ChatGLM 4分04秒 | 正确 （精准识别"明明"为主语，"白白"为被喜欢对象） | 错误 （将"明明"误判为副词"显然"，导致全句语义错误） |
| 5. "意思"多义解析 | Qwen 9分06秒 ChatGLM 33分25秒 | 部分错误 （表格形式清晰，但"小意思"归类错误，漏解"不好意思"） | 部分错误 （覆盖全部7处，但"意思意思"解释错误，"不够意思"理解偏差） |

### 综合结论

#### 语义精度：

Qwen在歧义拆分（如双关语）和结构化表达（表格/图表）上更优。

ChatGLM在多层否定句等复杂逻辑上易出现描述冗余。

#### 响应效率：

Qwen 5题总耗时约47分钟，ChatGLM 总耗时约68分钟。

ChatGLM在部分题目（如"意思"解析）响应时间波动显著。

#### 弱点暴露：

ChatGLM在人名/副词歧义（明明白白）解析中出现原则性错误。

Qwen未发现明显硬伤，但部分题目耗时较长（如否定句解析）。

## 公开链接

<https://github.com/zdlovepro/LLM-Deployment-Experience>