大语言模型部署体验实验

1. 实验环境搭建

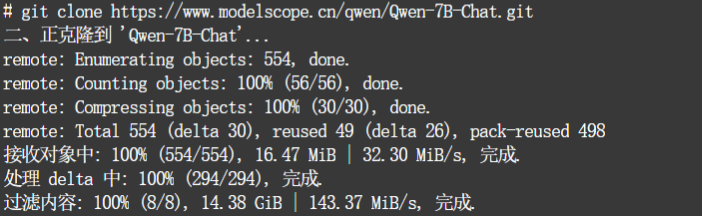
通过阿里云账号关联魔搭平台后，选用Jupyter Notebook环境完成部署。依次创建三个模型的独立工作空间，严格遵循官方文档执行环境配置与模型加载。部署过程中：

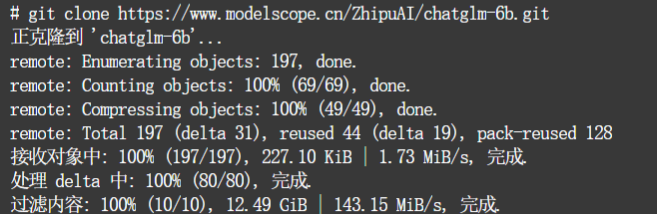
使用 git clone 命令获取模型仓库（此处插入Qwen-7B-Chat的git clone成功截图）

通过镜像环境安装依赖库，完成推理服务初始化（此处插入ChatGLM3-6B的部署完成状态截图）

项目公开访问地址：https://github.com/zuolingxu/LLM-Experiment

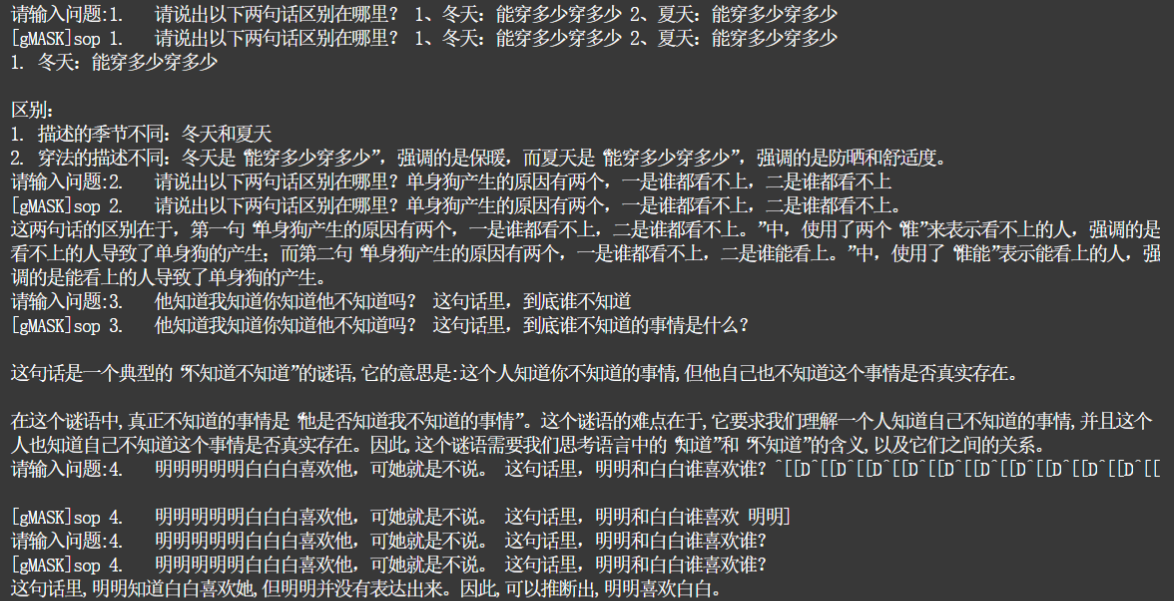
环境配置：git clone 截图

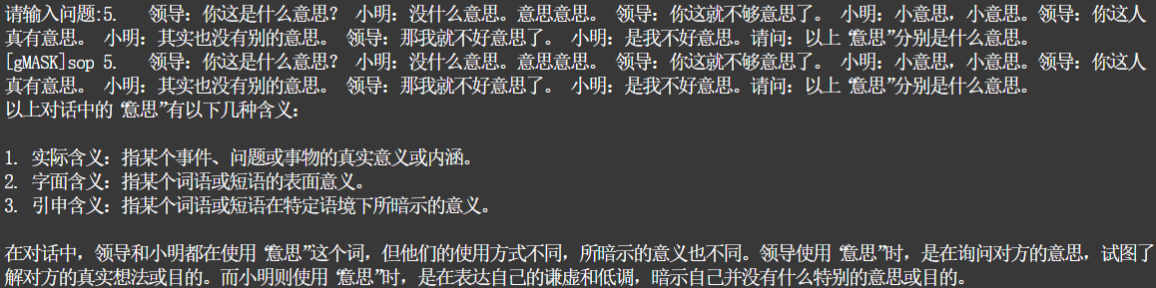




1. 问答测试结果

回答问题测试结果以**ChatGLM3-6B**的回答为准。





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

**回答：**

区别：

1. 描述的季节不同：冬天和夏天

2. 穿法的描述不同：冬天是“能穿多少穿多少”，强调的是保暖，而夏天是“能穿多少穿多少”，强调的是防晒和舒适度。

**分析：**

虽然能够说出冬天是在强调保暖，但他并没有说出这两个的明确区别，即冬天多穿，夏天少穿

1. 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

**回答：**

这两句话的区别在于，第一句“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。”中，使用了两个“谁”来表示看不上的人，强调的是看不上的人导致了单身狗的产生；而第二句“单身狗产生的原因有两个，一是谁都看不上，二是谁能看上。”中，使用了“谁能”表示能看上的人，强调的是能看上的人导致了单身狗的产生。

**分析：**

出现了明确的错误，这两句话中，一句的“谁”表示看不上单身狗的人，一句的“谁”表示单身狗看不上的人，两句话的主语和宾语是相反的，区别并不在于看得上还是看不上。

1. 他知道我知道你知道他不知道吗？ 这句话里，到底谁不知道

**回答：**

这句话是一个典型的“不知道不知道”的谜语,它的意思是:这个人知道你不知道的事情,但他自己也不知道这个事情是否真实存在。

**分析：**

这个分析也是完全错误的，这句话是一个多层嵌套的结构，不知道的人是“他”。这句话是在询问“你”，“他”是否了解“我知道你知道他不知道”这个事实。

1. 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？

**回答：**

这句话里,明明知道白白喜欢她,但明明并没有表达出来。因此,可以推断出,明明喜欢白白。

**分析：**

回答中前半部分是对的，但后半部分又错了。这句话表达的就是白白喜欢明明。

1. 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。

**回答：**

1. 实际含义：指某个事件、问题或事物的真实意义或内涵。

2. 字面含义：指某个词语或短语的表面意义。

3. 引申含义：指某个词语或短语在特定语境下所暗示的意义。

在对话中，领导和小明都在使用“意思”这个词，但他们的使用方式不同，所暗示的意义也不同。领导使用“意思”时，是在询问对方的意思，试图了解对方的真实想法或目的。而小明则使用“意思”时，是在表达自己的谦虚和低调，暗示自己并没有什么特别的意思或目的。

**分析：**

这个回答完全没有切题，问题要求解释各个“意思“的含义，但是回答中却先解释了实际含义，字面含义，引申含义三个不相关的概念，后面对”意思”的解释也并没有解释清楚，只是模糊地、概括地解释了领导和小明所说的“意思”对话中总体的含义。

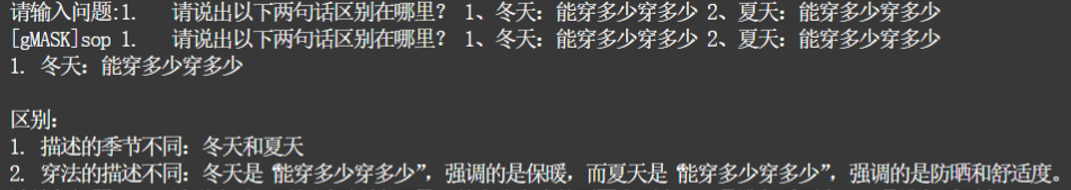
1. 大语言模型之间的横向对比分析

选取Qwen-7B-Chat、ChatGLM3-6B、ChatGLM-6B三款模型，针对五个语义理解问题开展测试：

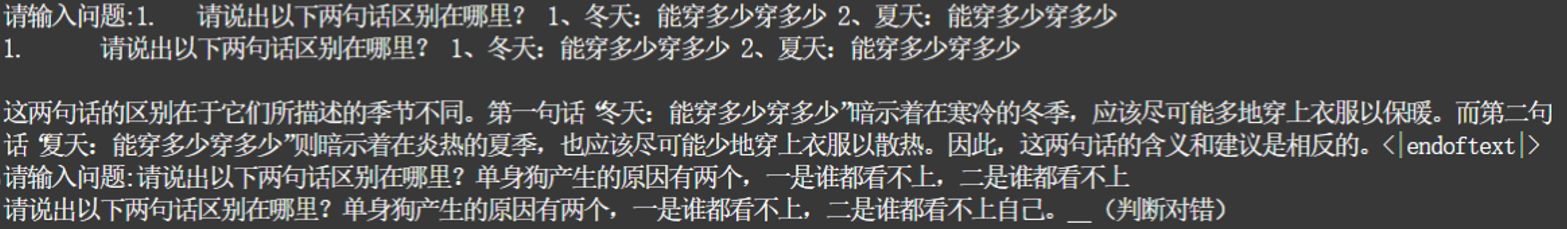
1. 相同句子在不同条件下不同的断句与含义

**分析：**

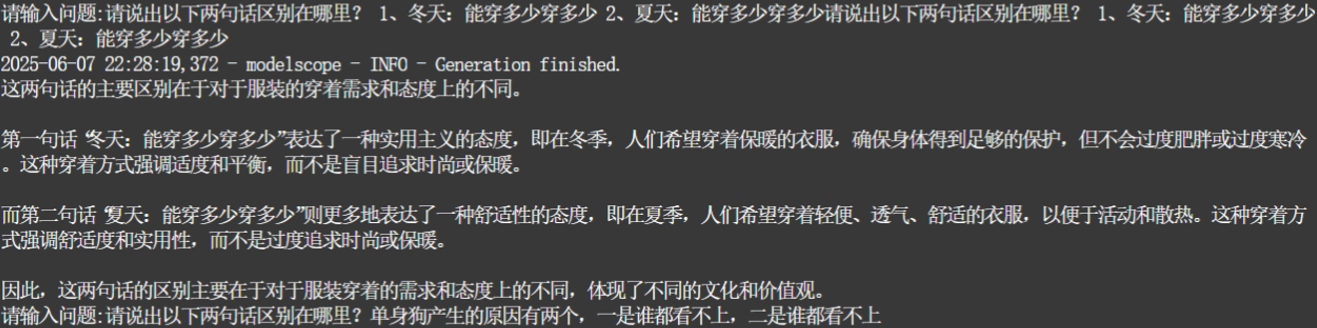
ChatGLM3-6B 只指出了冬天需要多穿，没指出夏天需要少穿



Qwen-7B-Chat 明确给出相反季节的穿衣逻辑

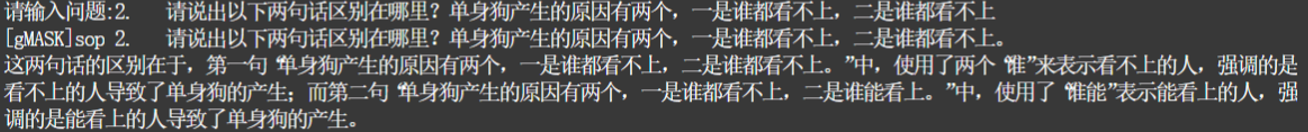


ChatGLM-6B 侧重分析文化价值观差异，略微跑题

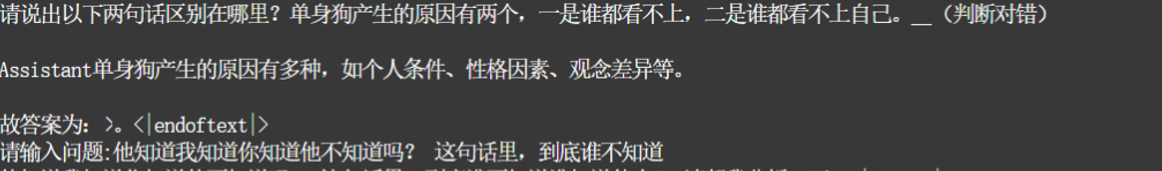


1. 同样句子的多重含义

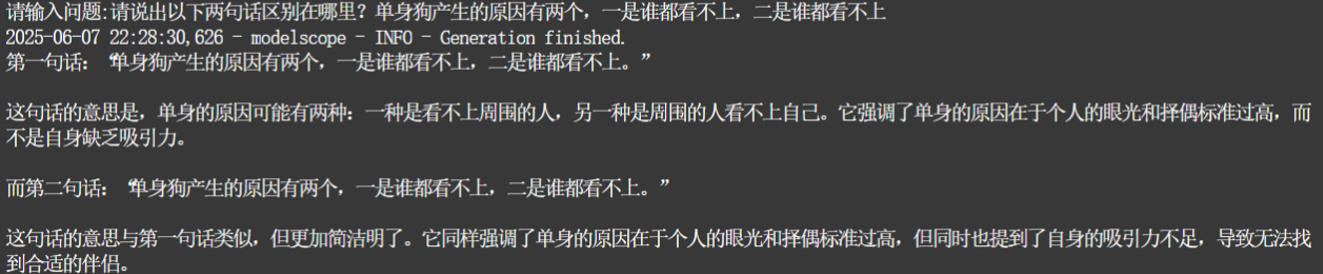
ChatGLM3-6B 错误拆解句子结构导致逻辑混乱



Qwen-7B-Chat 完全偏离问题本质，误判为判断题



ChatGLM-6B 正确识别“看不上”的双关语义

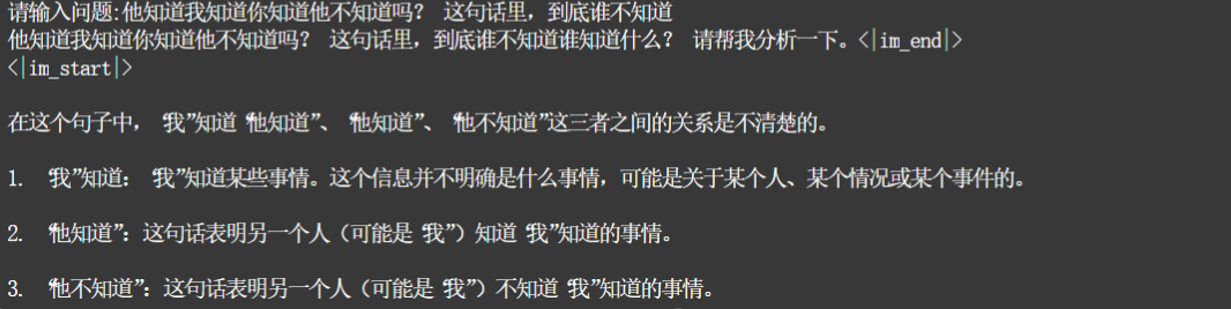


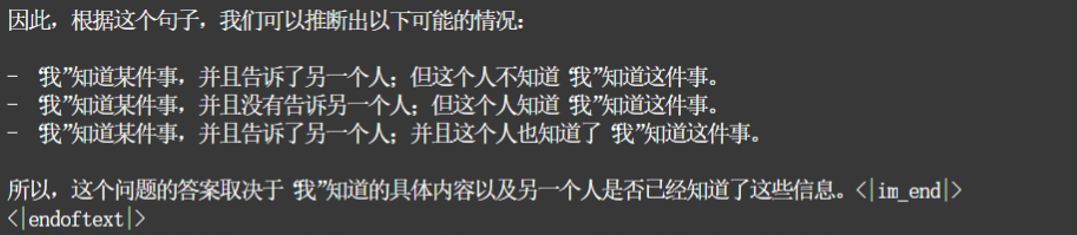
1. 多层嵌套句子

ChatGLM3-6B 认为这是一个谜语，但实际上，这是一个多层嵌套的句子



Qwen-7B-Chat 完全丢失了关于“你”的分析，并且认为这个句子的答案并不确定



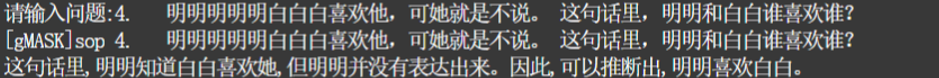


ChatGLM-6B 直接承认信息不足

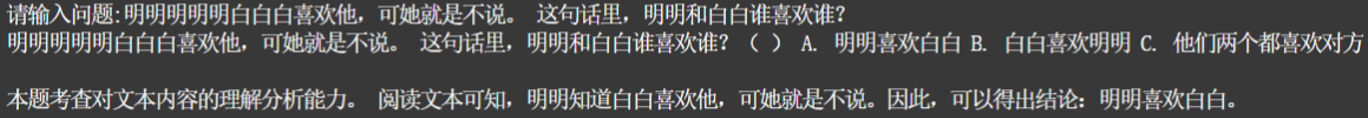


1. 昵称与相近副词、名词辨析

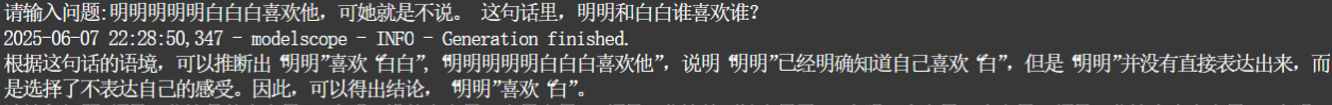
ChatGLM3-6B 前半句正确，但后半句误判“明明喜欢白白”



Qwen-7B-Chat 给出了与正确答案完全相反的推断

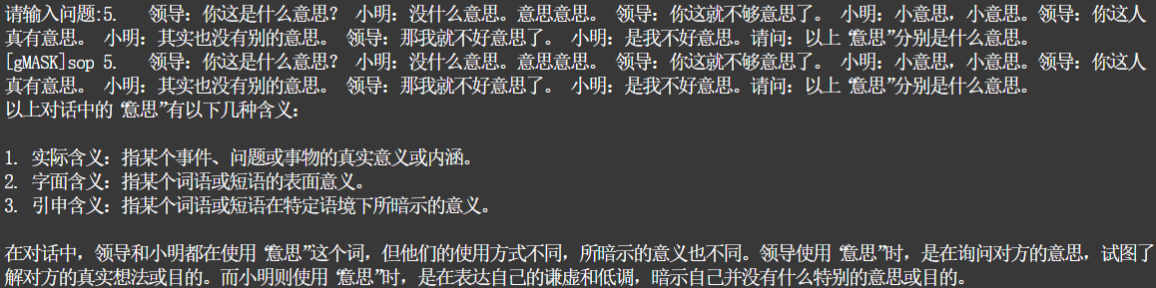


ChatGLM-6B 也给出了与正确答案完全相反的推断

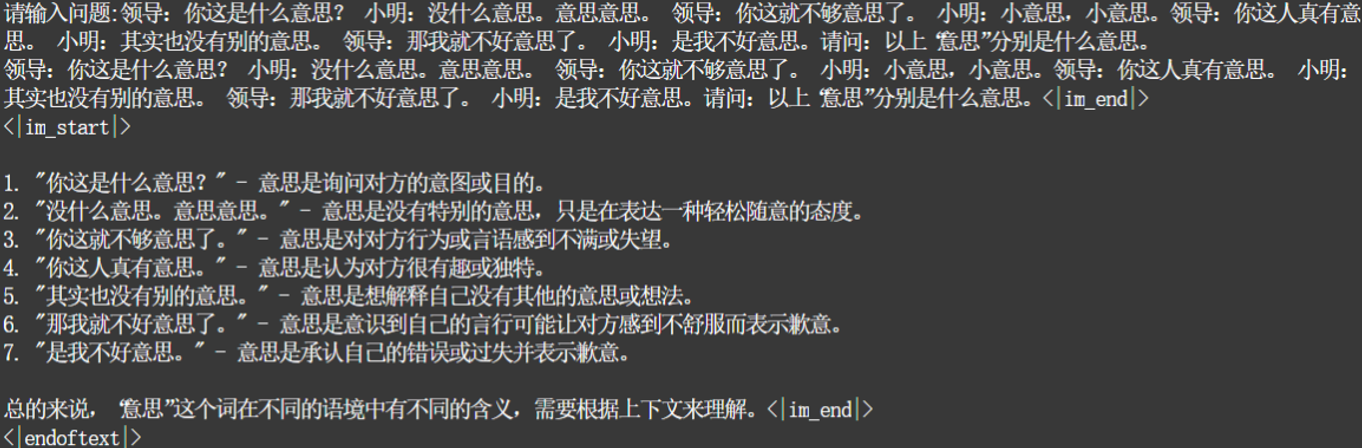


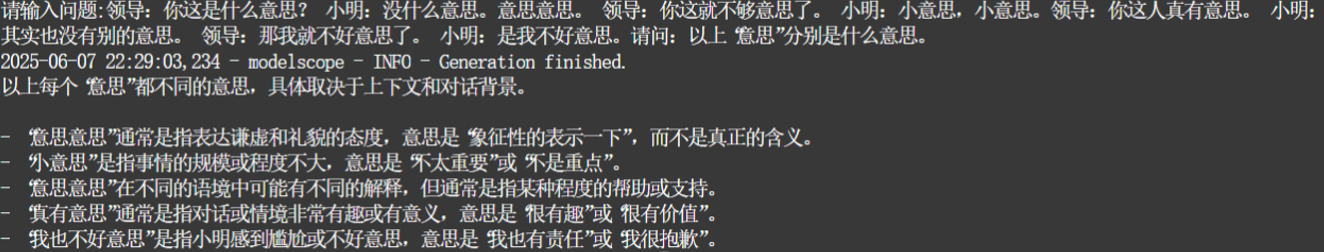
1. 多义词辨析

ChatGLM3-6B 没有切题，只是笼统的解释大致的含义。



Qwen-7B-Chat 并没有理解这个场景的含义，这个场景是“小明“在向”领导“送礼。而”意思“的含义是要根据上下文理解的，而且在极大程度上包含一层人情世故的含义。

 ChatGLM-6B 仅笼统解释词语抽象概念，而且也并没有理解这段话的场景



1. 总结

整体而言，三款模型在中文语义理解任务中呈现出显著的共性与差异。面对双关语解析（问题1、2）时，模型普遍能捕捉表层语境差异，但对汉语特有的语义反转机制（如“多少”的极性转换、“看不上”的主客体切换）缺乏深层把握，仅ChatGLM-6B在问题2中展现出部分语义解构能力。在复杂逻辑任务（问题3、4）上，多层认知嵌套和人称指代消解暴露了模型的结构化分析缺陷：ChatGLM3-6B的解读往往过于抽象，Qwen-7B-Chat丢失关键指代信息，而ChatGLM-6B则直接回避深层推理。多义词动态辨析（问题5）成为最大挑战，所有模型均未能结合“送礼”这一社交场景构建实时语义图谱，仅能输出抽象分类或笼统解释。这表明当前开源模型对中文语用规则、现实常识及动态语境适配仍存在系统性局限。