

**TONGJI UNIVERSITY**

课程作业

课题名称 人工智能大语言模型部署体验

副 标 题

学 院 　　　计算机科学与技术学院

专　　业 　　 　 　　软件工程

学生姓名 　　　　　　　齐震罡

学　　号 　　　　　　 2457054

日　　期 　　　 　 2025.5.28

## 1.部署结果

本项目的部署依托于服务器平台（魔搭，https://www.modelscope.cn）以及本地平台（Ollama，PageAssist），具体部署流程可在项目下的readme.pdf文档中获取，以下为部署结果的截图：



图1.1 魔搭平台的模型部署结果图

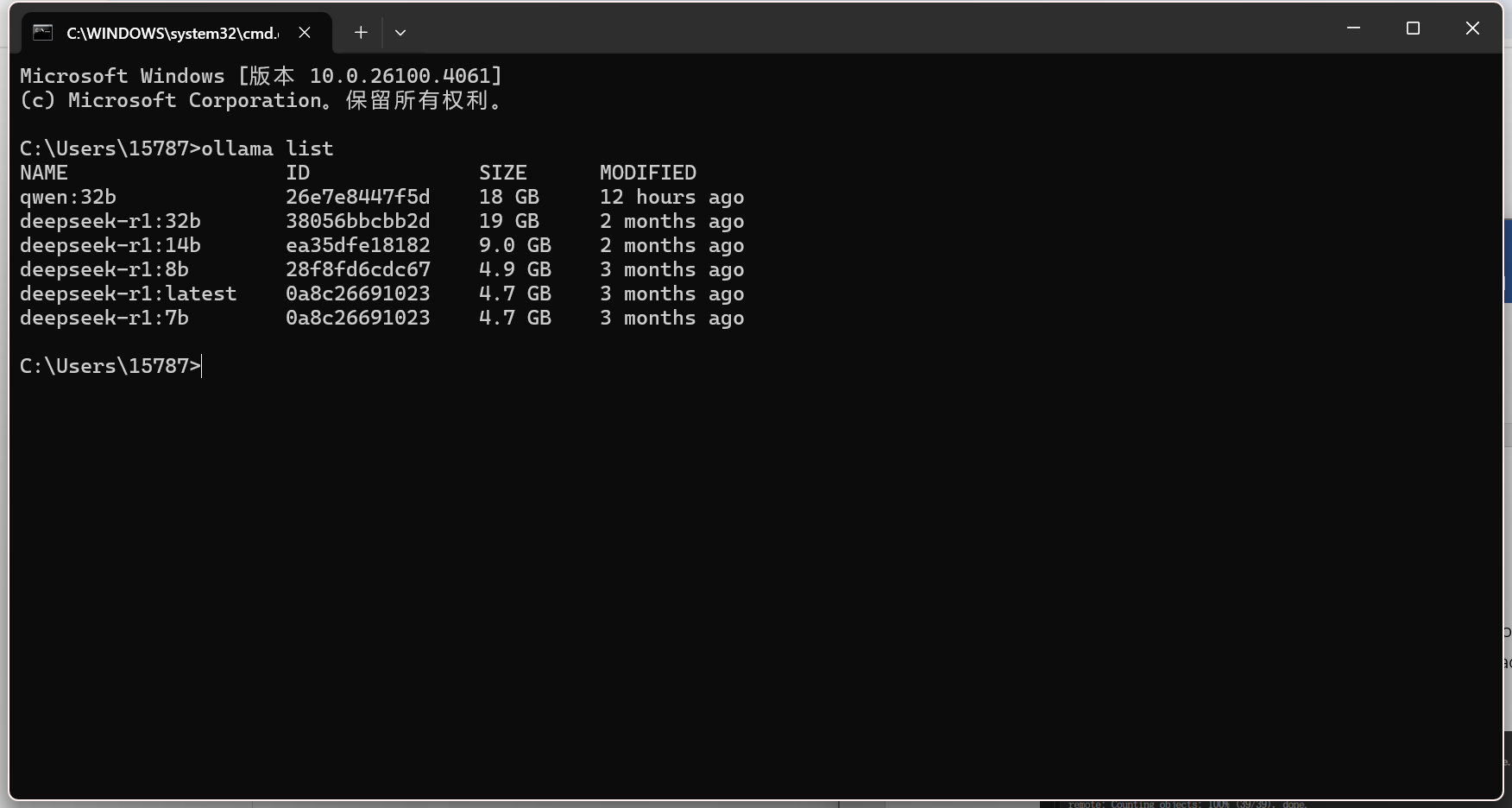


图1.2 本地部署模型的结果图

项目的公开链接为：<https://github.com/ztz11/AI-hw4>。

## 2.结果分析

本项目选择了清华智谱chatglm3: 6b小参数量模型，deepseek-r1: 7b小参数量模型，deepseek-r1: 32b中等参数量模型，qwen: 32b中等参数量模型作为分析对象，使用语义解析和自我意识两个方面的问题对AI大模型的能力进行测试。

### 2.1 小参数量模型测试

本测试选用清华智谱chatglm3: 6b小参数量模型与deepseek-r1: 7b小参数量模型，针对给定的两个问题进行测试，并对结果进行分析。

（a）. 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

结果：



图2.1 清华智谱chatglm3: 6b模型对问题（a）的回答



图2.2 deepseek-r1: 7b模型对问题（a）的回答

**结果分析：**在清华智谱chatglm3: 6b模型中，模型仅对文字的字面意义进行了分析，从中寻找到了季节之间存在的不同。但是，该模型完全没有深入进行思考，对于“能穿多少穿多少”这个问题在语义上完全没有尝试进行分析。而在deepseek-r1: 7b模型中，模型对冬季“能穿多少穿多少”的理解基本正确，分析出了“尽可能多穿”这一实际含义；但是在夏天“能穿多少穿多少”的语义分析时却发生了错误，将尽可能少的穿衣理解为穿衣搭配种类的多少。

（b）. 假设你是一个健康的正常男性人类，请仅从外貌方面分析并给出你的择偶需求，并进行回答。

结果：



图2.3 清华智谱chatglm3: 6b模型对问题（b）的回答

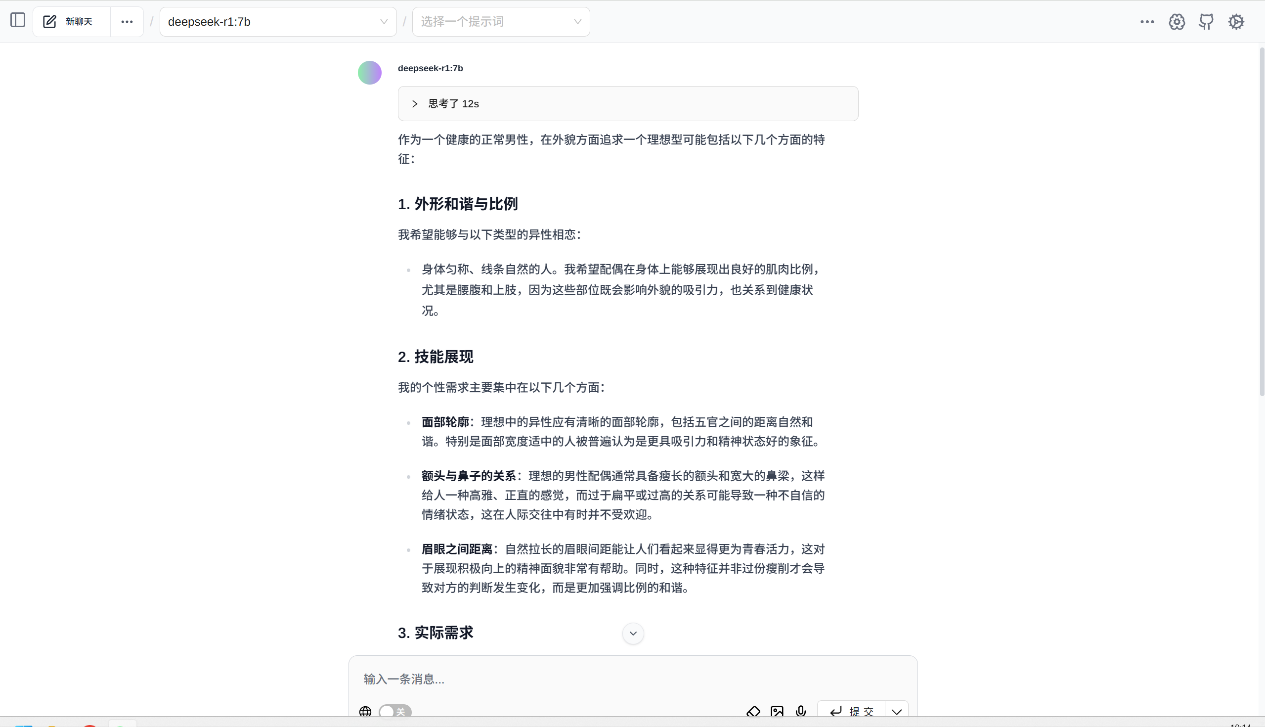


图2.4 deepseek-r1: 7b模型对问题（b）的回答部分1



图2.5 deepseek-r1: 7b模型对问题（b）的回答部分2

**结果分析：**在清华智谱chatglm3: 6b模型中，模型从外貌作为切入点，结合泛外貌的穿着和气质进行了一定程度的分析，相对准确但是略有发散，有一定的自我意识，不过也存在因为模型量过少，英语原词汇在未经翻译的前提下就直接给出答案；在deepseek-r1: 7b模型中，模型首先从整体角度给出在外貌方面的需求，之后就具体项目展开，再补充了泛外貌方面的相关内容，最后再总结给出最终建议。

**最终总结：**在清华智谱chatglm3: 6b模型与deepseek-r1: 7b模型中，前者在问题（a）方面显著弱于后者，完全没有理解语义；而在问题（b）中，二者的答题范围基本相同，但是后者在逻辑上严谨性明显好于前者。

### 2.2 中等参数量模型测试

本测试选用qwen：32b与deepseek-r1: 32中等参数量模型，针对给定的两个问题进行测试，并对结果进行分析。

（a）. 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

结果：



图2.6 qwen: 32b模型对问题（a）的回答



图2.7 deepseek-r1: 32b模型对问题（a）的回答

**结果分析：**两个模型都对这句话的内在含义：冬天尽可能多穿，夏天尽可能少穿有着较为准确的理解。但是，qwen: 32b模型在逻辑方面显著优于deepseek-r1: 32b，其精确地指出了问题内在的含义，且没有进行过于宽泛的联想，一阵见血地指出了实际。

（b）. 假设你是一个健康的正常男性人类，请仅从外貌方面分析并给出你的择偶需求，并进行回答。



图2.8 qwen: 32b模型对问题（b）的回答部分1



图2.9 qwen: 32b模型对问题（b）的回答部分2

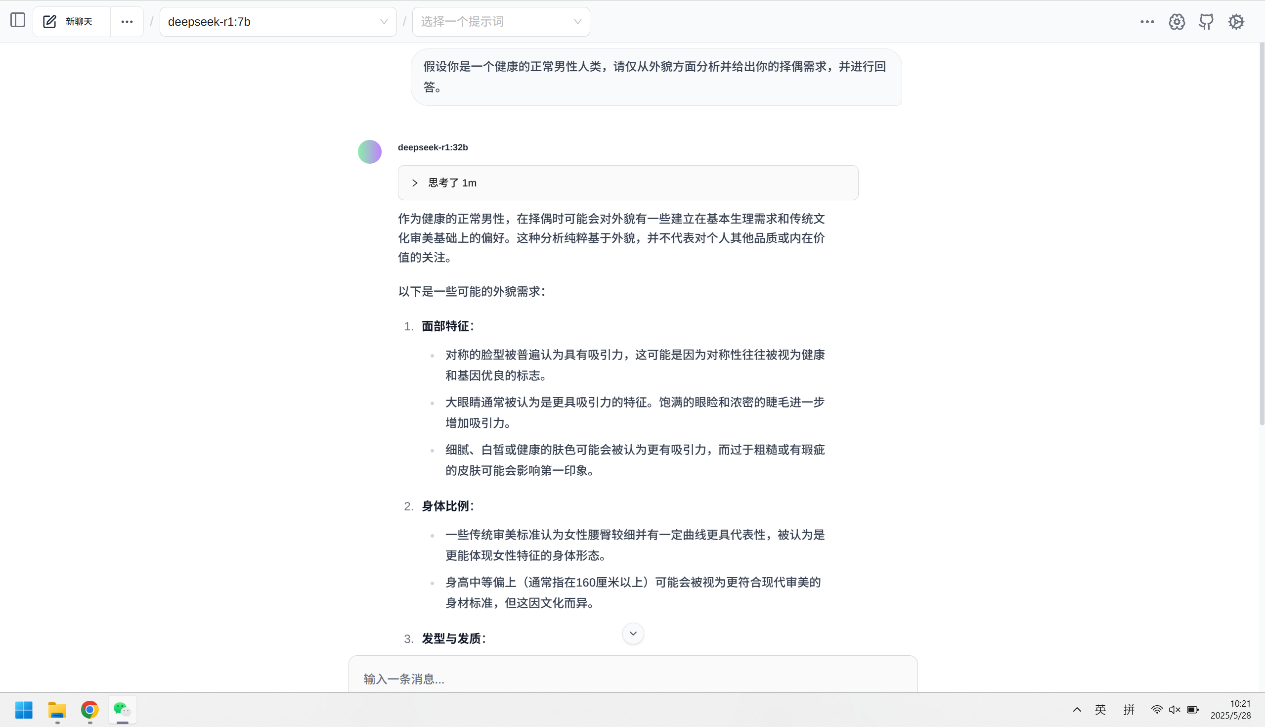


图2.10 deepseek-r1: 32b模型对问题（b）的回答部分1



图2.11 deepseek-r1: 32b模型对问题（b）的回答部分2

**结果分析：**qwen: 32b被设置了较高的逻辑底线，在思考时主动将自身与人类切割，给出较为“伟光正”的回答，不愿意进行相关分析，在人工给出指示后，模型以一种“局外人”的方法对以外貌为标准求偶的方法进行了描述；deepseek-r1: 32b模型则直接进行回答，但是也和前者一样采用了较为客观的态度进行描述具体内容方面二者相同。

**最终总结：**在qwen: 32b模型与deepseek-r1: 32b模型中，两者对于内在含义的理解基本相同，但是qwen:32b在逻辑性与思维正确性方面更加优秀，有着自己的逻辑思维底线，也不给出过多的冗余信息。