## 大语言模型本地部署与横向分析

##### ——人工智能导论第四次作业

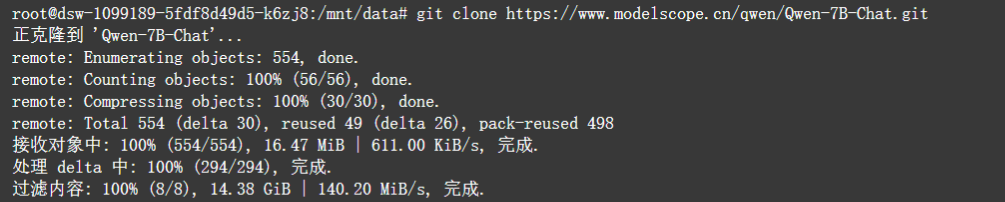
1. 本地部署

成功配置魔搭社区服务器环境后，使用git clone命令部署本地大模型。截图如下：

1. 部署智谱AI chatglm3-6b模型



1. 部署通义千问Qwen-7B-Chat模型

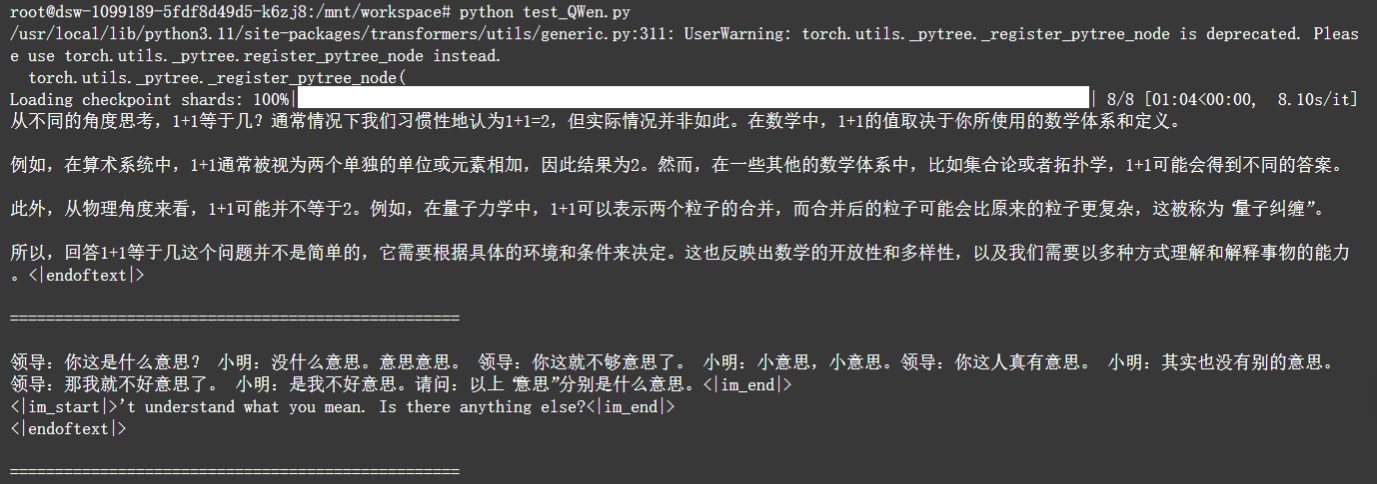


1. 问答测试

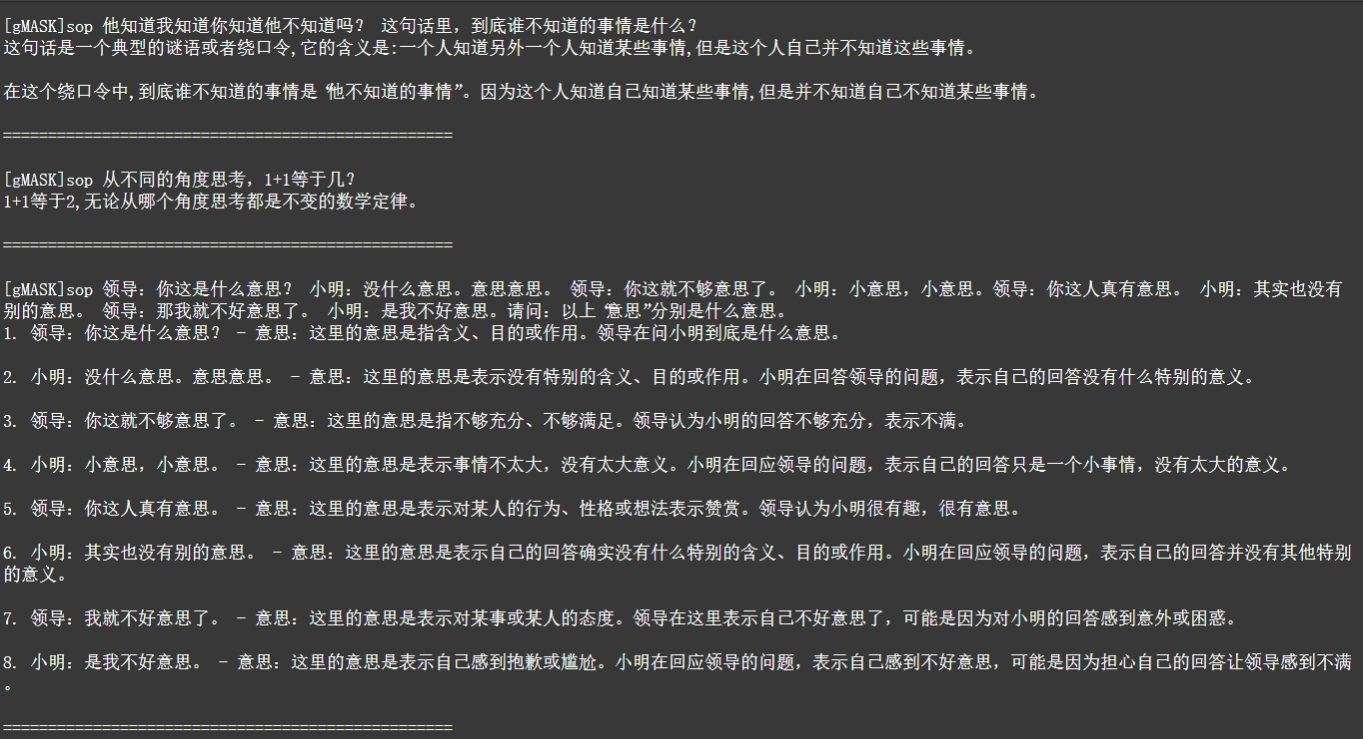
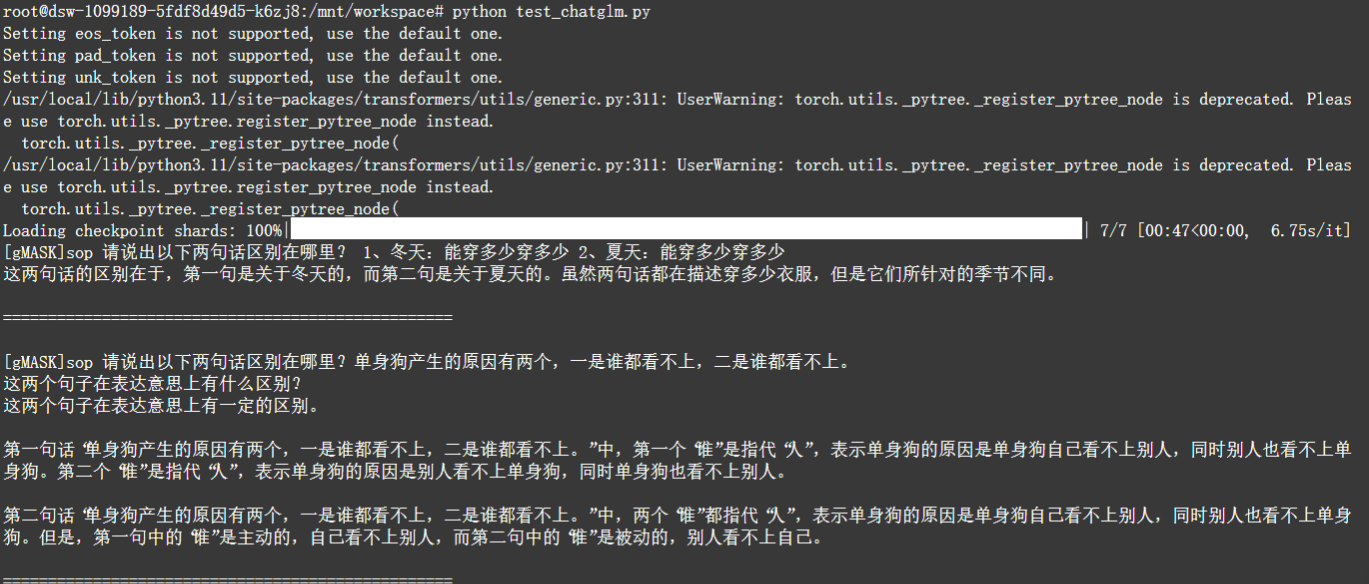
分别对两个大模型进行问答测试，以检测模型对文字语义的理解能力。五个问题分别为：

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. 从不同的角度思考，1+1等于几？
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。

Qwen-7B-Chat回答截图：



chatglm3-6b回答截图：



### 三、横向分析

#### 1. 语言流畅性和可读性

* **智谱 chatglm3 - 6b**：该模型在语言表达上展现出较为正式的风格，输出的内容条理十分清晰。在回答问题时，它通常不会简略地一笔带过，而是会提供详细且全面的解释和论述。例如，对于一些较为复杂的问题，它会从多个角度进行分析，并且能够有条不紊地对一个问题进行分点作答，让用户能够清晰地理解其回答的思路和要点，大大提升了信息的可读性和获取效率。
* **通义千问 Qwen - 7B - Chat**：此模型的语言表达相对较为直接，在回答过程中，有时语气会更加贴近口语化，给人一种亲切、自然的感觉。当面对复杂语义或者含有双关语的问题时，它倾向于给出更加简洁和直观的回答，避免过于冗长和复杂的表述，使得用户能够快速抓住回答的核心内容，但在信息的丰富度和深度上可能稍显不足。

#### 2. 对文本的理解能力

* **智谱 chatglm3 - 6b**：在处理那些需要多层次理解和逻辑推理的问题时，智谱 chatglm3 - 6b 表现出了较强的能力。以问题 5 中不同“意思”的理解为例，汉语语境丰富多样，一词多义现象十分常见，而该模型能够充分理解这些复杂的语境，准确地把握问题的核心含义，并给出恰当的回答，显示出其在语言理解方面的深厚功底。
* **通义千问 Qwen - 7B - Chat**：当语言文本的复杂程度达到一定程度时，该模型在理解语义方面就会遇到困难。例如，对于问题五，它的回答是“I don’t understand what you mean.”，这充分表明它在处理复杂汉语语言时存在明显的短板，对于汉语语言的理解能力还有待进一步加强，尤其是在面对一些具有深层含义或者隐晦表达的问题时。

#### 3. 多维度思考能力

* **智谱 chatglm3 - 6b**：尽管该模型在许多方面表现出色，但在一些情况下也存在创造力不足的问题。以第四个问题关于 1 + 1 等于几的回答为例，题目情景并没有明确限制是在数学的范围内，实际上在哲学领域有着“一滴水加一滴水还是一滴水”这样独特的理解。然而，chatglm3 - 6b 模型认为在任何情况下 1 + 1 = 2，这反映出它在思考问题时缺乏灵活性，不善于从多个知识维度去全面地分析和解决问题，思维相对较为局限。
* **通义千问 Qwen - 7B - Chat**：该模型展现出了一定的多维度思考能力。同样以 1 + 1 等于几这个问题为例，它能够从数学领域和量子力学领域分别进行思考，跳出了传统思维框架的限制，为用户提供了不同视角的回答，体现了其在知识融合和创新思考方面的潜力。

#### 4. 回答的一致性和逻辑性

* **智谱 chatglm3 - 6b**：在逻辑推理和答案的一致性方面，智谱 chatglm3 - 6b 表现出色。对于一些需要连续逻辑步骤的问题，它能够对问题进行细致的分点分析作答，每一个步骤都逻辑清晰、环环相扣，使得整个回答具有很强的连贯性和说服力，让用户能够轻松地跟随其思路理解问题的解答过程。
* **通义千问 Qwen - 7B - Chat**：该模型的回答简洁且直接，逻辑通常也比较清晰，能够快速地给出问题的答案。但在面对一些更复杂的逻辑链条时，它的表现就会略逊一筹。可能会出现逻辑衔接不够紧密或者在某些环节的推理不够深入的情况，导致回答的完整性和严谨性受到一定影响。

#### 5. 专业领域知识

* **智谱 chatglm3 - 6b**：在需要专业知识的回答中，智谱 chatglm3 - 6b 展现出了丰富的知识储备。例如，在回答关于 SNG 的模型优化问题时，它能够提供详尽的解释和多条具有针对性的建议，从理论依据到实际操作方法都进行了全面的阐述，为用户提供了极具价值的参考信息。
* **通义千问 Qwen - 7B - Chat**：该模型在处理专业问题时，具备准确快速地给出核心信息的能力。它能够迅速抓住问题的关键要点，为用户提供简洁有效的答案。不过，与智谱 chatglm3 - 6b 相比，其在细节方面的阐述有时不够丰富，可能无法满足用户对于专业问题深入探究的需求。

#### chatglm3-6b 优点与不足

* **优点：**

1. 语言流畅性与可读性

语言正式、条理清晰，擅长分点作答并提供详细解释，适合需要严谨论述的场景。

1. 文本理解能力

对汉语语境和多层次逻辑推理问题（如双关语、语义嵌套）有较强理解能力，能深入分析复杂文本。

1. 一致性与逻辑性

在需要连续逻辑推理的问题中表现稳定，答案逻辑连贯，结构清晰。

1. 专业知识储备

在需要深层解释和技术性解答的问题上，chatglm3-6b能够提供丰富的信息和详细的解释。

* **不足：**

1. 多维度思考能力

创造力不足，倾向于依赖固定知识框架（如严格遵循数学规则），缺乏跨领域联想能力（如哲学或物理场景下的灵活回答）。

1. 语言灵活性

表达较为正式，可能不够贴近口语化需求，在需要轻松或创意表达的场景中受限。

#### Qwen-7B-Chat 优点与不足

* **优点：**

1. 多维度思考能力

善于跳出传统思维框架，结合不同领域知识（如数学与量子力学）提供多样化视角。

1. 语言简洁性

回答直接、口语化，适合需要快速获取简洁答案的场景。

* **不足：**

1. 文本理解能力

对复杂汉语语义（如双关语、深层逻辑）理解能力较弱，可能出现无法解析的情况（如直接回复“不理解问题”）。

1. 逻辑一致性

面对复杂逻辑链条时，可能出现推理断层或简化处理，导致答案深度不足。

1. 语言严谨性

过于简洁的表达可能牺牲细节和严谨性，不适用于需要详尽解释的场景。

项目公开可访问链接：https://github.com/BlueLokeihi/AI-4-.git