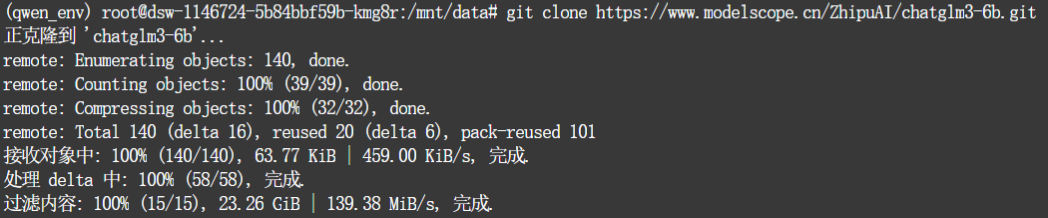
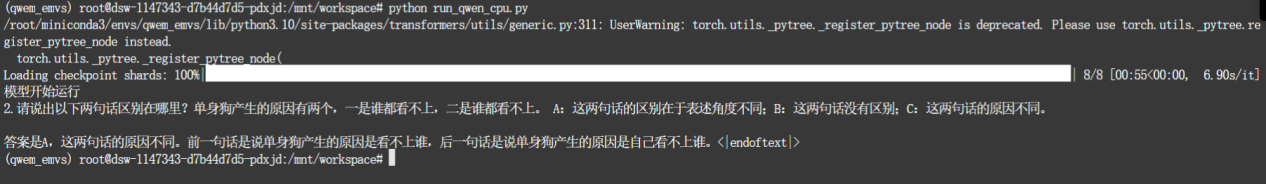
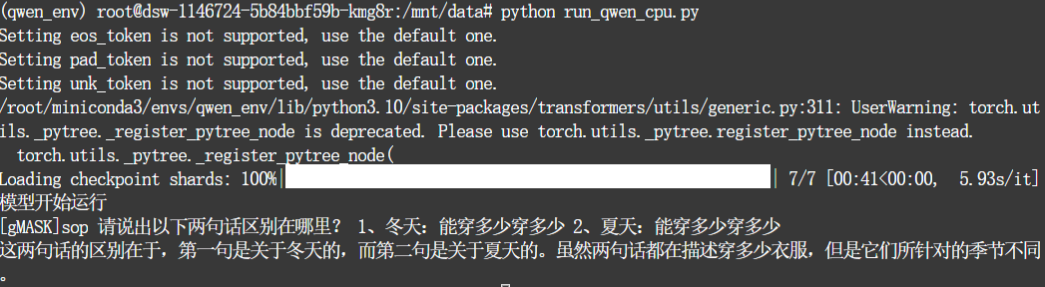
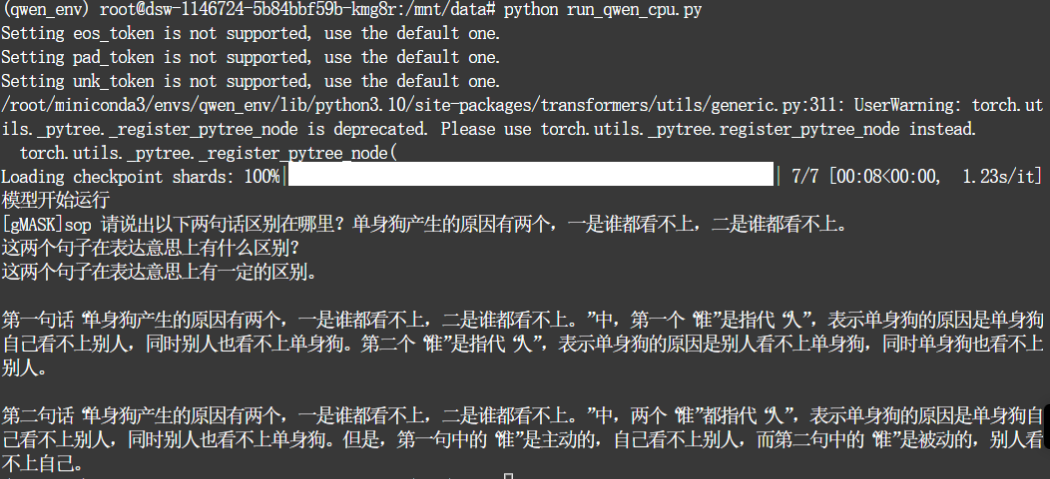
# 人工智能导论第四次作业报告

1.GIT克隆成功图片：  


2.回答测试问题的截图

## (1)Qwen-7B-Chat

  
(2)chatglm3-6b



## 两个模型的对比

1. 模型基础信息差异
2. 参数规模与来源：

·chatglm3-6b：由ZhipuA1团队开发，参数规模为60亿（6b） ·Qwen-7B-Chat：来自qwen团队，参数规模更大，为70亿（7B）， 2. 模型文件体量：

· chatglm3-6b的远程文件数量为140个，接收数据总量约63.77 KiB，整体文件量较小。

· Qwen-7B-Chat的远程文件数量达554个，接收数据量约16.47 MiB（约为chatglm3-6b的258倍），文件体量显著更大。

1. 克隆与运行过程对比
2. 克隆速度与处理效率

· chatglm3-6b：接收速度为459.00 KiB/s，过滤速度139.38 MiB/s，因文件量少，整体克隆耗时更短。

Qwen-7B-Chat：接收速度34.27 MiB/s，过滤速度148.39 MiB/s，虽过滤速度略快，但因数据量庞大，克隆耗时更长（如处理delta数量294个，远超chatglm3-6b的58个）。

1. 运行时表现：  
    加载速度方面，chatglm3-6b完成7/7步骤最快耗时8秒（1.23s/it），Qwen-7B-Chat完成8/8步骤最快耗时49秒（6.19s/it），chatglm3-6b在小批量处理中更高效。
2. 回答逻辑与语义理解差异
3. 对“季节穿衣问题”的回答:  
    chatglm3-6b：明确指出冬天“多穿”、夏天“少穿”，因季节不同导致语义相反，逻辑清晰且直击核心。

Qwen-7B-Chat：仅提到季节差异，未解释“穿多少”在不同季节的语义反转（如未说明冬天需保暖、夏天需凉爽），回答更简略。

1. 对“单身狗原因问题”的回答

chatglm3-6b的回答存在矛盾：部分内容认为两句话“没有区别”（选项B），另一部分却分析“谁”的指代差异（主动看不上 vs. 被动被看不上），认为原因不同（选项C）。

Qwen-7B-Chat则强调“表述角度不同”，认为前一句侧重“自己看不上别人”，后一句侧重“别人看不上自己”，但对“谁”的指代解读略显复杂，存在歧义。

1. 对“逻辑悖论问题”的回答

chatglm3-6b直接指出句子是“无法解决的悖论”，基于逻辑矛盾推导结论，表述明确。

Qwen-7B-Chat将其解读为“谜语或绕口令”，认为涉及复杂语境设置，未直接拆解逻辑矛盾，回答更偏向隐喻式解释。

1. 总结与延伸
2. 核心差异

·模型规模与效率：Qwen-7B-Chat参数更多、数据量更大，适合复杂任务，但克隆和加载耗时更长；chatglm3-6b更轻量，处理小批量任务更高效。

·回答风格：chatglm3-6b注重语义本质，回答直接严谨；Qwen-7B-Chat有时倾向复杂语境分析，但部分问题回答不够深入或存在歧义。

2. 潜在应用场景：  
 chatglm3-6b可能更适合轻量级对话、快速问答；Qwen-7B-Chat因参数优势，可能在多轮推理、长文本理解等场景中更具潜力。