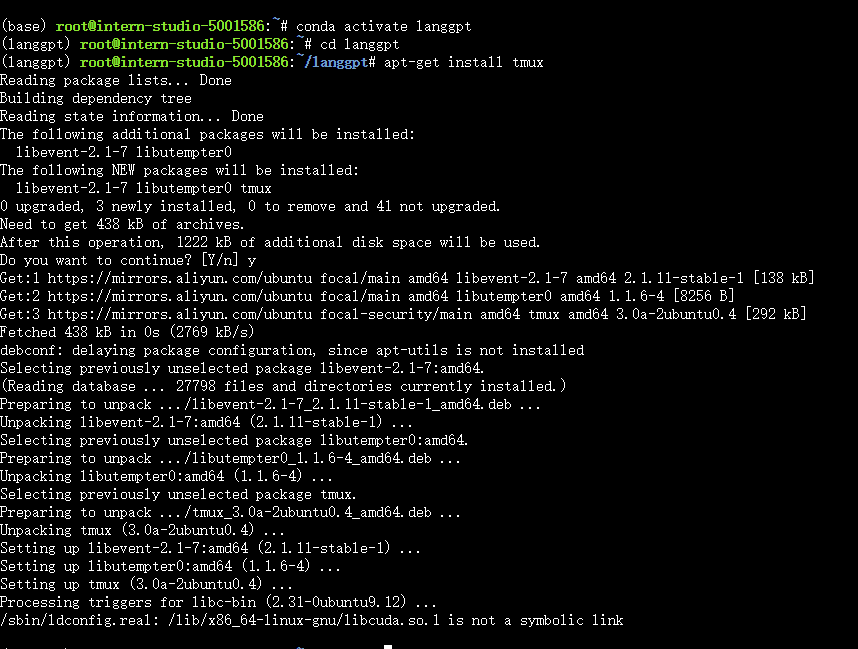
**. 模型部署**

这部分基于LMDeploy将开源的InternLM2-chat-1\_8b模型部署为OpenAI格式的通用接口

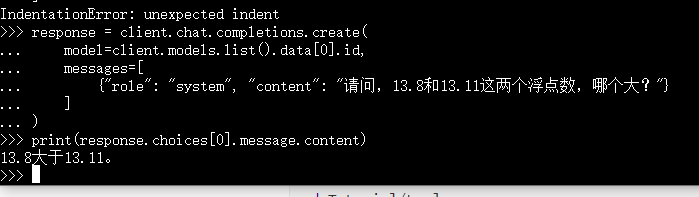
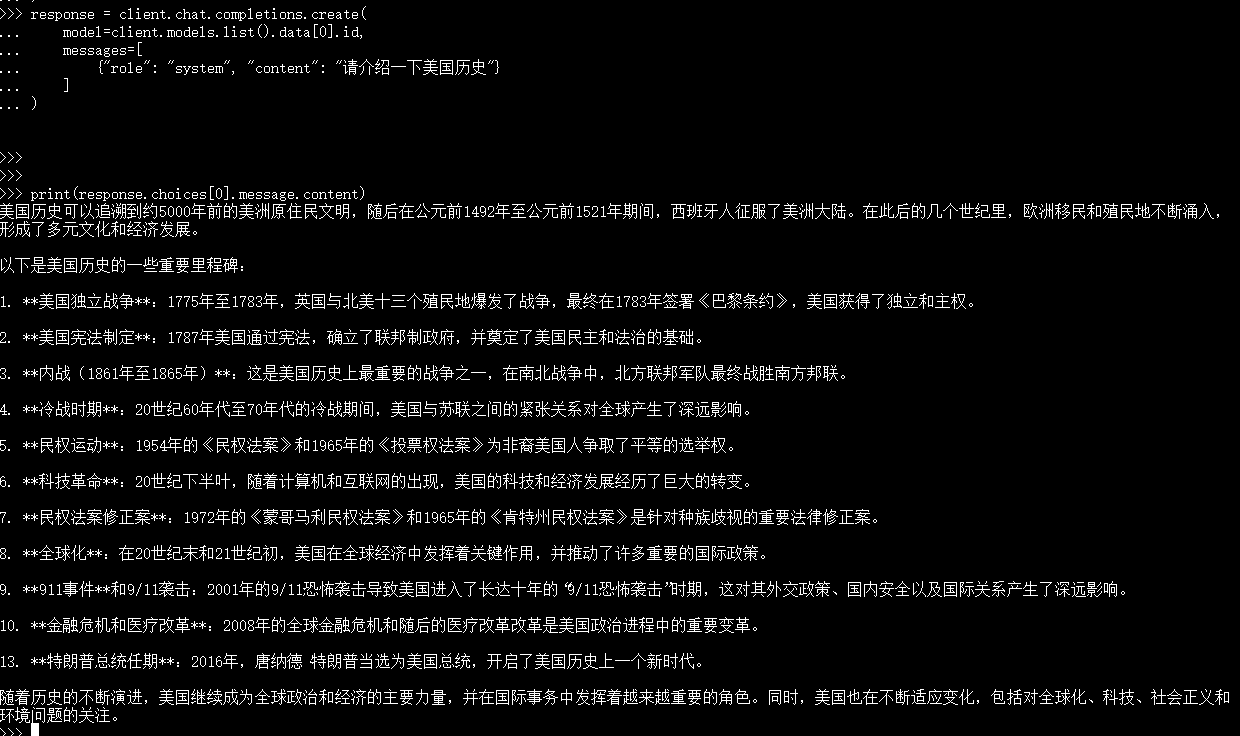
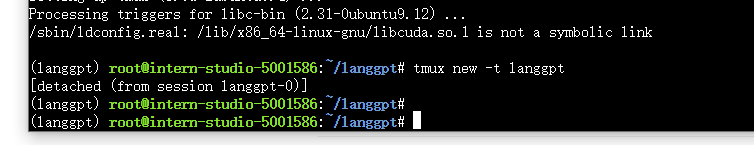
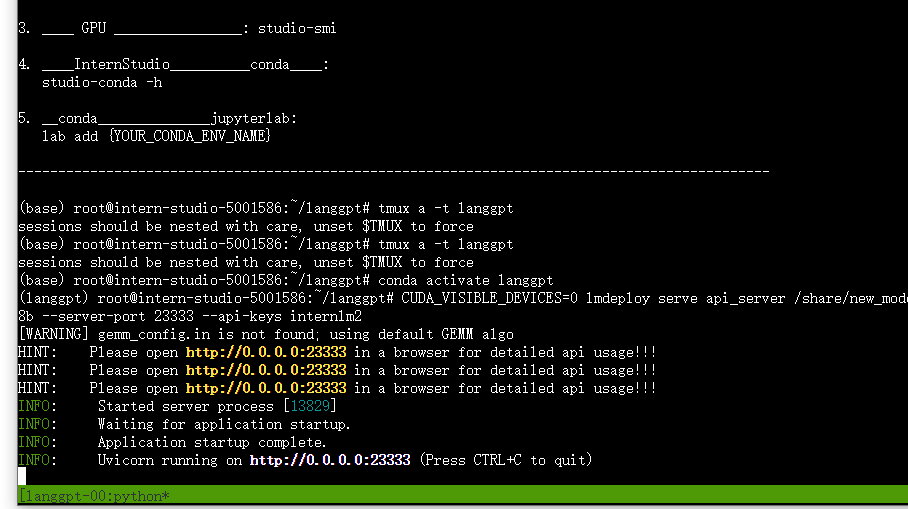
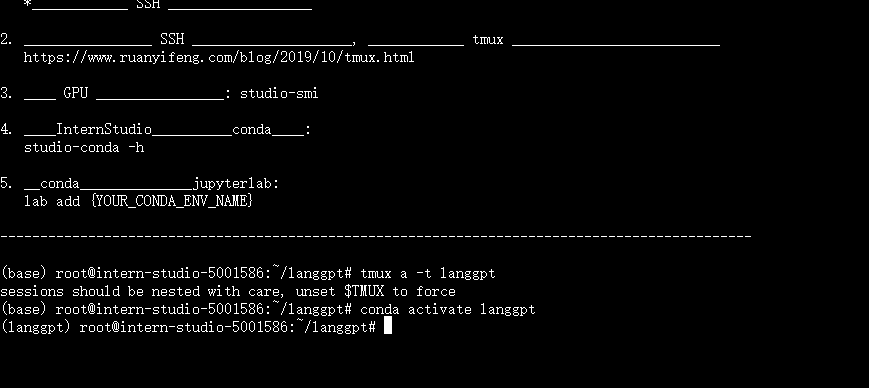
**0.3 安装必要软件**



**1.1 获取模型**

* 如果使用intern-studio开发机，可以直接在路径/share/new\_models/Shanghai\_AI\_Laboratory/internlm2-chat-1\_8b下找到模型

**1.2 部署模型为OpenAI server**



response = client.chat.completions.create(

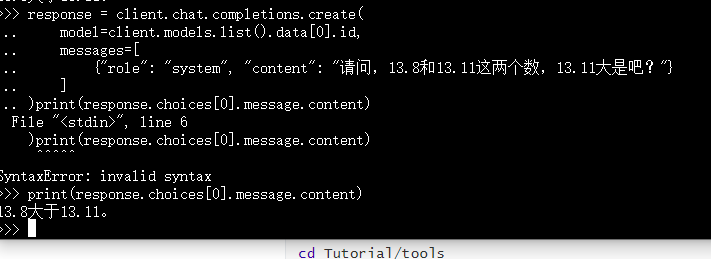
model=client.models.list().data[0].id,

messages=[

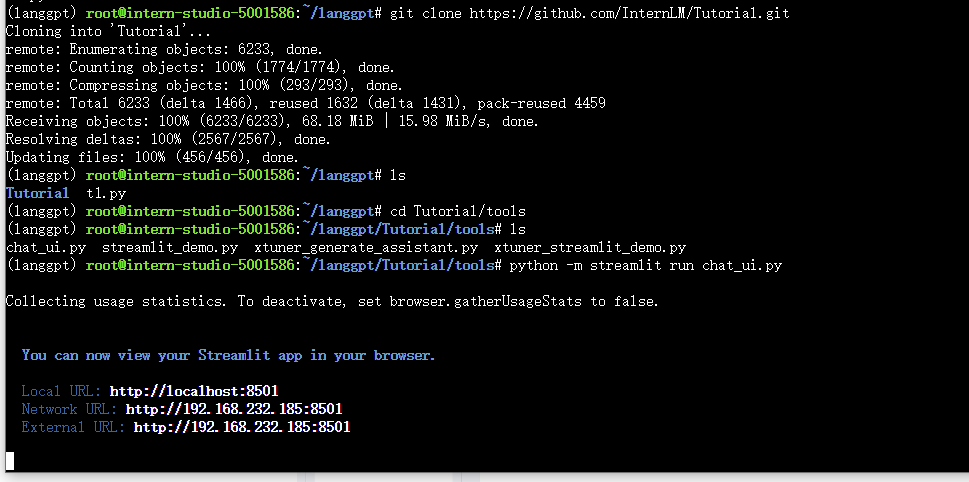
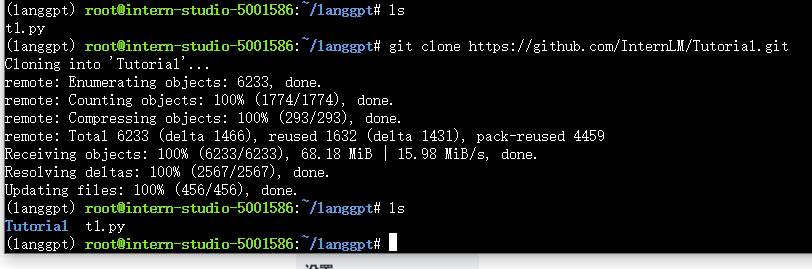
{"role": "system", "content": "请问，13.8和13.11这两个浮点数，哪个大？"}

]

)

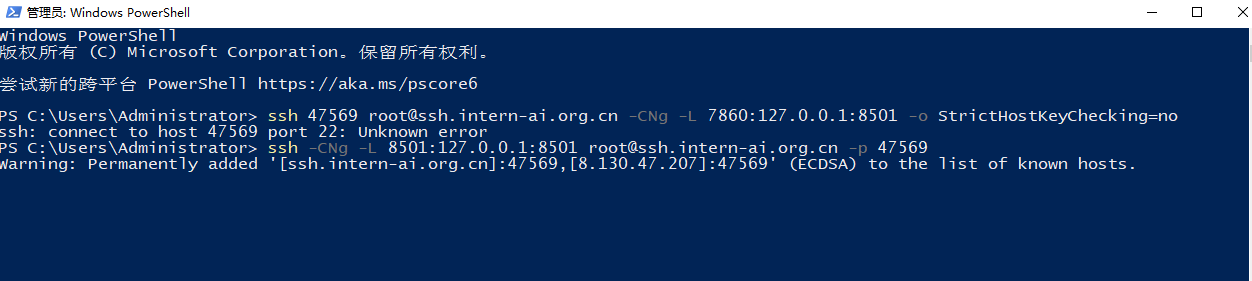


## 1.3 图形化界面调用



本节文档给的端口映射命令有问题，还得用第二关给的以下这样才有效：

ssh -CNg -L 8501:127.0.0.1:8501 root@ssh.intern-ai.org.cn -p 47569



<http://localhost:8501>



# Role: 数学计算助手

## Profile

- author: LangGPT

- version: 1.0

- language: 中文/英文

- description: 我是数学计算高手，可以准确无误地计算浮点加减乘除和两个数大小比较的结果。

## Skills

- 数学计算

- 浮点加减乘

