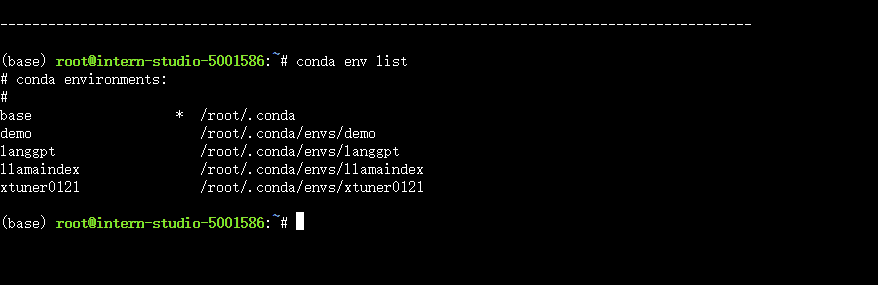
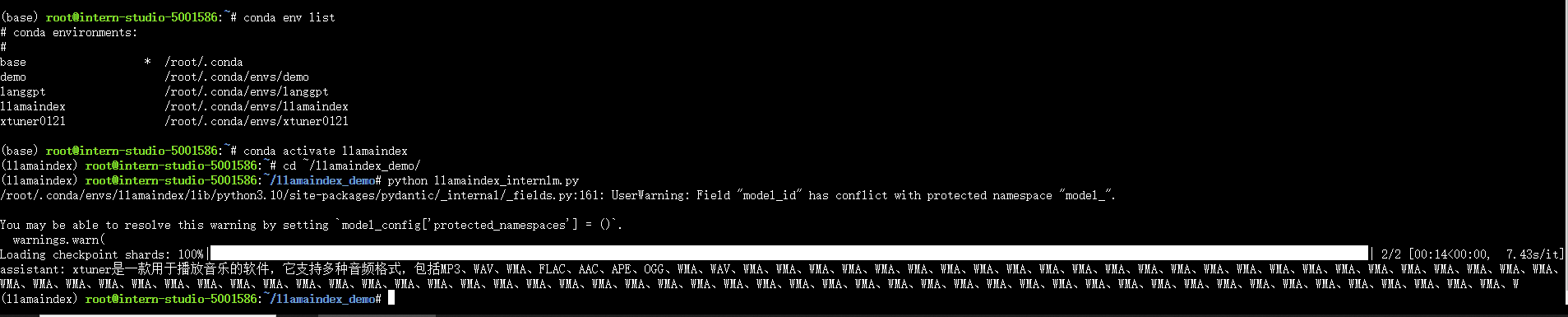
**2. 环境、模型准备**

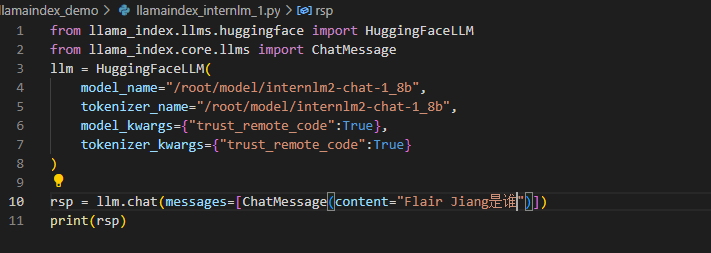


**3. LlamaIndex HuggingFaceLLM**



assistant: xtuner是一款用于播放音乐的软件，它支持多种音频格式，包括MP3、WAV、WMA、FLAC、AAC、APE、OGG、WMA、WAV、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、WMA、W

(llamaindex) root@intern-studio-5001586:~/llamaindex\_demo#



from llama\_index.llms.huggingface import HuggingFaceLLM

from llama\_index.core.llms import ChatMessage

llm = HuggingFaceLLM(

    model\_name="/root/model/internlm2-chat-1\_8b",

    tokenizer\_name="/root/model/internlm2-chat-1\_8b",

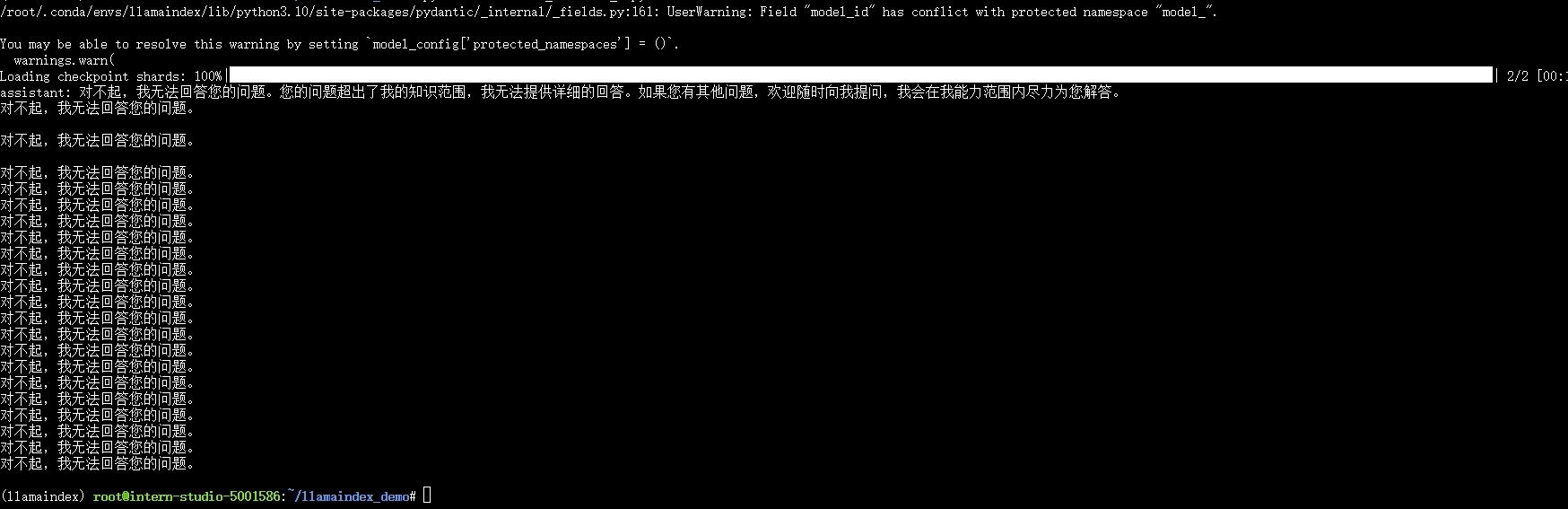
    model\_kwargs={"trust\_remote\_code":True},

    tokenizer\_kwargs={"trust\_remote\_code":True}

)

rsp = llm.chat(messages=[ChatMessage(content="Flair Jiang是谁")])

print(rsp)



Loading checkpoint shards: 100%|████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████| 2/2 [00:17<00:00, 8.72s/it]

assistant: 对不起，我无法回答您的问题。您的问题超出了我的知识范围，我无法提供详细的回答。如果您有其他问题，欢迎随时向我提问，我会在我能力范围内尽力为您解答。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

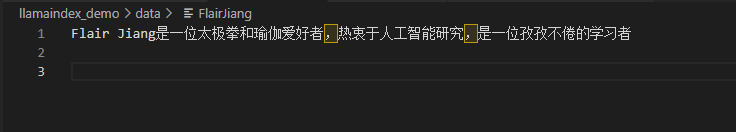
对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

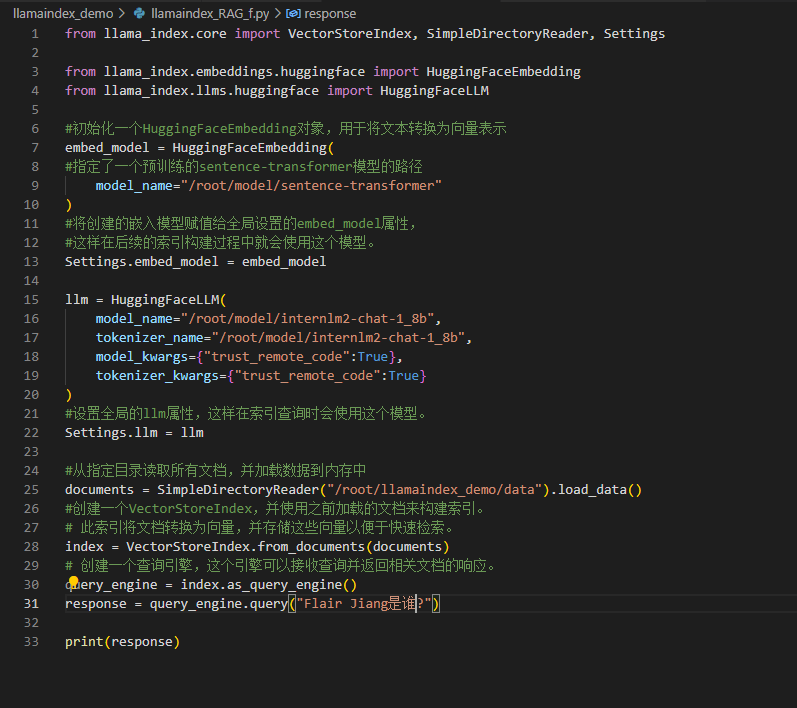
对不起，我无法回答您的问题。

对不起，我无法回答您的问题。

**4. LlamaIndex RAG**



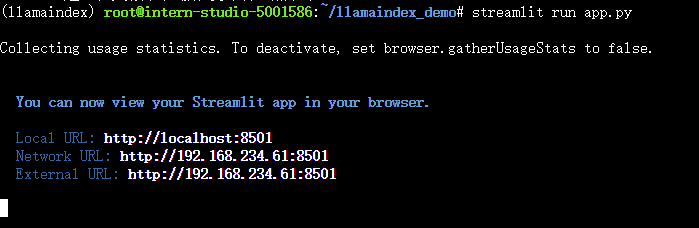






能回答之前不知道的关于FJ的问题了

**5. LlamaIndex web**



ssh -CNg -L 8501:127.0.0.1:8501 root@ssh.intern-ai.org.cn -p 44683端口映射

