**终端当前环境和目录：**

**conda activate eecs449; cd ~/EECS449/EECS449\_Warm-Up**

**Task 1:**

curl --location 'http://localhost:8000/user/register' \

--header 'Content-Type: application/json' \

--header 'Accept: application/json' \

--data '{

"password": "haha",

"email": "test22@mail.com"

}'

curl --location 'http://localhost:8000/user/login' \

--header 'Content-Type: application/json' \

--header 'Accept: application/json' \

--data '{

"password": "haha",

"email": "test22@mail.com"

}'

curl -X POST http://localhost:8000/walker/interact -H "Authorization: Bearer ey..."

curl -X POST http://localhost:8000/walker/time\_greeting -H "Authorization: Bearer **<token>**" -H "Content-Type: application/json" -d '{"name": "Alice", "hour": 15}'

curl -X POST http://localhost:8000/walker/reverse\_string -H "Authorization: Bearer **<token>**" -H "Content-Type: application/json" -d '{"text": "Hello World!"}'

**Server:**

OpenAI API(付费的):

~~export OPENAI\_API\_KEY=”sk-w6O\_NRYWMH3fl1YZiaztdbvzu-j0o8DDMeXKhxXLgyT3BlbkFJzk3ok4Q4epvFaM2x3\_9VRGQWc6l-tRVTddO\_VnUhMA”~~

~~DATABASE\_HOST=mongodb://localhost:27017/?replicaSet=my-rs jac serve server-openai.jac~~

**Ollama(本地的):**

Step 0: **docker run --name mongodb -p 27017:27017 -d mongodb/mongodb-community-server:latest --replSet my-rs**

In another terminal (第一次)：

**mongosh --port 27017**

**rs.initiate({\_id: "my-rs", members: [{\_id: 0, host: "localhost"}]})**

**exit**

(另外，如果有其他的占用可以进行如下的清理)

docker stop aa3d2ac9a046

docker rm aa3d2ac9a046

sudo lsof -i :27017

sudo kill -9 577

**注意千万不要使用这个systemctl mongod！！**（禁用方法：sudo systemctl disable mongod）

**3个终端分别运行：**

**(1) ollama serve （这个只需在后台运行即可，也就是可以终端意外退出，比如下面这样）**

(eecs449) qilong@MARSCAT:~/EECS449/EECS449\_Warm-Up$ sudo lsof -i :11434

[sudo] password for qilong:

COMMAND PID USER FD TYPE DEVICE SIZE/OFF NODE NAME

ollama 24630 ollama 3u IPv4 63178 0t0 TCP localhost:11434 (LISTEN)

ollama 24630 ollama 8u IPv4 95633 0t0 TCP localhost:11434->localhost:49632 (ESTABLISHED)

jac 98961 qilong 22u IPv4 99584 0t0 TCP localhost:49632->localhost:11434 (ESTABLISHED)

**(2) DATABASE\_HOST=mongodb://localhost:27017/?replicaSet=my-rs jac serve server-ollama.jac（其中server-ollama.jac 可替换成其他的LLM模型，如server-qwen\_0-5b.jac）**

**Client:**

**(3) jac streamlit client.jac**

(另外，如果有其他的占用可以进行如下的清理)

**docker ps**

**sudo lsof -i :8000**

**sudo lsof -i :11434**

**sudo kill -9 19161**

**此课程使用的专用语言Jac不支持Hugging Face’s Models**

**~~使用新的LLM模型的方法~~** ~~- 例如[FallenMerick/Chewy-Lemon-Cookie-11B · Hugging Face](https://huggingface.co/FallenMerick/Chewy-Lemon-Cookie-11B)：~~

~~├── server-Huggingface-Chewy-Lemon-Cookie-11B.jac~~

~~├── server\_Huggingface\_Chewy\_Lemon\_Cookie\_11B\_model.py~~

[**~~unsloth/Llama-3.2-1B · Hugging Face~~**](https://huggingface.co/unsloth/Llama-3.2-1B)

**如何卸载HuggineFace下载过的LLM模型：**

**1. 找到模型存储位置**

**$ echo $HOME/.cache/huggingface/hub/\***

/home/qilong/.cache/huggingface/hub/models--FallenMerick--Chewy-Lemon-Cookie-11B

/home/qilong/.cache/huggingface/hub/models--gpt2 /home/qilong/.cache/huggingface/hub/version.txt

**2. 删除特定模型**

如果你知道要删除的模型名称，可以直接导航到缓存目录并删除相应的文件夹。例如：

**rm -rf ~/.cache/huggingface/transformers/{模型名称}**

例如：

rm -rf ~/.cache/huggingface/hub/models--FallenMerick--Chewy-Lemon-Cookie-11B

4. 使用 transformers-cli（可选）

如果你安装了 transformers-cli，你也可以使用它来清除缓存。你可以通过以下命令查看缓存信息：

transformers-cli cache list

然后使用以下命令清除缓存：

transformers-cli cache clear