deepseek 环境搭建

这篇文章是入门文章，没有什么实用价值，仅是将运行deepseek的环境运行起来了。

主要软件

- ollama https://ollama.com/download
- open-webui: 简单的ui页面 http://github.com/open-webui/open-webui
- dify 大型编排平台, https://github.com/langgenius/dify
- docker
- python
  - 3.11: open-webui需要的运行时版本
- 上网: https://iovhm.com/book/books/cee63/page/9872e

ollama可以直接在容器运行

不用担心在容器会对模型和网络的损失，除了网络方面会有损失外，其他方面并没有损失，可以将网络设置为

相反，如果在主机直接运行，各种版本的依赖，新版本的升级，环境更加容易出问题。

```
version: "3"
services:
  ollama:
    image: harbor.iovhm.com/hub/ollama/ollama:0.5.12
    container_name: ollama
    restart: always
    privileged: true
    ports:
      - "11434:11434"
    volumes:
      - ./ollama:/root/.ollama
    networks:
      - vpclub-bridge
```

安装模型运行平台ollama

https://ollama.com/download

模型备份

每次下载模型都需要很久，可以将模型备份出来

C:\Users\Admin.ollama\models ，将模型复制出来，复制到新的机器对应的目录（未验证）

下载和运行模型

```
# 显示所有命令行参数
ollama

# 所有的命令行

serve       Start ollama
create      Create a model from a Modelfile
show        Show information for a model
run         Run a model
stop        Stop a running model
pull        Pull a model from a registry
push        Push a model to a registry
list        List models
ps          List running models
cp          Copy a model
rm          Remove a model
help        Help about any command
```

```
# 运行服务，如果服务没有运行，可以用这个启动服务，非必须，一般安装完成后都会自动运行
ollama serve

# 验证安装
ollama -v


# 下载模型
ollama pull deepseek-r1:1.5b
```

```
# 运行模型
# 如果本地没有这个模型，则会自动下载并运行
# 运行后，此时会出现一个对话窗口，可以进行输入文字进行对话
# 键入/bye 或者ctrl+d可以退出对话窗口，模型在后台运行
# 退出对话框后，如果需要再次进入对话框，可以在此run模型
ollama run deepseek-r1:1.5b

# 显示已经安装的模型列表
ollama list

# 显示所有在运行的模型列表
ollama ps

# 显示模型信息
ollama show deepseek-r1:1.5b
```

使用API对话

ollama默认修改[1]行[B43]端口，可以使用命令行或者postmain测试

这种方式太简陋，要实现的内容太多，不推荐

```
curl -X POST -H "Content-Type: application/json" \
    -d '{"model": "deepseek-r1:1.5b", "prompt": "你好，世界！"}' \
    http://localhost:11434/api/generate
```

使用open-webui

这是一个简单的兼容多个模型的、兼容openai接口调用方式的可视化界面

https://github.com/open-webui/open-webui

https://docs.openwebui.com/

可以直接使用docker运行，docker的安装方式自行百度，优先推荐使用docker安装

```
# GPU版
```

```
docker run -d -p 3000:8080 --gpus all --add-host=host.docker.internal:host-gateway -v open-
webui:/app/backend/data --name open-webui --restart always ghcr.io/open-webui/open-webui:cuda

# CPU
docker run -d -p 3000:8080 --add-host=host.docker.internal:host-gateway -v open-
webui:/app/backend/data --name open-webui --restart always ghcr.io/open-webui/open-webui:main

# pip
pip install open-webui -i https://pypi.tuna.tsinghua.edu.cn/simple

# python
# http://localhost:8080
open-webui serve
```

https://iovhm.com/book/books/bbcbf/page/deepseek-docker

#27
3 2025 03:32:31
9 2025 01:35:27