

docker deepseek open-

ollama ollama

docker-compose.yml

ollama open-webui

```
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:
      - vpcclub-bridge

# docker-compose --profile open-webui up -d
open-webui:
  # CPU
  image: harbor.iovhm.com/public/open-webui/open-webui:main
  # GPU
  # image: harbor.iovhm.com/public/open-webui/open-webui:main-gpu
  container_name: open-webui
  restart: always
  privileged: true
  ports:
    - "3000:8080"
  volumes:
    - ./open-webui:/app/backend/data
  environment:
```

```

# [ ] ollama [ ]OLLAMA_BASE_URLS[ ]OLLAMA_BASE_URLS
# - OLLAMA_BASE_URL=http://ollama:11434
# [ ] ollama [ ]
- OLLAMA_BASE_URLS=http://ollama:11434
# [ ] openai api [ ]openai[ ]
- ENABLE_OPENAI_API=false
# [ ] Arena Model[ ]
- ENABLE_EVALUATION_ARENA_MODELS=false
# [ ]
- ENABLE_COMMUNITY_SHARING=false
# [ ]Internet[ ]( [ ])
# - HF_HUB_OFFLINE=true
# [ ]( [ ])
# - RAG_EMBEDDING_ENGINE=ollama
# [ ]( [ ])
# - RAG_EMBEDDING_MODEL=nomic-embed-text:latest
networks:
- vpclub-bridge

```

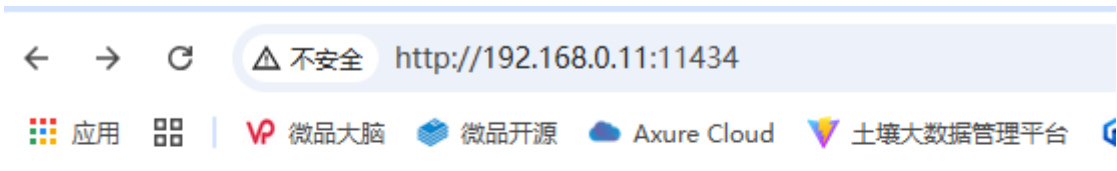
```

networks:
  vpclub-bridge:
    external:
      name: vpclub-bridge

```

[]

[] <http://ip:11434> , [] Ollama is running , [] ollama []



Ollama is running

[] open-web <http://ip:3000> , [] open-webui []

点燃好奇心 无论你在哪里



开始使用

open-webui

- open-webui

openai


```
# open-webui

# openai api
- ENABLE_OPENAI_API=false

# RAG
- RAG_EMBEDDING_ENGINE=ollama

# 
- RAG_EMBEDDING_MODEL=nomic-embed-text:latest
```

openai

☰ **oi** 新对话 

🔗 工作空间

🔍 搜索

∨ 对话

用户 竞技场评估 函数 设置

⚙️ 通用

🔗 外部连接

📦 模型

📄 竞技场评估

📄 文档

🌐 联网搜索

📄 代码执行

🗨️ 界面

🔊 语音

🖼️ 图像

📡 Pipeline

🗄️ 数据库

关闭openai



OpenAI API

Ollama API



管理Ollama API连接

配置ollama地址



http://ollama:11434



访问 Ollama 时遇到问题? [点击这里获取帮助。](#)

直接连接



直接连接功能允许用户连接至其自有的、兼容 OpenAI 的 API 端点。



新对话 工作空间 搜索 对话

用户 竞技场评估 函数 **设置**

通用

Content Extraction Engine 默认

PDF 图像处理 (使用 OCR)

Bypass Embedding and Retrieval

文本分切器 默认 (字符)

块大小 (Chunk Size) 1000 块重叠 (Chunk Overlap) 100

Embedding

语义向量模型引擎 Ollama

http://ollama:11434 API 密钥

语义向量模型 nomic-embed-text

警告: 如果您修改了语义向量模型, 则需要重新导入所有文档。

嵌入层批处理大小 (Embedding Batch Size) 1

完整上下文模式

混合搜索

Retrieval

Top K 3

RAG 提示词模板

```
### Task:
Respond to the user query using the provided context, incorporating inline citations in the format [source_id] **only when the <source_id> tag is explicitly provided** in the context.

### Guidelines:
- If you don't know the answer, clearly state that.
- If uncertain, ask the user for clarification.
- Respond in the same language as the user's query.
- If the context is unreadable or of poor quality, inform the user and provide the
```

保存

设置 已归档对话 AI 对话游乐场 **管理员面板** 登出 当前在线用户: 1 admin

ollama

```
# ollama
docker exec -it ollama /bin/bash

# ollama
ollama -v

# ollama
ollama run deepseek-r1:7b

# ollama
ollama pull nomic-embed-text
```



☰ oi 新对话

🔧 工作空间

🔍 搜索

⌵ 对话

⚙️ 设置

📁 已归档对话

🎮 AI 对话游乐场

👤 管理员面板

🚪 登出

● 当前在线用户: 1

A admin

用户 竞技场评估 函数 **设置**

- ⚙️ 通用
- 🔗 外部连接
- 📦 模型**
- 🏆 竞技场评估
- 📄 文档
- 🌐 联网搜索
- 📄 代码执行
- 🖥️ 界面
- 🔊 语音
- 🖼️ 图像
- 🔗 Pipeline
- 🗄️ 数据库

模型 5

🔍 搜索模型

oi	bge-m3:latest bge-m3:latest...		<input checked="" type="checkbox"/>
oi	deepseek-r1:1.5b deepseek-r1:1.5b...		<input checked="" type="checkbox"/>
oi	deepseek-r1:latest deepseek-r1:latest...		<input checked="" type="checkbox"/>
oi	nomiic-embed-text:latest nomiic-embed-text:latest...		<input checked="" type="checkbox"/>
oi	qllama/bge-reranker-v2-m3:latest qllama/bge-reranker-v2-m3:latest...		<input checked="" type="checkbox"/>

导入预设 导出预设



☰ oi 新对话

🔧 工作空间

🔍 搜索

⌵ 对话

模型 **知识库** 提示词 工具

知识库 1

🔍 搜索知识



文件集 ...

威海智慧谷
威海智慧谷

由 Admin 提供 已更新 11 小时前

① 在输入框中输入"#号来加载你需要的知识库内容。

markdown ### [Progress bar with 20 empty slots]

COLLECTION 威海智慧谷

威海智慧谷

FILE 威海智慧谷.md

威海智慧谷 - 威海智慧谷

#

+

代码解释器

🎤

⬆️

Open WebUI · v0.5.18

deepseekdeepseekB

☰ 新对话

✎

deepseek-r1:1.5b +

⋮

🟡

🔗 工作空间

🔍 搜索

∨ 对话

今天

宫廷玉液酒 🌈 @@ 网页流芳 ...

宫廷玉液酒

🟡 deepseek-r1:1.5b Today at 23:46

思考用时 8 秒 ∨

“宫廷玉液酒”是一个源自中国传统文化的词汇,同时也因1996年春晚小品《打工奇遇》而广为人知。在1996年春晚的小品中,赵丽蓉和巩汉林等演员表演了一个关于“宫廷玉液酒”的搞笑片段。这一设定讽刺了商家的虚假宣传和哄抬物价的行为。小品中的经典台词“宫廷玉液酒,一百八一杯”成为了广为流传的网络梗。随着小品的流行,“宫廷玉液酒”在网络上成为了一个经典的暗号和调侃用语。它常被用来测试对方是否了解中国传统文化或春晚节目,甚至被用作网络社交中的“接头暗号”。

1 威海智慧谷.txt 2 宫廷玉液酒.txt

✎ 🗑️ 🔊 🌐 📌 📄 🔄

输入消息

+

🌐 联网搜索

📄 代码解释器

🎤

⬆️

🟡 董列涛

□□ #24

□□□ □□ 3 □ 2025 10:09:39

□□□ □□ 18 □ 2025 03:45:38