

Ollama Open-Webui

Notes only for now:

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

<https://docs.openwebui.com/getting-started/env-configuration/#general>

Nvidia GPU, linux:

```
mkdir /opt/ollama
mkdir /opt/open-webui
```

[docker-ollama.yml](#)

```
services:
  ollama:
    image: ollama/ollama:latest
    container_name: ollama
    volumes:
      - /opt/ollama:/root/.ollama
    ports:
      - 11434:11434
    #runtime: nvidia
    deploy:
      resources:
        reservations:
          devices:
            - driver: nvidia
              device_ids: ['0']
              capabilities: [gpu]
```

[docker-openwebui.yml](#)

```
services:
  open-webui:
    image: ghcr.io/open-webui/open-webui:main
    container_name: open-webui
    deploy:
      resources:
        reservations:
          devices:
            - driver: nvidia
              device_ids: ['0']
              capabilities: [gpu]
    ports:
      - 3000:8080
    volumes:
      - /opt/open-webui:/app/backend/data
```

```
restart: always
extra_hosts:
  host.docker.internal: host-gateway
environment:
  - WEBUI_NAME="CustomGPT"
```

[docker-openedai-speech.yml](#)

```
services:
  openedai-speech:
    image: ghcr.io/matatonic/openedai-speech
    container_name: openedai-speech
    ports:
      - "8060:8000"
    volumes:
      - /opt/openedai-speech/tts-voices:/app/voices
      - /opt/openedai-speech/tts-config:/app/config
    restart: unless-stopped
    deploy:
      resources:
        reservations:
          devices:
            - driver: nvidia
              device_ids: ['0']
              capabilities: [gpu]
    environment:
      - TZ=Europe/London
      - TTS_HOME=voices
      - HF_HOME=voices
      # - PRELOAD_MODEL=xtts
      # - PRELOAD_MODEL=xtts_v2.0.2
      # - PRELOAD_MODEL=parler-tts/parler_tts_mini_v0.1
```

In open-webui under settings → audio

1. set to openai
2. API Base URL: <http://host.docker.internal:8060/v1>
3. API Key: sk-111111111 (note: this is a dummy API key, no key required)
4. Under TTS Voice within the same audio settings menu in the admin panel, you can set the TTS Model to use from the following choices below that openedai-speech supports. The voices of these models are optimized for the English language.
tts-1 or tts-1-hd: alloy, echo, echo-alt, fable, onyx, nova, and shimmer (tts-1-hd is configurable; uses OpenAI samples by default)

AMD GPU on Windows:

[docker-ollama.yml](#)

```
name: ollama
```

```

services:
  ollama:
    image: ollama/ollama:rocm
    container_name: ollama
    volumes:
      - /p/Docker_Volumes/ollama:/root/.ollama
    ports:
      - 11434:11434
    deploy:
      resources:
        reservations:
          devices:
            - capabilities: [gpu]

```

[docker-openwebui.yml](#)

```

name: webui
services:
  open-webui:
    image: ghcr.io/open-webui/open-webui:main
    container_name: open-webui
    deploy:
      resources:
        reservations:
          devices:
            - capabilities: [gpu]
    ports:
      - 3000:8080
    volumes:
      - /p/Docker_Volumes/openwebui:/app/backend/data
    restart: always
    extra_hosts:
      host.docker.internal: host-gateway
    environment:
      - WEBUI_AUTH=false

```

docker install - WSL2 backend cmd line

```

docker compose -f docker-openwebui.yml up -d
docker compose -f docker-ollama.yml up -d

```

mkdir p/Docker_Volumes = P:\Docker_Volumes

From:

<http://wuff.dyndns.org/> - **Wulf's Various Things**

Permanent link:

<http://wuff.dyndns.org/doku.php?id=ai:ollama-openwebui&rev=1719488606>



Last update: **2024/06/27 12:43**