

Ollama avec Open WebUI

1. [Installation de la boîte à outils NVIDIA Container](#)
2. Crée un volume pour le conteneur :

```
docker volume create ollama
docker volume create open-webui
```

```
services:
  ollama:
    container_name: ollama
    image: ollama/ollama:latest
    restart: 'unless-stopped'
    ports:
      - '11434:11434'
    volumes:
      - 'ollama:/root/.ollama'
    environment:
      TZ: 'Europe/Zurich'
      PUID: '1000'
      PGID: '1000'
      OLLAMA_HOST: '0.0.0.0'
      NVIDIA_DRIVER_CAPABILITIES: 'all'
      NVIDIA_VISIBLE_DEVICES: 'all'
    deploy:
      resources:
        limits:
          cpus: '2.00'
          memory: '4096M'
        reservations:
          cpus: '1.00'
          memory: '4096M'
        devices:
          - driver: nvidia
            count: all
            capabilities: [gpu]
  open-webui:
```

```
container_name: open-webui
image: ghcr.io/open-webui/open-webui:${WEBUI_DOCKER_TAG-main}
restart: 'unless-stopped'
volumes:
  - 'open-webui:/app/backend/data'
volumes:
  ollama:
  open-webui:
    external: true
```

Pour voir l'utilisation en temps réelle sur l'hôte :

```
nvidia-smi dmon
```

Révision #5

Créé 2025-04-02 22:42:52 CEST par Philippe Favre

Mis à jour 2025-05-22 01:57:23 CEST par Philippe Favre