

Docker deployment

```
services:
  ollama:
    image: ollama/ollama:latest
    container_name: ollama
#   ports: ["11434:11434"]
    volumes:
      - ./ollama_data:/root/.ollama
    restart: unless-stopped

# Uncomment below to run on GPU support
deploy:
  resources:
    reservations:
      devices:
        - driver: ${OLLAMA_GPU_DRIVER-nvidia}
          count: ${OLLAMA_GPU_COUNT-1}
          capabilities:
            - gpu
  networks:
    - ai-network
  labels:
    - "traefik.enable=false" # Explicitly disable for Ollama

litellm-db:
  image: postgres:16-alpine
  container_name: litellm-db
  restart: unless-stopped
  environment:
    POSTGRES_USER: litellm
    POSTGRES_PASSWORD: ${LITELLM_DB_PASSWORD}
    POSTGRES_DB: litellm
  volumes:
    - ./litellm_pgdata:/var/lib/postgresql/data
  healthcheck:
```

```

test: ["CMD-SHELL", "pg_isready -U litellm"]
interval: 5s
timeout: 5s
retries: 10
networks:
  - ai-network

litellm:
  image: ghcr.io/berriai/litellm:main-stable
  container_name: litellm
  depends_on:
    ollama:
      condition: service_started
    litellm-db:
      condition: service_healthy
  volumes:
    - ./litellm/config.yaml:/app/config.yaml
  command: ["--config", "/app/config.yaml", "--port", "4000"]
# ports: ["4000:4000"]
  restart: unless-stopped
  networks:
    - ai-network
  labels:
    - "traefik.enable=true"
    - "traefik.http.routers.ollama-api.rule=Host(`litellm.ai.alexwtz.ch`)"
    - "traefik.http.routers.ollama-api.entrypoints=web"
    - "traefik.http.services.ollama-api.loadbalancer.server.port=4000"
    - "traefik.docker.network=ai-network"

environment:
  - LITELLM_MASTER_KEY=${LITELLM_MASTER_KEY}
  - UI_USERNAME=${UI_USERNAME}
  - UI_PASSWORD=${UI_PASSWORD}
  - DATABASE_URL=${DATABASE_URL}
  - STORE_MODEL_IN_DB="True"
  - LITELLM_SALT_KEY=${LITELLM_SALT_KEY}
  healthcheck:
    test: ["CMD-SHELL", "python3 -c \"import urllib.request;
urllib.request.urlopen('http://localhost:4000/health/liveliness')\" || exit 1"]
    interval: 30s

```

```
timeout: 10s
retries: 5
start_period: 40s
```

networks:

ai-network:

driver: bridge

external: true # Use existing network if it exists

Environement file

```
LITELLM_MASTER_KEY=sk-...
UI_USERNAME="admin"
UI_PASSWORD="..."
LITELLM_DB_PASSWORD="..."
LITELLM_SALT_KEY="..." # 32 char salt
DATABASE_URL="postgresql://litellm:${LITELLM_DB_PASSWORD}@litellm-db:5432/litellm"
```

Generate a random sk-...

```
echo "sk-$(openssl rand -hex 32)"
```

Revision #1

Created 16 September 2026 18:02:12 by Admin

Updated 16 September 2026 18:06:11 by Admin