# Настройка Open WebUI + MLX на Apple Silicon

Эта инструкция связывает локальный сервер моделей MLX (Apple) с интерфейсом Open WebUI (Docker).

## Шаг 1: Подготовка (Docker)

Если у вас еще не установлен Docker, вам понадобится **Docker Desktop** для Mac (версия Apple Silicon) или **OrbStack** (более легкая и быстрая альтернатива для Mac).

* *Рекомендация:* Установите [OrbStack](https://orbstack.dev/) — он потребляет меньше памяти и батареи, чем Docker Desktop.

## Шаг 2: Запуск MLX модели в режиме сервера

Чтобы Open WebUI "увидел" модель, она должна работать как API-сервер.

1. Откройте терминал.
2. Установите серверную версию MLX (если нет):  
   pip install mlx-lm
3. Запустите модель (например, Llama-3.1-8B-Abliterated) на порту 8080:  
   mlx\_lm.server --model mlx-community/Llama-3.1-8B-Instruct-Abliterated-4bit --port 8080  
     
   *Оставьте это окно терминала открытым. Теперь ваш Mac раздает модель по адресу http://localhost:8080.*

## Шаг 3: Запуск Open WebUI в контейнере

Теперь запускаем интерфейс и говорим ему, где искать модель.

**Важный нюанс Apple:** Изнутри контейнера localhost — это сам контейнер. Чтобы достучаться до вашего Mac, нужно использовать специальный адрес: host.docker.internal.

Выполните команду в **новом** окне терминала:

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

### Разбор команды (для безопасности):

* -d: Запустить в фоне.
* -p 3000:8080: Интерфейс будет доступен в браузере по порту 3000.
* --add-host=host.docker.internal:host-gateway: **Ключевой момент.** Это разрешает контейнеру видеть порты вашего Mac.
* -v open-webui:/app/backend/data: Создает виртуальный том для сохранения ваших чатов и истории (чтобы они не пропали после перезагрузки).
* ghcr.io/...: Официальный проверенный образ разработчиков.

## Шаг 4: Настройка подключения

1. Откройте браузер и перейдите: http://localhost:3000
2. Зарегистрируйтесь (первый аккаунт станет администратором, данные хранятся локально у вас).
3. Зайдите в **Settings (Настройки)** -> **Connections (Подключения)**.
4. В поле **OpenAI API** (MLX имитирует OpenAI):
   * URL: http://host.docker.internal:8080/v1
   * Key: введите любой символ (например, sk-123), MLX не проверяет ключи, но поле не должно быть пустым.
5. Нажмите "Save".

Теперь в выпадающем списке моделей вы увидите вашу запущенную MLX модель.

## Полезные команды

**Остановить интерфейс:**

docker stop open-webui

**Запустить снова:**

docker start open-webui

**Обновить Open WebUI (когда выйдет новая версия):**

docker run --rm -v /var/run/docker.sock:/var/run/docker.sock containrrr/watchtower --run-once open-webui