

OpenWebUI API 活用ガイド（OpenAI互換API対応）

OpenWebUI は、OpenAI API と互換のエンドポイントを提供することで、**ChatGPTのようなローカルAPI**として利用可能です。

Ollama と連携することで、無料で高性能なチャットAPIが実現できます。

☑ 前提条件

- OpenWebUI が起動していること（例: `http://localhost:3000`）
- Ollama が起動しており、使用したいモデルが pull 済みであること
- OpenWebUI の「設定 → API設定」で **OpenAI API互換モード** を有効化済みであること

☑ 利用できるエンドポイント一覧

エンドポイント	内容
<code>/api/v1/chat/completions</code>	Chat形式のやりとり（ChatGPT互換）
<code>/api/v1/completions</code>	旧OpenAI Completion API互換
<code>/api/v1/embeddings</code>	ベクトル埋め込み生成API

☑ 実行例（curl）

```
curl -X POST http://localhost:3000/api/v1/chat/completions -H "Content-Type: application/json" -H "Authorization: Bearer sk-anything" -d "{
  \"model\": \"hf.co/alfredplpl/gemma-2-2b-jpn-it-gguf\",
  \"messages\": [
    { \"role\": \"user\", \"content\": \"こんにちは\" }
  ]
}"
```

説明

- `"Authorization"` は形式上必要ですが、OpenWebUIでは認証は省略される場合が多い
- `"model"` には **Ollama で登録済みのモデル名** を指定
- `"messages"` は ChatGPT API と同じ形式

☑ 応用例

☑ LangChain や LlamaIndex の **OpenAI** モジュールで使う

```
from langchain.chat_models import ChatOpenAI

llm = ChatOpenAI(
    model_name="hf.co/alfredplpl/gemma-2-2b-jpn-it-gguf",
    openai_api_base="http://localhost:3000/api/v1",
    openai_api_key="sk-local-use-anything"
)
```

→ これで LangChain から **OpenAIと同じコード**で OpenWebUI を使える！

☑ メリットまとめ

項目	内容
無料	OpenAIの従量課金が不要
オフライン	完全ローカル動作が可能
高互換性	既存のOpenAI対応ツールがそのまま使える
LLM切り替え自由	Ollama管理のあらゆるモデルを使える

☑ 注意点

- レスポンス速度や性能は使用するモデルに依存します
- 一部 OpenAI 固有機能（関数呼び出しなど）は未対応の可能性があります