

会話補完

Chat Completion API エンドポイントは、提供されたメッセージのリストに基づいて会話を構築し、その後モデルを使用して応答を生成します。

例題プログラム

PC 側では、OpenAI API にメッセージのリストを渡して対話を構築します。プログラムを実行する前に、以下の **base_url** の IP 部分を実際のデバイスの IP アドレスに変更し、デバイスに対応するモデルパッケージをインストールしてください。モデルパッケージのインストール手順は [モデル一覧](#) セクションを参照してください。

```
from openai import OpenAI
client = OpenAI(
    api_key="sk-",
    base_url="http://192.168.20.186:8000/v1"
)

completion = client.chat.completions.create(
    model="qwen2.5-0.5B-p256-ax630c",
    messages=[
        {"role": "developer", "content": "You are a helpful assistant."},
        {"role": "user", "content": "Hello!"}
    ]
)

print(completion.choices[0].message)
```

リクエストパラメータ

パラメータ名	型	必須	例	説明
messages	array	はい	[{"role": "user", "content": "你好"}]	会話履歴。複数のメッセージで構成され、モデルに応じてテキスト、画像、音声などのモダリティをサポートします。
model	string	はい	qwen2.5-0.5B-p256-ax630c	応答生成に使用するモデル ID。複数のモデルをサポートしており、 モデル一覧 を参照してください。
audio	–	いいえ	–	現在オーディオ出力はサポートされていません。
function_call	–	いいえ	–	現在関数呼び出し機能はサポートされていません。
max_tokens	integer	いいえ	1024	モデルが生成を許可される最大トークン数。これを超えると切り捨てられます。
response_format	object	いいえ	"json_object"	モデル出力のフォーマットを指定します。現在は "json_object" のみサポートされています。

レスポンス例

```
ChatCompletionMessage(content='Hello! How can I assist you today?', refusal=None, role='assistant', ann
```