Gerado a 26/11/2026 07:24

BASE DE CONHECIMENTO · DOCUMENTO PÚBLICO

Usar o SDK da OpenAI

O que você pode automatizar

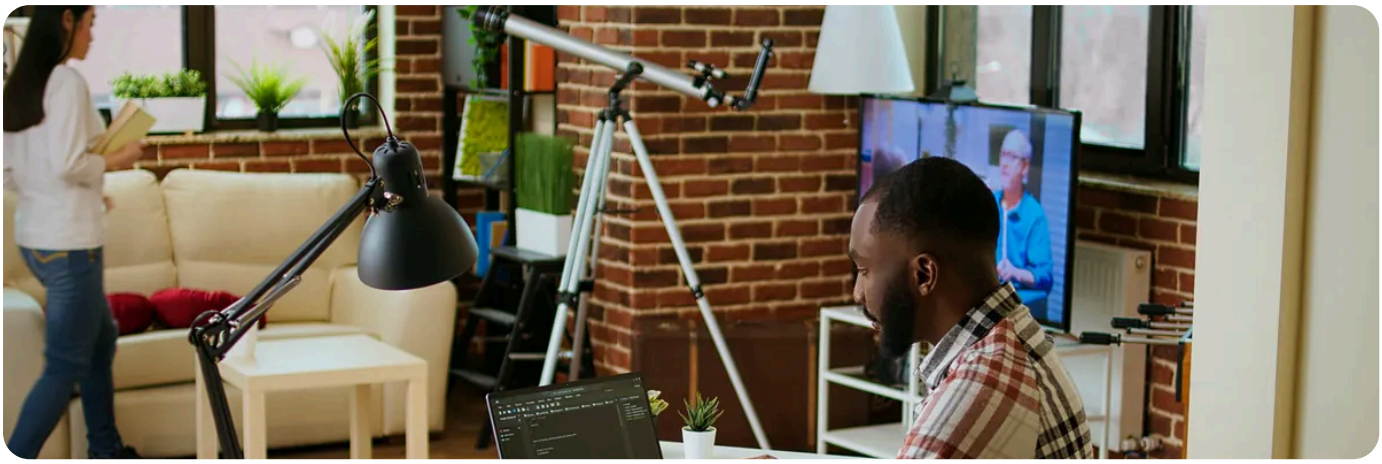
CÓDIGO	VERSÃO	PAÍS · IDIOMA
RLX-BR-KB-8670	**v2026.11.26**	**Brasil · Português (Brasil)**

MozOut Brasil
RLX Holdings (Pty) Ltd. · Grupo RLX Holdings
N.º fiscal (SARS) 943 907 6234 · N.º de registo (CIPC) 2025/667953/07

https://www.my.mozout.com/knowledgebase/8670

Usar o SDK da OpenAI

O que você pode automatizar Criado em **26 de novembro de 2026** 3 min de leitura



A resposta curta: o chat do All in One AI é compatível com a API da OpenAI. Um programa que já usa o SDK da OpenAI passa a usar os nossos modelos mudando **só o endereço e a chave**. É por isso que a **revenda de IA** serve a quem já tem código feito.

Instalar e apontar o SDK

1. Instale o SDK oficial: `pip install openai` em Python ou `npm install openai` em Node.js.
2. Crie uma chave na **página API do My MozOut**, com `ai:read` e `ai:write`. Use uma `sandbox_` enquanto escreve o programa.
3. Guarde a chave na variável de ambiente `MOZOUT_API_KEY`.
4. Indique ao SDK o endereço `https://api.mozout.com/v1` e a chave.

Em Python:

```
import os
from openai import OpenAI

client = OpenAI(base_url="https://api.mozout.com/v1", api_key=os.environ["MOZOUT_API_KEY"])

completion = client.chat.completions.create(
model="deepseek-ai/deepseek-v4-flash",
messages=[{"role": "user", "content": "Escreve duas frases de boas-vindas para o nosso site."}],
```

```
max_tokens=200,
)
print(completion.choices[0].message.content)
```

Em Node.js:

```
import OpenAI from "openai";

const client = new OpenAI({ baseURL: "https://api.mozout.com/v1", apiKey: process.env.MOZOUT_API_KEY });

const completion = await client.chat.completions.create({
model: "deepseek-ai/deepseek-v4-flash",
messages: [{ role: "user", content: "Escreve duas frases de boas-vindas para o nosso site." }],
max_tokens: 200,
});
console.log(completion.choices[0].message.content);
```

O `model` leva sempre uma barra, como no exemplo. `GET /v1/models` dá a lista toda, com o preço de cada modelo em créditos, e filtra por tipo com `type`.

Streaming, para o texto aparecer aos poucos

Com `stream=True`, a resposta chega por partes, à medida que o modelo escreve. Peça sempre `stream_options={"include_usage": True}`: a contagem de tokens vem no último evento, e sem ela fica cobrada a reserva inteira.

Como se cobra cada pedido

Antes de chamar o modelo, reservamos créditos para a entrada e para o `max_tokens` de saída. Sem `max_tokens`, a reserva conta 4096 tokens de saída. No fim, cobramos os tokens reais e devolvemos o resto ao saldo. Se o saldo não chegar para a reserva, a resposta é `402` e nada é cobrado. Com o Coding Plan, os modelos do plano gastam pontos em vez de créditos.

> **Defina sempre `max_tokens`.**
> Limita o tamanho da resposta e baixa a reserva, o que evita o `402` quando o saldo está baixo.

Agentes, com o mesmo código

Um agente criado no All in One AI usa-se no mesmo chat. Em `model`, envie `agent/` seguido do id do agente. Não envie as instruções nem o conhecimento: o agente junta-os sozinho. Se quer saber até onde um agente vai, veja **o que é um agente de IA**.

O que o SDK não faz

O SDK da OpenAI não tem as operações de imagem, vídeo e voz. Para essas, use pedidos HTTP normais a `POST /v1/media/generations` e consulte o estado em `GET /v1/media/generations/{id}`. As ferramentas de programação também funcionam: o Claude Code usa o formato Messages e o Codex CLI o formato Responses, com a mesma chave.

Os erros, no SDK

Os SDK da OpenAI lêem a mensagem de `error.message` e lançam uma exceção com o estado HTTP. Em Python é `openai.APIStatusError`, com `status_code`.

Estado	Quando	O que fazer
`402 insufficient_credits`	Os créditos não chegam para a reserva.	Compre créditos no All in One AI ou baixe o `max_tokens`.
`403 insufficient_scope`	A chave não tem `ai:read` ou `ai:write`.	Crie uma chave com a permissão certa.
`404 model_not_found`	O modelo não existe ou não é de texto.	Confirme o `id` em `GET /v1/models`.
`429 rate_limited`	Passou o limite de pedidos por minuto.	Espere os segundos de `Retry-After`.
`502 upstream_error`	O serviço de IA não respondeu como esperado.	Tente de novo daqui a pouco.

A lista completa está em **os códigos de erro da nossa API**.

RECOMENDADO

Já tem código com o SDK da OpenAI? Veja a página da revenda de IA e crie a sua chave.

Ver a revenda de IA →

VEJA TAMBÉM

- **A mesma chave da API para os seus clientes** →
- **Testar na Sandbox sem custo** →

- Os códigos de erro da nossa API: o que quer dizer cada um, e quais vale a pena repetir
- O que é um agente de IA, e o que ainda não consegue fazer de forma fiável
- Revenda de IA no Reseller Program

MozOut Brasil

RLX Holdings (Pty) Ltd. · Grupo RLX Holdings

N.º FISCAL (SARS)	943 907 6234
N.º DE REGISTO (CIPC)	2025/667953/07
SEDE	72 York Road, Kensington Joanesburgo 2094, Gauteng, África do Sul
CONTACTOS	info@mozout.com · +27 78 823 3325 · www.mozout.com
HORÁRIO	2ª. à 6ª. Feira - 08 à 16 Horas Sábado & Domingo - Fechado

Versão online deste documento

RLX-BR-KB-8670 · v2026.11.26 · Gerado a 26/11/2026 07:24