interweb

Gerado a 27/11/2026 09:56

BASE DE CONHECIMENTO · DOCUMENTO PÚBLICO

Cursor, Codex CLI, Cline e outras ferramentas: o endereço e o modelo certos em cada uma

All in One AI

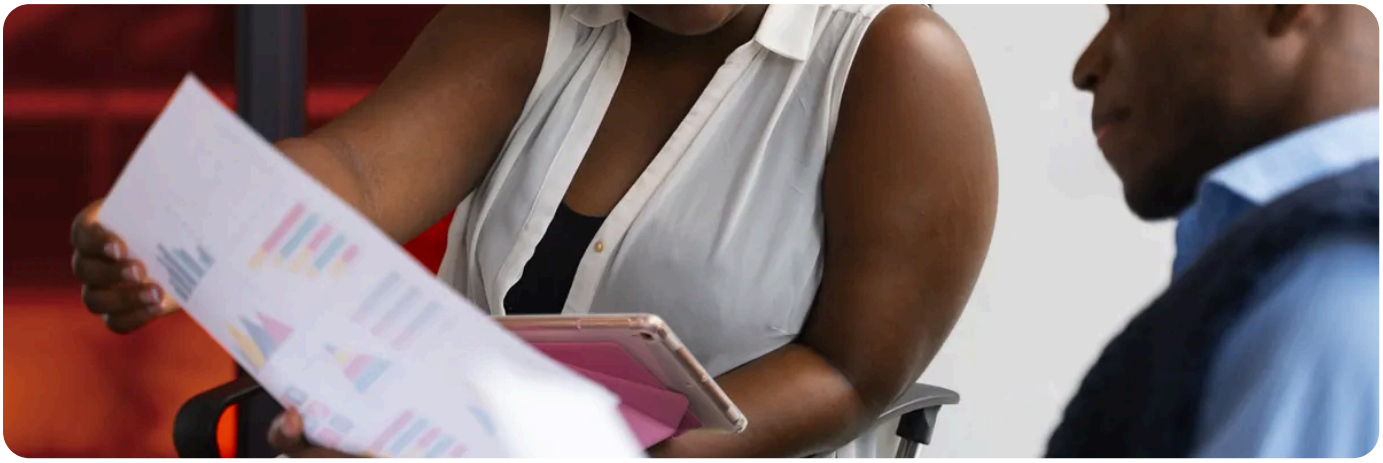
CÓDIGO	VERSÃO	PAÍS · IDIOMA
RLX-MZ-KB-2896	v2026.11.26	Moçambique · Português

Interweb Moçambique
RLX Holdings, S.A. · Grupo RLX Holdings
NUIT 401 191 607 · NUEL 105 045 974 · CAE 62022

https://www.meu.interweb.co.mz/knowledgebase/2896

Cursor, Codex CLI, Cline e outras ferramentas: o endereço e o modelo certos em cada uma

All in One AI Criado em **26 de novembro de 2026** 3 min de leitura



Cada ferramenta de programação fala um formato, e o endereço muda com ele. O erro mais comum é este: `/v1` a mais ou a menos. A chave é sempre a mesma, com `ai:read` e `ai:write`, e paga-se com créditos de IA ou com o Coding Plan. O Claude Code tem um artigo só dele: **usar o Claude Code com o All in One AI**.

Formato	Endereço base	Ferramentas
Anthropic (Messages)	`https://api.interweb.co.mz`	Claude Code
OpenAI (Responses)	`https://api.interweb.co.mz/v1`	Codex CLI
OpenAI (Chat Completions)	`https://api.interweb.co.mz/v1`	Cursor, Cline, Roo Code, OpenCode e Kilo Code

Os modelos recomendados

Todos aceitam ferramentas (function calling) e entram no Coding Plan. Use o id tal como está, com a barra:

Modelo	Contexto	Para quê
`moonshotai/kimi-k2.7-code`	262 144 tokens	Feito para programar: o melhor ponto de partida para o dia-a-dia.
`deepseek-ai/deepseek-v4-pro`	1 048 576 tokens	Raciocínio forte, para mudanças grandes e erros difíceis.
`zai-org/glm-5.2`	1 048 576 tokens	Pensado para agentes: usa bem as ferramentas em tarefas longas.

Modelo	Contexto	Para quê
`qwen/qwen3.8-max`	1 000 000 tokens	Modelo grande, para raciocínio e código de projectos inteiros.
`deepseek-ai/deepseek-v4-flash`	1 048 576 tokens	Rápido e barato, para tarefas de fundo e perguntas curtas.
`minimaxai/minimax-m2.5`	196 608 tokens	Leve e barato, bom para edições simples.

Codex CLI

Crie um fornecedor no `~/.codex/config.toml`:

```
model = "moonshotai/kimi-k2.7-code"
model_provider = "api"

[model_providers.api]
name = "Interweb"
base_url = "https://api.interweb.co.mz/v1"
env_key = "INTERWEB_API_KEY"
wire_api = "responses"
```

O `env_key` é o nome da variável com a chave. O `wire_api = "responses"` é obrigatório. O id do fornecedor pode ter outro nome, mas não `openai`, `ollama` nem `lmstudio`. O fornecedor só se define no ficheiro do utilizador: o de um projecto não o pode mudar. Para outro modelo numa só sessão: `codex -m zai-org/glm-5.2`.

Cursor

1. Abra **Cursor Settings** e depois **Models**.
2. Em **API Keys**, active **OpenAI API Key** e cole a chave `prod_`.
3. Active **Override OpenAI Base URL** e escreva `https://api.interweb.co.mz/v1`.
4. Em **Add Custom Model**, junte o id do modelo, por exemplo `moonshotai/kimi-k2.7-code`, e escolha-o no selector.

Os pedidos passam pelos servidores do Cursor antes de chegarem à nossa API. Segundo a documentação do Cursor, as chaves próprias servem só para os modelos de chat: o Tab continua com os modelos do Cursor. Enquanto o endereço estiver trocado, os pedidos aos modelos da OpenAI no Cursor também vêm para nós. Desligue **Override OpenAI Base URL** para voltar a eles.

Cline e Roo Code

Nos dois, escolha **OpenAI Compatible** em **API Provider**, ponha `https://api.interweb.co.mz/v1` em **Base URL**, a chave em **API Key** e o id do modelo em **Model ID** (no Roo Code, **Model**). Em **Model Configuration**, indique o contexto e o máximo de saída do modelo. O Roo Code só funciona com modelos que aceitam ferramentas.

OpenCode e Kilo Code

Os dois juntam um fornecedor *OpenAI compatible* num ficheiro de configuração, com o endereço `https://api.interweb.co.mz/v1` e a chave numa variável de ambiente. Na extensão do Kilo Code para o VS Code, a lista de modelos vem de `GET /v1/models`. Os ficheiros completos estão na **página de ferramentas de programação das docs**.

Nunca ponha a chave num ficheiro que vai para o repositório.
Use a variável de ambiente. Se uma chave for parar ao Git, revogue-a e veja **o que fazer, pela ordem certa**.

Os agentes de código fazem muitos pedidos seguidos. Se aparecer `429 rate_limited`, espere os segundos de `Retry-After`: as ferramentas costumam repetir sozinhas. Para testar a configuração sem gastar, use uma chave `sandbox_`.

RECOMENDADO

Quer usar a sua ferramenta preferida com os nossos modelos? Veja a página Programar com IA.

Ver Programar com IA →

VEJA TAMBÉM

- **Usar o Claude Code com o All in One AI: as variáveis, o modelo e os erros comuns**
- **Coding Plan: qual escolher, quanto custa e como assinar**
- **O servidor MCP do All in One AI: imagens, vídeos e voz no Claude Code, no Codex e no ChatGPT**
- **Uma chave de API ou um token ficou exposto: o que fazer, pela ordem certa**
- **Programar com IA no All in One AI**

interweb

Interweb Moçambique

RLX Holdings, S.A. · Grupo RLX Holdings

NUIT	401 191 607
NUEL	105 045 974
CAE	62022
SEDE	Bairro 7 de Setembro Vilankulo 1304 · CEP 0411-02, Inhambane, Moçambique
CONTACTOS	info@interweb.co.mz · +258 84 635 1111 · +258 86 635 1111 · www.interweb.co.mz
HORÁRIO	2ª. à 6ª. Feira - 08 à 16 Horas Sábado & Domingo - Fechado

Versão online deste documento

RLX-MZ-KB-2896 · v2026.11.26 · Gerado a 27/11/2026 09:56