interweb

Gerado a 28/11/2026 22:26

BASE DE CONHECIMENTO · DOCUMENTO PÚBLICO

Os modelos recomendados para programar: o contexto e para que serve cada um

Escolher a ferramenta e resolver erros

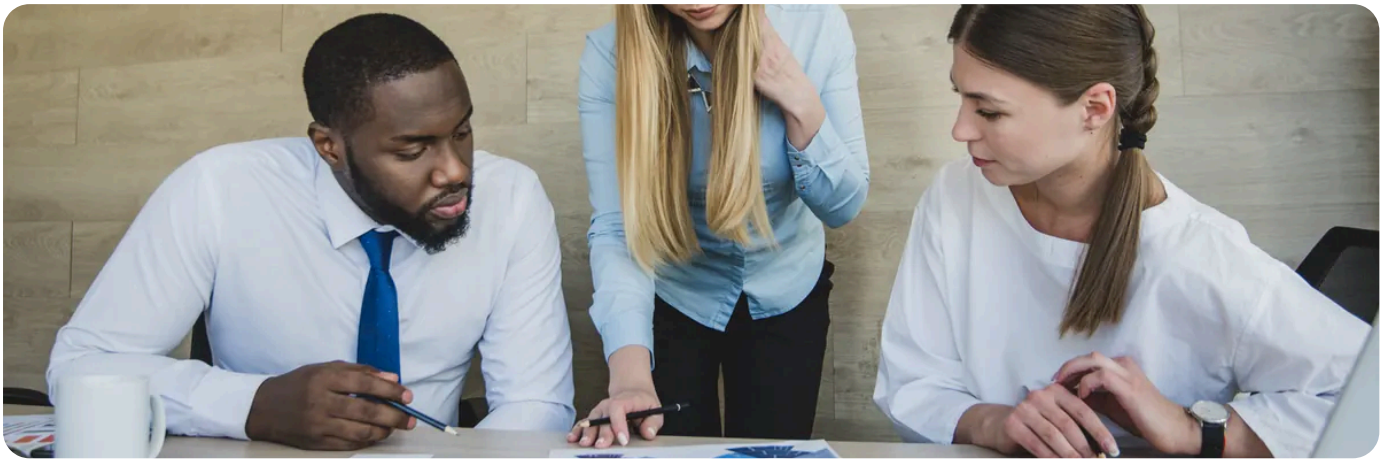
CÓDIGO	VERSÃO	PAÍS · IDIOMA
RLX-MZ-KB-3258	v2026.11.28	Moçambique · Português

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/3258

Os modelos recomendados para programar: o contexto e para que serve cada um

Escolher a ferramenta e resolver erros · Criado em **28 de novembro de 2026** · 3 min de leitura



O catálogo do All in One AI tem muitos modelos de texto, mas nem todos servem para programar. Estes seis estão no catálogo, 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.
`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.
`zai-org/glm-5.3-flash`	1 048 576 tokens	Leve e barato, bom para edições simples.

Qual escolher

1. Para o dia-a-dia, comece pelo `moonshotai/kimi-k2.7-code`: é feito para programar.

2. Para uma mudança grande ou um erro difícil, passe ao `deepseek-ai/deepseek-v4-pro`.

3. Para tarefas longas, em que o agente usa muitas ferramentas seguidas, use o `zai-org/glm-5.2`.

4. Para raciocinar sobre um projecto inteiro, o `qwen/qwen3.8-max`.

5. Para tarefas de fundo e perguntas curtas, o `deepseek-ai/deepseek-v4-flash`. É o que a nossa configuração do Claude Code põe em `ANTHROPIC_DEFAULT_HAIKU_MODEL`.

6. Para edições simples e baratas, o `zai-org/glm-5.3-flash`.

O contexto, e onde o indicar

O contexto (`context_length`) é o máximo de tokens que o modelo lê num pedido, entrada e saída juntas. O `moonshotai/kimi-k2.7-code` lê 262 144 tokens; os outros cinco, cerca de um milhão. Indique na ferramenta o contexto do modelo que escolheu, para ela saber quando a conversa chega ao limite:

Ferramenta	Onde se indica o contexto
Claude Code	`CLAUDE_CODE_MAX_CONTEXT_TOKENS`
Cline	**Model Configuration**, com o máximo de saída
OpenCode e Kilo Code (ficheiro)	`limit.context` de cada modelo
Aider	`max_input_tokens` no `.aider.model.metadata.json`

Um contexto maior não torna o pedido mais barato. Cada pedido paga os tokens de entrada que leva: quanto mais longa a conversa, mais texto vai em cada pedido.

A saída, o preço e o Coding Plan

O máximo de saída de cada modelo está em `max_output_tokens` e o preço em créditos em `pricing`. Reservamos créditos pelo máximo de saída que a ferramenta pede em cada resposta. Por isso os exemplos pedem 16 000 tokens: chega para programar e precisa de menos saldo do que um valor maior. Com um Coding Plan activo, estes seis gastam pontos em vez de créditos, e os modelos mais fortes gastam mais pontos por token. Os modelos que gastam pontos do plano estão na aba **«Plano» do All in One AI**.

Os ids estão na aba «Modelos» do **All in One AI** ou na lista da API. No Linux e no macOS:

```
curl "https://api.interweb.co.mz/v1/models?type=text" -H "Authorization: Bearer $INTERWEB_API_KEY"
```

No PowerShell do Windows:

```
curl.exe "https://api.interweb.co.mz/v1/models?type=text" -H "Authorization: Bearer $env:INTERWEB_API_KEY"
```

Cada modelo vem com o `id` (sempre com a barra), o `context_length` e o preço em `pricing`. Ver a lista não gasta créditos.

Trocar de modelo a meio

No Claude Code, escreva `/model zai-org/glm-5.2`. No Codex, `codex -m zai-org/glm-5.2` abre uma sessão com outro modelo. No OpenCode, use `/models`, no formato fornecedor/modelo, por exemplo `api/zai-org/glm-5.2`. Se passar para o `moonshotai/kimi-k2.7-code`, baixe também o contexto que indicou na ferramenta.

Há mais modelos que aceitam ferramentas: o bloco de modelos do **artigo do OpenCode** traz os 79, do mais caro ao mais barato.

RECOMENDADO

Quer programar com os modelos do All in One AI? Veja a página Programar com IA.

Ver Programar com IA →

VEJA TAMBÉM

- **Porque é que um modelo para programar tem de aceitar ferramentas** →
- **Que ferramenta de programação usar** →
- **Escolher um modelo de texto** →
- **Coding Plan: qual escolher** →
- **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-3258 · v2026.11.28 · Gerado a 28/11/2026 22:26