interweb

Gerado a 28/11/2026 21:59

BASE DE CONHECIMENTO · DOCUMENTO PÚBLICO

Modelo de texto barato ou forte: comparar o preço por milhão de tokens

Playground e modelos de texto

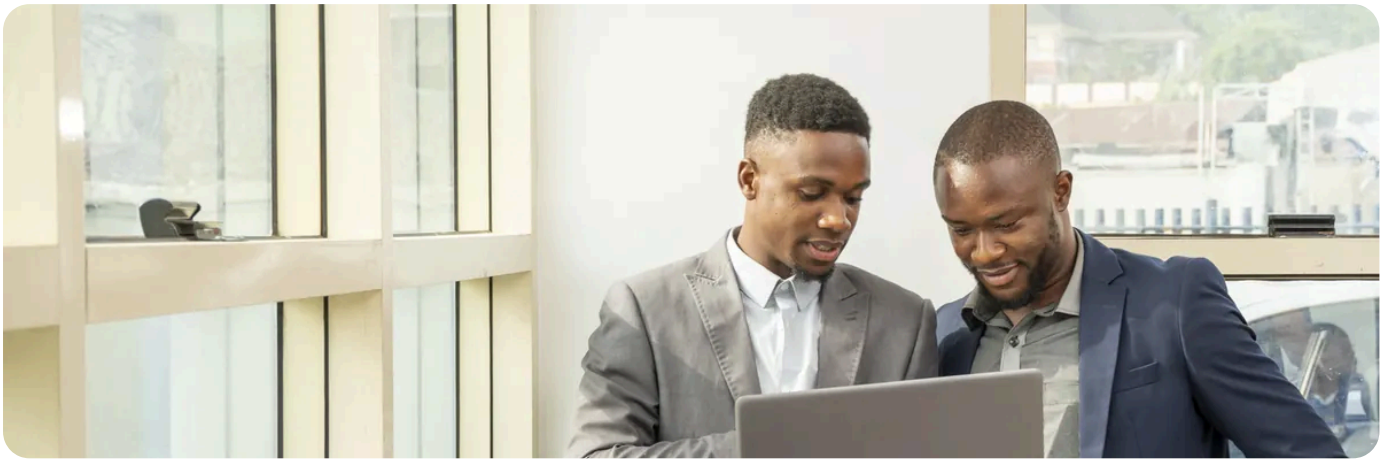
CÓDIGO	VERSÃO	PAÍS · IDIOMA
RLX-MZ-KB-2988	v2026.11.28	Moçambique · Português

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/2988

Modelo de texto barato ou forte: comparar o preço por milhão de tokens

Playground e modelos de texto Criado em **28 de novembro de 2026** 2 min de leitura



Entre um modelo de texto barato e um dos mais fortes, o preço pode mudar muito. Comparar o preço por milhão de tokens e testar os dois com o mesmo pedido mostra se a diferença compensa para o seu trabalho.

Dois modelos lado a lado

Com os preços que a documentação dá como exemplo, em créditos por milhão de tokens:

Modelo	Entrada	Saída	1.000 de entrada e 500 de saída
`deepseek-ai/deepseek-v4-flash`	56	112	0,112, cobrado 0,12
`moonshotai/kimi-k2.6`	380	1.600	0,38 + 0,8 = 1,18

O mesmo pedido custa quase dez vezes mais no segundo. Os preços actuais estão na aba **Modelos**.

Olhe sobretudo para a saída

A saída costuma ser mais cara do que a entrada, e numa resposta longa é ela que pesa. No exemplo, a saída do segundo modelo custa mais de 14 vezes a do primeiro. Num trabalho com respostas curtas e textos longos, como resumos, pesa mais a entrada.

Testar com o mesmo pedido

1. Abra o **Playground** e active **Comparar modelos**.
2. Escolha o modelo principal, o segundo e, se quiser, um terceiro.

3. Escreva a mensagem e veja o custo estimado dos pedidos, somados. Cada modelo é cobrado à parte, ao seu preço.

4. Envie e compare as respostas, lado a lado. Cada uma fica guardada na sua conversa.

Quando vale a pena o mais forte

Os modelos rápidos e baratos servem bem para perguntas curtas, resumos, classificação e tarefas de fundo. Os mais fortes compensam no raciocínio difícil, no código grande e nos textos que têm de sair certos à primeira. A documentação dá `deepseek-ai/deepseek-v4-flash` como rápido e barato, `deepseek-ai/deepseek-v4-pro` como forte no raciocínio e `moonshotai/kimi-k2.7-code` como feito para programar.

Comece pelo modelo mais barato que faz o trabalho e só suba se a diferença se notar. O custo real de cada teste fica no Histórico. Veja **como escolher um modelo de texto**.

RECOMENDADO

Compare dois modelos com o mesmo pedido no Playground.

Ver o All in One AI →

VEJA TAMBÉM

- **Escolher um modelo de texto: preço, contexto e para que serve cada um**
- **Ler o preço de um modelo: créditos por milhão de tokens, por imagem, por segundo ou por geração**
- **Ver quanto vai custar uma mensagem antes de a enviar no Playground**
- **Entrada em cache: porque é que o texto que o modelo já leu gasta menos créditos**
- **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-2988 · v2026.11.28 · Gerado a 28/11/2026 21:59