interweb

Generated on 28/11/2026 21:59

KNOWLEDGE BASE · PUBLIC DOCUMENT

Cheap or strong text model: comparing the price per million tokens

Playground & Text Models

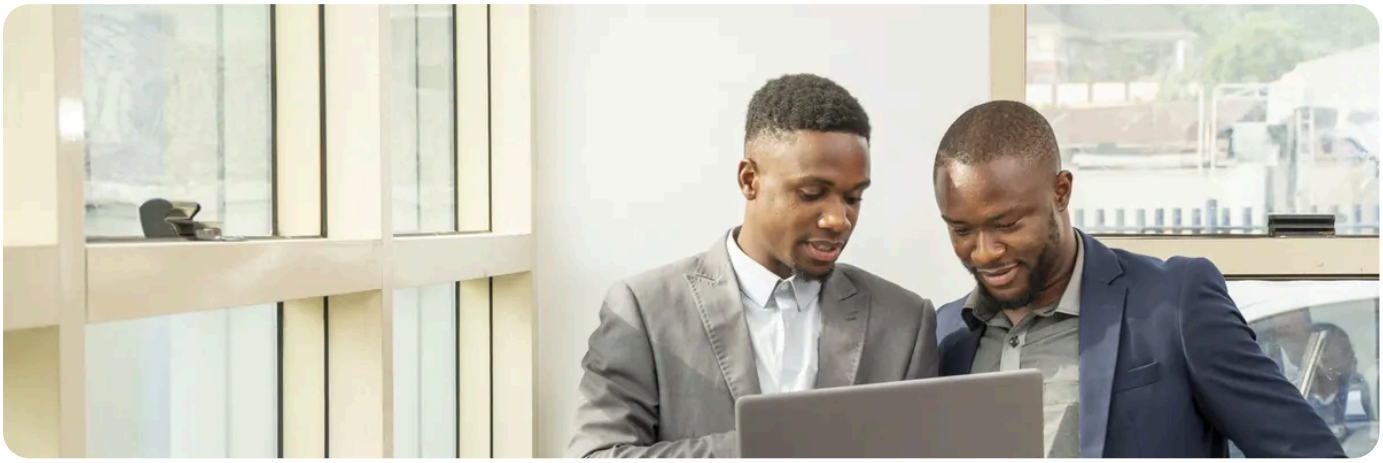
CODE	VERSION	COUNTRY · LANGUAGE
RLX-MZ-KB-2988	v2026.11.28	Mozambique · English

Interweb Mozambique
RLX Holdings, S.A. · Group RLX Holdings
Tax ID (NUIT) 401 191 607 · Legal entity no. (NUEL) 105 045 974 · Economic activity code (CAE) 62022

https://www.meu.interweb.co.mz/knowledgebase/2988

Cheap or strong text model: comparing the price per million tokens

Playground & Text Models Created on **November 28, 2026** 2 min read



Between a cheap text model and one of the strongest, the price can change a lot. Comparing the price per million tokens and testing both with the same request shows whether the difference is worth it for your work.

Two models side by side

With the prices the documentation gives as examples, in credits per million tokens:

Model	Input	Output	1,000 input and 500 output
`deepseek-ai/deepseek-v4-flash`	56	112	0.112, charged 0.12
`moonshotai/kimi-k2.6`	380	1,600	0.38 + 0.8 = 1.18

The same request costs almost ten times more on the second one. Current prices are on the **Models** tab.

Look mostly at the output

Output is usually dearer than input, and in a long answer it is what weighs most. In the example, the second model's output costs more than 14 times the first one's. In work with short answers and long texts, such as summaries, the input weighs more.

Testing with the same request

1. Open the **Playground** and turn on **Compare models**.
2. Choose the main model, the second one and, if you like, a third.

3. Type the message and check the estimated cost of the requests, added up. Each model is charged separately, at its own price.

4. Send it and compare the answers side by side. Each one is saved in its own conversation.

When the stronger model is worth it

Fast, cheap models do well with short questions, summaries, sorting and background tasks. The stronger ones pay off for hard reasoning, large code and texts that must come out right first time. The documentation describes `deepseek-ai/deepseek-v4-flash` as fast and cheap, `deepseek-ai/deepseek-v4-pro` as strong at reasoning and `moonshotai/kimi-k2.7-code` as built for coding.

Start with the cheapest model that does the job and only move up if the difference shows. The real cost of each test is in History. See **how to choose a text model**.

RECOMMENDED

Compare two models with the same request in the Playground.

See All in One AI →

SEE ALSO

- **Choosing a text model: price, context window and what each one is for** →
- **Reading a model's price: credits per million tokens, per image, per second or per generation** →
- **Seeing what a message will cost before you send it in the Playground** →
- **Cached input: why text the model has already read costs fewer credits** →
- **All in One AI** →

interweb

Interweb Mozambique

RLX Holdings, S.A. · Group RLX Holdings

TAX ID (NUIT)	401 191 607
LEGAL ENTITY NO. (NUEL)	105 045 974
ECONOMIC ACTIVITY CODE (CAE)	62022
REGISTERED OFFICE	Bairro 7 de Setembro Vilankulo 1304 · CEP 0411-02, Inhambane, Mozambique
CONTACT	info@interweb.co.mz · +258 84 635 1111 · +258 86 635 1111 · www.interweb.co.mz
OPENING HOURS	Monday to Friday - 08 AM to 04 PM Saturday & Sunday - Closed

Online version of this document

RLX-MZ-KB-2988 · v2026.11.28 · Generated on 28/11/2026 21:59