interweb

Gerado a 29/11/2026 11:26

BASE DE CONHECIMENTO · DOCUMENTO PÚBLICO

Experimentar o mesmo pedido em vários modelos no gateway de IA antes de escolher

Playground e modelos de texto

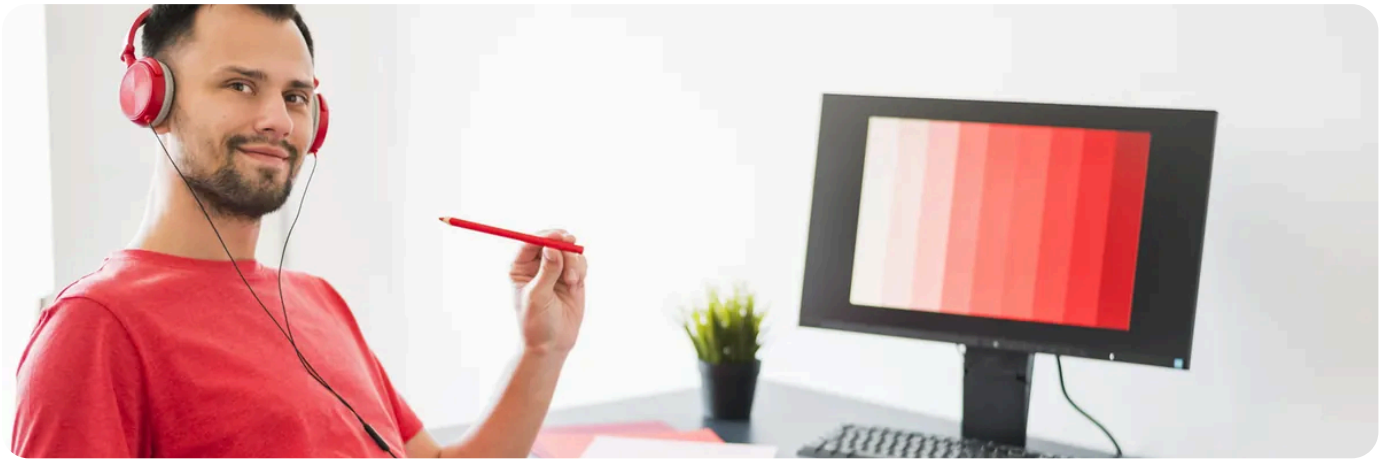
CÓDIGO	VERSÃO	PAÍS · IDIOMA
RLX-MZ-KB-3284	v2026.11.29	Moçambique · Português

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/3284

Experimentar o mesmo pedido em vários modelos no gateway de IA antes de escolher

Playground e modelos de texto Criado em **29 de novembro de 2026** 2 min de leitura



Uma das vantagens de um gateway de IA é pôr vários modelos à prova com o mesmo pedido, a mesma chave e o mesmo código. Antes de escolher o modelo do seu programa, vale a pena ver como três ou quatro respondem ao mesmo texto e quanto gasta cada um.

No Playground, sem escrever código

1. Abra o **All in One AI** e vá ao Playground.
2. Escreva o pedido que o seu programa vai fazer, com um texto real.
3. Escolha um modelo, veja quanto vai custar antes de enviar e envie. Veja **como ver o custo antes de enviar**.
4. Mude de modelo e envie o mesmo pedido. Compare as respostas e o custo de cada uma.

Pela API, com um ciclo

```
import os
from openai import OpenAI

client = OpenAI(base_url="api.interweb.co.mz/v1", api_key=os.environ["INTERWEB_API_KEY"])
pedido = [{"role": "user", "content": "Escreva uma resposta curta a um cliente que pergunta pelo horário de sábado."}]

for modelo in ["deepseek-ai/deepseek-v4-flash", "deepseek-ai/deepseek-v4-pro", "zai-org/glm-5.2"]:
```

```
    r = client.chat.completions.create(model=modelo, messages=pedido, max_tokens=300)
        print("==", modelo, r.usage.prompt_tokens, "+", r.usage.completion_tokens,
"tokens")
    print(r.choices[0].message.content)
```

O custo de cada pedido fica na aba Histórico, linha a linha, com o modelo usado. Veja **a aba Histórico**.

A Sandbox não serve para comparar.
Com uma chave `sandbox_` nenhum modelo é chamado: todos respondem «Resposta de teste (sandbox).». Para comparar respostas reais precisa da chave `prod_` e de créditos. Use textos curtos e `max_tokens` baixo para gastar pouco.

O que comparar

O quê	Porquê
A qualidade da resposta	É o que o seu cliente vai ler. Use textos reais, não exemplos inventados.
Os tokens e o custo	Dois modelos podem dar respostas parecidas a preços muito diferentes.
O tempo de resposta	Meça-o no seu programa: depende do modelo e do tamanho da resposta.
O comprimento	Alguns modelos escrevem mais do que precisa, e a saída também se paga.

RECOMENDADO

Quer comparar modelos? Abra o Playground do All in One AI.

Ver o All in One AI →

VEJA TAMBÉM

- **Conversar no Playground**
- **Modelo barato ou forte**
- **Mudar de modelo sem mexer no código**
- **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-3284 · v2026.11.29 · Gerado a 29/11/2026 11:26