interweb

Gerado a 27/11/2026 09:56

BASE DE CONHECIMENTO · DOCUMENTO PÚBLICO

# Quanto custa um agente de IA, e como o testar sem gastar

All in One AI

| CÓDIGO | VERSÃO | PAÍS · IDIOMA |
| --- | --- | --- |
| RLX-MZ-KB-2892 | 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/2892

# Quanto custa um agente de IA, e como o testar sem gastar

All in One AI Criado em **26 de novembro de 2026** 2 min de leitura



Um agente não tem preço próprio. Paga-se o modelo que ele usa, em créditos de IA, ao preço que está em `GET /v1/models`. Cada volta ao modelo gasta créditos como um pedido de chat.

## O que entra em cada resposta

| Parte | Como conta |
|---|---|
| **Instruções e conhecimento** | Tokens de entrada, em todos os pedidos |
| **As mensagens da conversa** | Tokens de entrada |
| **A resposta** | Tokens de saída |
| **Cada uso de uma ferramenta** | Mais uma chamada ao modelo, que leva também o texto das anteriores |

Por isso, uma resposta que usa duas ferramentas faz três chamadas ao modelo e custa mais do que uma sem ferramentas. Para perceber os tokens, veja **tokens explicados**.

## A reserva e o acerto

1. Antes de chamar o modelo, reservamos créditos para a entrada e para o `max_tokens` de saída. Sem `max_tokens`, a reserva conta 4096 tokens de saída.
2. Se o saldo não chegar para a reserva, a resposta é `402 insufficient_credits` e nada é cobrado.
3. No fim, cobramos os tokens reais e devolvemos o resto da reserva ao saldo. O mínimo por pedido é 0,01 créditos.

## Testar sem gastar, e testar a sério

| Forma | Gasta créditos? | Para que serve |
|---|---|---|
| **Chave** `sandbox_` | Não | Escrever e testar o seu programa. O agente responde no mesmo formato, mas não chama o modelo nem as ferramentas: o texto é sempre «Resposta de teste (sandbox).» |
| **Botão «Experimentar»** | Sim | Ver as respostas reais do agente, com o custo de cada uma. Os pedidos aparecem no histórico como «Playground do All in One AI». |

## Como baixar a conta

1. Escreva instruções e conhecimento curtos. Vão em todas as respostas. Veja **o que juntar ao conhecimento**.
2. Defina o `max_tokens`. Limita o tamanho da resposta e baixa a reserva, o que evita o `402` quando o saldo está baixo.
3. Active só as ferramentas de que o agente precisa.
4. Para perguntas simples, escolha um modelo mais barato. O preço de cada um está em `GET /v1/models`, no campo `pricing`.

**Antes de abrir o agente a muita gente,**
confirme o saldo com `GET /v1/ai/balance` ou no **All in One AI**, e faça meia dúzia de perguntas no «Experimentar» para ver quanto custa uma conversa normal.

RECOMENDADO

Quer um agente com custos que controla? Veja a página Agentes de IA.

Ver os agentes de IA →

VEJA TAMBÉM

**Criar um agente de IA no All in One AI, passo a passo** →

**O conhecimento do agente: que textos juntar e como manter cada resposta barata**

**As ferramentas do agente: ler páginas da internet e saber a data de hoje**

**Tokens explicados: como se contam as APIs de modelos de IA**

**Agentes de 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<br>Vilankulo 1304 · CEP 0411-02, Inhambane, Moçambique |
| CONTACTOS | info@interweb.co.mz · +258 84 635 1111 · +258 86 635 1111 ·<br>www.interweb.co.mz |
| HORÁRIO | 2ª. à 6ª. Feira - 08 à 16 Horas<br>Sábado & Domingo - Fechado |

Versão online deste documento

RLX-MZ-KB-2892 · v2026.11.26 · Gerado a 27/11/2026 09:56