interweb

Gerado a 28/11/2026 22:08

BASE DE CONHECIMENTO · DOCUMENTO PÚBLICO

Os limites do agente: 5 rondas de ferramentas, e max_tokens e temperature que só podem descer

Agentes pela API

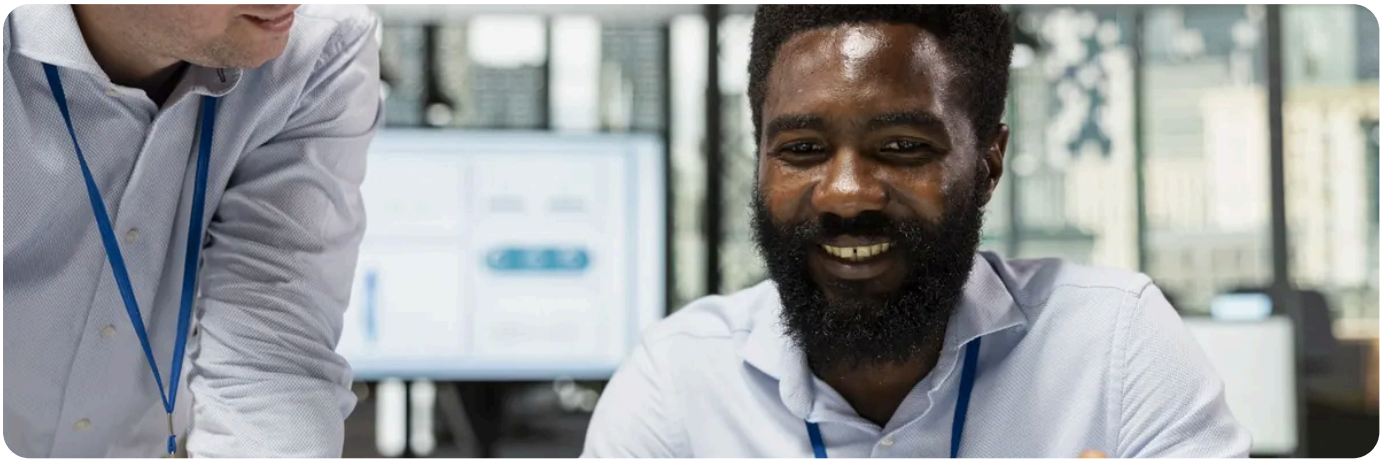
CÓDIGO	VERSÃO	PAÍS · IDIOMA
RLX-MZ-KB-3056	v2026.11.28	Moçambique · Português

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/3056

Os limites do agente: 5 rondas de ferramentas, e max_tokens e temperature que só podem descer

Agentes pela API Criado em **28 de novembro de 2026** 2 min de leitura



Um agente leva as suas definições em cada pedido. O seu programa pode baixar algumas, mas não as pode ultrapassar. Estes são os limites, e o que fazer quando um deles atrapalha.

Os limites de cada pedido

Limite	Valor
Voltas de ferramentas por pedido	No máximo 5
Pesquisas na web por resposta	No máximo 3, a 0,5 créditos cada
Uma página lida com `web_fetch`	Até 20.000 caracteres de texto; espera no máximo 10 segundos e lê até 1 MB
Modelo	Sempre o do agente
`max_tokens` e `temperature`	O pedido pode baixar os do agente, nunca subir

Os limites do agente

O quê	Limite
Agentes por conta	50
Instruções	20.000 caracteres
Conhecimento	200 KB por agente; cada ficheiro até 10 MB
Temperatura	De 0 a 2

O quê	Limite
«Máximo de tokens na resposta»	De 16 a 32.768

max_tokens e temperature só descem

Imagine um agente com 1024 tokens de máximo e temperatura 0,7, os valores com que a aba «Agentes» abre um agente novo:

O pedido envia	O agente usa
`"max_tokens": 300`	300
`"max_tokens": 4000`	1024, o do agente
`"temperature": 0.2`	0,2
`"temperature": 1.5`	0,7, a do agente
Nenhum dos dois	Os do agente

Para respostas maiores, suba o máximo nas Definições do agente, no **All in One AI**. Um máximo maior também sobe a reserva de créditos de cada pedido. A reserva está explicada em **quanto custa um pedido de chat**.

As 5 voltas

Cada uso de uma ferramenta é uma volta ao modelo, e em cada pedido o agente faz no máximo 5. Para perguntas que precisariam de muitas páginas, divida o trabalho em pedidos mais pequenos. Cada volta também conta no custo e no tempo de resposta: veja **a resposta do agente com streaming**.

Além disto, cada chave tem um limite de pedidos por minuto, como no resto da API. Está em **limites de pedidos da API**.

RECOMENDADO

Quer saber até onde vai o seu agente? Veja o All in One AI.

Ver o All in One AI →

VEJA TAMBÉM

Usar o seu agente pela API: agent/ no campo model →

Quanto custa um pedido de chat pela API: a reserva, o max_tokens e o acerto

Limites de pedidos da API: por chave, por conta e por tarefa, os cabeçalhos e o que fazer num 429

A resposta do agente com streaming

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-3056 · v2026.11.28 · Gerado a 28/11/2026 22:08