Mozout

Generated on 28/11/2026 21:54

KNOWLEDGE BASE · PUBLIC DOCUMENT

429 rate_limited in a coding tool: why agents reach the limit and what to do

Choosing a Tool & Fixing Errors

CODE	VERSION	COUNTRY · LANGUAGE
RLX-INT-KB-9870	v2026.11.28	International · English

MozOut International
RLX Holdings (Pty) Ltd. · Group RLX Holdings
Tax no. (SARS) 943 907 6234 · Registration no. (CIPC) 2025/667953/07

https://www.my.mozout.com/knowledgebase/9870

429 rate_limited in a coding tool: why agents reach the limit and what to do

Choosing a Tool & Fixing Errors Created on **November 28, 2026** 2 min read



`429` with the code `rate_limited` means you went over the limit of requests per minute. It is not a lack of credits or points, and it clears when the minute ends. Coding agents get there faster than an ordinary program, because they make many requests in a row.

The limits

Who is asking	Limit
Each key	90 requests per minute
The whole account, adding up its keys	300 requests per minute

Reseller accounts get 300 per key and 1,000 for the account. These are today's values and they can change: the request at the end of this page shows yours.

Why agents get there

1. Each tool round is a new request: read a file, edit another, run a command, and back to the model. See **why a coding model uses tools**.

2. Claude Code also uses the background task model (`ANTHROPIC_DEFAULT_HAIKU_MODEL`) and asks for a token count to manage the context. The count costs nothing, but it is a request like any other.

3. Two windows, two agents or one agent with sub-agents, on the same key, share the same 90 requests.

What to do

1. Wait the number of seconds in `Retry-After`. The tools usually retry on their own: Claude Code retries failed requests up to 10 times by default (`CLAUDE_CODE_MAX_RETRIES`) and Codex up to 4 (`request_max_retries`, in the provider in `config.toml`).
2. Do not run several sessions at once on the same key. If you need several, give each tool its own key: each one has its 90 requests, within the account's 300.
3. See how many requests are left with the request below. It spends no credits.
4. If your work needs more, **open a support ticket**: our team can give an account its own limits.

On Linux and macOS:

```
curl "https://api.mozout.com/v1/limits" -H "Authorization: Bearer $MOZOUT_API_KEY"
```

In Windows PowerShell:

```
curl.exe "https://api.mozout.com/v1/limits" -H "Authorization: Bearer $env:MOZOUT_API_KEY"
```

The answer has `requests.limit`, `requests.remaining` and `requests.reset` (the end of the minute, in Unix seconds). The same numbers come in the `X-RateLimit-Limit`, `X-RateLimit-Remaining` and `X-RateLimit-Reset` headers of every answer. In Claude Code the error arrives as `rate_limit_error`; in Codex, with the `code` `rate_limited`.

Not to be confused with plan_limit_reached

There is another `429`, with the code `plan_limit_reached`: the Coding Plan points ran out and the request cannot be paid with credits. Waiting a few seconds does not help. Wait for `resets_at`, turn on Use credits when the plan runs out in the **Plan tab of All in One AI** or buy credits. The points are explained in **Coding Plan: which plan to choose**.

RECOMMENDED

Want to code with All in One AI models? Visit the Coding with AI page.

See Coding with AI →

SEE ALSO

- API request limits
- Why a coding model must accept tools
- Coding Plan: which plan to choose
- Errors when setting up a tool
- Coding with AI in All in One AI

Mozout

MozOut International

RLX Holdings (Pty) Ltd. · Group RLX Holdings

TAX NO. (SARS)	943 907 6234
REGISTRATION NO. (CIPC)	2025/667953/07
REGISTERED OFFICE	1111 Olive Street Saint Louis 63101, Missouri, United States
CONTACT	info@mozout.com · +27 78 823 3325 · www.mozout.com
OPENING HOURS	Monday to Friday - 08 AM to 04 PM Saturday & Sunday - Closed

Online version of this document

RLX-INT-KB-9870 · v2026.11.28 · Generated on 28/11/2026 21:54