Mozout

Generated on 28/11/2026 21:01

KNOWLEDGE BASE · PUBLIC DOCUMENT

The 402 insufficient_credits error: what required and balance mean, and how to fix it

AI Credits & Costs

CODE	VERSION	COUNTRY · LANGUAGE
RLX-INT-KB-8994	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/8994

The 402 insufficient_credits error: what required and balance mean, and how to fix it

AI Credits & Costs Created on **November 28, 2026** 2 min read



A `402` with `insufficient_credits` means your AI credits do not cover the reservation for the request. **Nothing was charged.** It almost always has two fixes: buy credits or ask for a shorter answer.

What the error contains

```
{
  "error": {
    "code": "insufficient_credits",
    "message": "Not enough credits for this request. Nothing was charged.",
    "correlation_id": "0f8fad5b-d9cb-469f-a165-70867728950e",
    "details": { "required": 46.12, "balance": 12.5 }
  }
}
```

Field	What it means
`required`	The credits the request needs to reserve before calling the model.
`balance`	The credits the account has right now.

In the example, the request needed 46.12 credits and the account had 12.5.

Why the reservation can be bigger than the cost

Before calling the model, we reserve credits for the input (about one token per 3 characters) and for `max_tokens` of output. Without `max_tokens`, the reservation counts 4,096 output tokens. At the end we charge only what was used and the rest goes back to your balance. That is why a cheap request can be refused if the maximum output is high.

An example, with the price the documentation gives for `moonshotai/kimi-k2.6` (1,600 credits per million output tokens): without `max_tokens`, the output alone reserves 4,096 × 1,600 / 1,000,000 = 6.55 credits. With `max_tokens` at 500, it drops to 0.8. Current prices are on the Models tab.

How to fix it

1. Buy credits on the **Buy credits** tab, or through the API.
2. Lower the maximum output: `max_tokens` in the API, "Maximum tokens in the reply" in the Playground, `--max-tokens` in `aio`, `CLAUDE_CODE_MAX_OUTPUT_TOKENS` in Claude Code or **Max Output Tokens** in the tools that have it.
3. Send less text: a long conversation or a big document all count as input.
4. Or choose a cheaper model.

In the Playground, the same case shows as "This request holds X credits and your account has Y. Buy credits and try again." For images, video and voice, the reservation comes in `credits_reserved` and a generation that fails costs nothing.

Do not confuse it with `insufficient_credit`, without the "s".
That one is the account balance in USD, for purchases through the API. And with a Coding Plan, when the points run out and there are no credits, the answer is `429` with `plan_limit_reached`, not `402`.

Repeating the request without changing anything does not help. Before a batch, check your balance with `GET /v1/ai/balance`. See also **what max_tokens changes in an answer**.

RECOMMENDED

Out of credits? Buy more on the Buy credits tab.

See All in One AI →

SEE ALSO

- **What a chat request costs through the API: the reservation, max_tokens and the final adjustment** →
- **Checking your AI credit balance: the page, GET /v1/ai/balance and aio balance** →

- Before a large batch of AI requests: checking the balance and the price first
- Errors when setting up a tool with All in One AI: 401, model not found, no credits, /v1 and rejected config files
- 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

RLX-INT-KB-8994 · v2026.11.28 · Generated on 28/11/2026 21:01

Online version of this document