Mozout

Generated on 29/11/2026 09:28

KNOWLEDGE BASE · PUBLIC DOCUMENT

Moving an app from a direct AI account to the AI gateway: what to change, step by step

Using the AI API

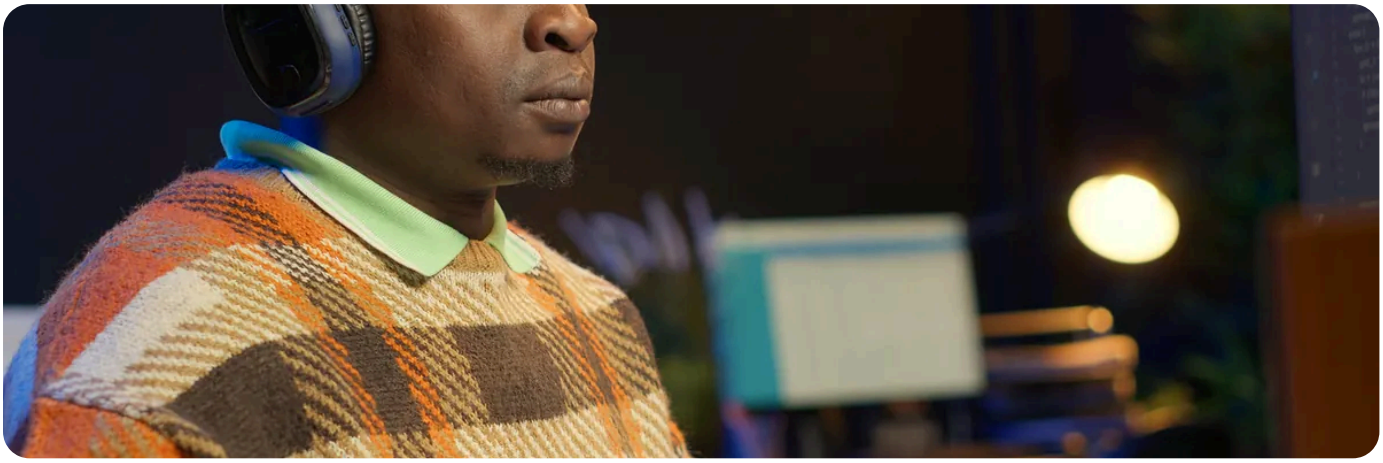
CODE	VERSION	COUNTRY · LANGUAGE
RLX-INT-KB-9918	v2026.11.29	International · English

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/9918

Moving an app from a direct AI account to the AI gateway: what to change, step by step

Using the AI API Created on **November 29, 2026** 2 min read



Do you have a program that already talks directly to an AI company and want to move it to the MozOut gateway? In most cases the change is small. This is the order we recommend.

Step by step

1. Create a `sandbox_` key in the **API Center**, with `ai:read` and `ai:write`, and keep it in an environment variable, `MOZOUT_API_KEY`.

2. Change the base URL to `api.mozout.com/v1`. In the OpenAI SDK it is `base_url`. For coding tools, follow the article for each one.

3. Change the model name to an id from our catalog. Our ids always have a slash, such as `deepseek-ai/deepseek-v4-flash`. A name without a slash gives `404 model_not_found`.

4. Run the program with the `sandbox_` key. Chat answers "Resposta de teste (sandbox)." and spends nothing: it confirms the address, the key and the format are right.

5. Buy credits, switch to the `prod_` key and make a small real request, with a low `max_tokens`.

What may need more care

If your program uses	What to do
Streaming	Ask for `stream_options` with `include_usage`. Without it, the whole credit reservation stays charged.

If your program uses	What to do
Requests without `max_tokens`	Always set `max_tokens`. Without it, we reserve 4096 output tokens, and with a low balance that gives `402`.
Images through the OpenAI SDK	Images, videos and voice use `POST /v1/media/generations`, with plain HTTP requests.
Tools (function calling)	They work in chat, but check that the chosen model accepts them. See **sending tools in a chat request**.
Error handling	Errors carry `code` and `message` inside `error`. Add `402 insufficient_credits` and `429 rate_limited`, which comes with the `Retry-After` header.

Make the change through environment variables (address, key and model), not in the code. That way you can go back in a minute if you need to. See **how to change models without touching your code**.

RECOMMENDED

Ready to move? Create the key and open All in One AI.

See All in One AI →

SEE ALSO

- Using the OpenAI SDK →
- Why every model id has a slash →
- The AI API errors →
- 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-9918 · v2026.11.29 · Generated on 29/11/2026 09:28