Mozout

Generated on 29/11/2026 04:04

KNOWLEDGE BASE · PUBLIC DOCUMENT

Apps made for Claude: the Anthropic Messages format at /v1/messages

Using the AI API

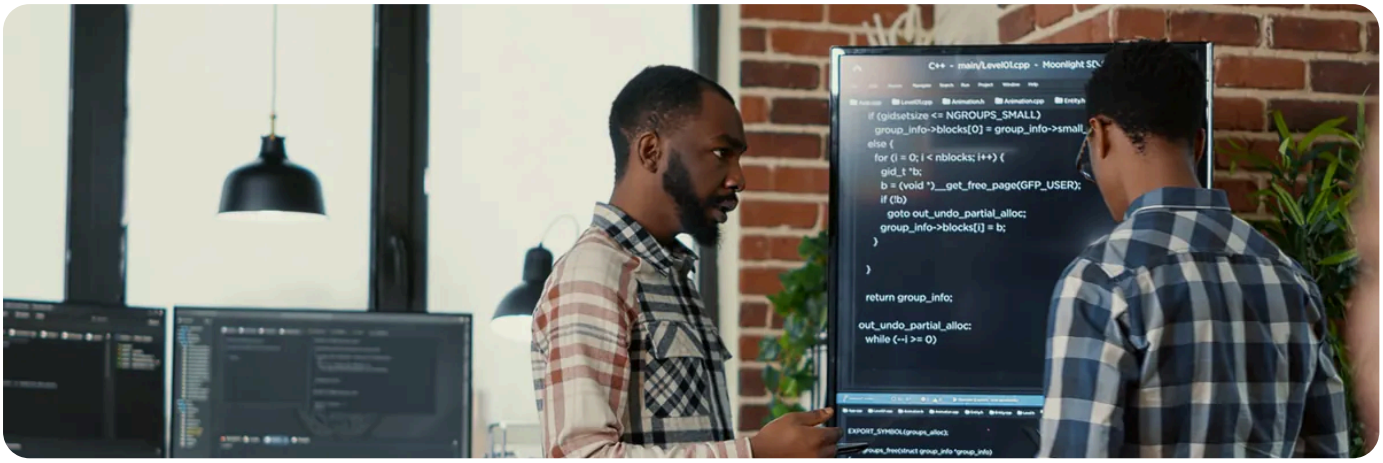
CODE	VERSION	COUNTRY · LANGUAGE
RLX-INT-KB-9084	v2026.11.28	International · English

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/9084

Apps made for Claude: the Anthropic Messages format at /v1/messages

Using the AI API Created on **November 28, 2026** 3 min read



In short: if your program was built with the Anthropic SDK, or you use Claude Code, you do not have to rewrite it. The API accepts the Anthropic Messages format at `POST /v1/messages`: you change the base address and the key, and everything else stays the same. The models, the agents, the credit reservation, the Coding Plan points and the Sandbox are the same as in chat.

The base address has no /v1

The base address is `https://api.mozout.com`, without `/v1`. The SDK and Claude Code add `/v1/messages` on their own. The key can go in `x-api-key`, as the Anthropic SDK sends it, or in `Authorization: Bearer`: we accept both. The `anthropic-version` header goes through as it is and you do not need to send it. You need a key with `ai:write`.

A request with curl

```
curl "https://api.mozout.com/v1/messages" \
  -H "x-api-key: $MOZOUT_API_KEY" \
  -H "Content-Type: application/json" \
  -d '{"model": "moonshotai/kimi-k2.7-code", "max_tokens": 1024, "system": "You are a careful senior developer.", "messages": [{"role": "user", "content": "Write a Python function that checks if a string is a palindrome."}]}'
```

The answer has the text in `content`, a list of blocks, the reason it ended in `stop_reason` and the tokens in `usage` (`input_tokens` and `output_tokens`).

With the Anthropic SDK, in Python

Install the SDK with `pip install anthropic` and give it our address in `base_url` (or in the `ANTHROPIC_BASE_URL` variable):

```
import os
from anthropic import Anthropic

client                  =                  Anthropic(base_url="https://api.mozout.com",
api_key=os.environ["MOZOUT_API_KEY"])

message = client.messages.create(
    model="moonshotai/kimi-k2.7-code",
    max_tokens=1024,
     messages=[{"role": "user", "content": "Write a Python function that checks if a
string is a palindrome."}],
)
for block in message.content:
    if block.type == "text":
        print(block.text)
```

What is different from OpenAI chat

Field	In the Anthropic Messages format
`max_tokens`	Required. We reserve credits for this number of output tokens. Without it, or without `model` or `messages`, the answer is `400`.
`system`	The instructions go in a separate field, not in a message with the `system` `role`.
`content`	Text, or a list of blocks: `text`, `image`, `tool_use` and `tool_result`.
`tools`	Each tool has `name`, `description` and `input_schema`. The model asks for it in a `tool_use` block and your program answers with a `tool_result` block.
`temperature`	From 0 to 1: lower is more predictable.
`stream`	With `true`, the answer comes as Anthropic events, from `message_start` to `message_stop`.
Errors	They come in the Anthropic format, with `error.type` (for example `rate_limit_error` on a 429), `error.message` and `request_id`, which is the same as the `X-Correlation-ID` header.

The request body can be up to 8 MB.

Always use a model from the catalog

A name that is not in the catalog, such as `claude-sonnet-...`, does not give an error: it uses our default coding model, `moonshotai/kimi-k2.7-code`, and charges that model. The answer brings back the name you sent, so it may look like another model. To know what you pay for, send in `model` an id from the `GET /v1/models` list, with the slash, or `agent/` followed by the id of one of your agents.

To count the input tokens before you send, there is `POST /v1/messages/count_tokens`. It is an estimate: it does not call the model, costs nothing and spends no points.

With a `sandbox_` key, the answer is always "Resposta de teste (sandbox).", with `usage` at zero. Use it to check the address and the key without spending. For Claude Code, the variables are in **using Claude Code with All in One AI**.

RECOMMENDED

Want to choose the model for your app? See the catalog in All in One AI.

See All in One AI →

SEE ALSO

- **Using Claude Code with All in One AI: the variables, the model and common errors**
- **The OpenAI Responses format: /v1/responses and responses kept for 30 days**
- **Testing All in One AI in the Sandbox at no cost: what each answer looks like**
- **Sending tools (function calling) in a chat request**
- **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-9084 · v2026.11.28 · Generated on 29/11/2026 04:04