interweb

Generated on 28/11/2026 15:47

KNOWLEDGE BASE · PUBLIC DOCUMENT

Using Hermes Agent with All in One AI: the custom endpoint, step by step

All in One AI

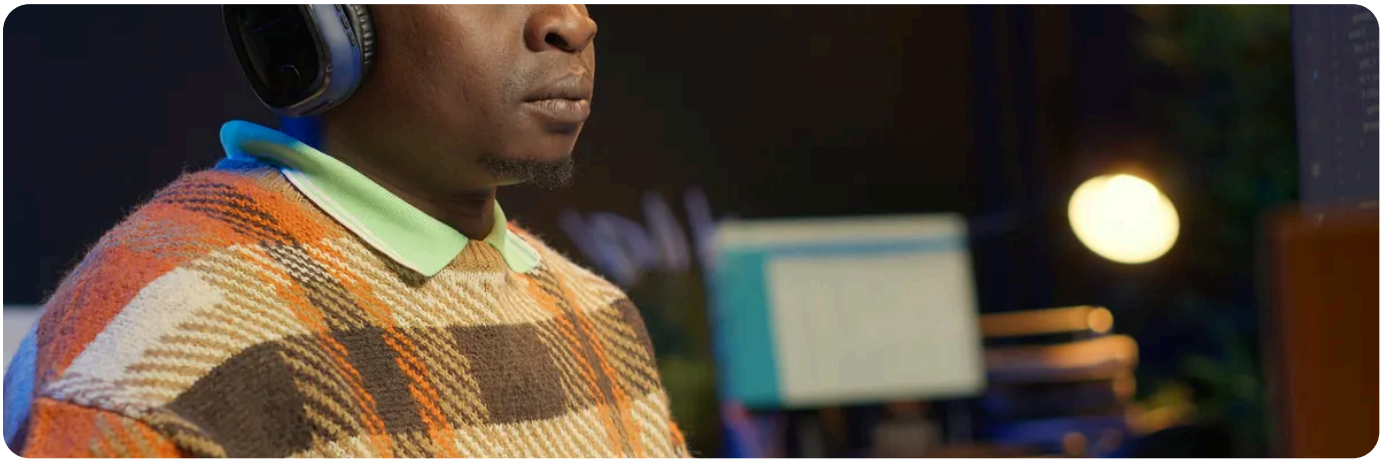
CODE	VERSION	COUNTRY · LANGUAGE
RLX-MZ-KB-2926	v2026.11.28	Mozambique · English

Interweb Mozambique
RLX Holdings, S.A. · Group RLX Holdings
Tax ID (NUIT) 401 191 607 · Legal entity no. (NUEL) 105 045 974 · Economic activity code (CAE) 62022

https://www.meu.interweb.co.mz/knowledgebase/2926

Using Hermes Agent with All in One AI: the custom endpoint, step by step

All in One AI Created on **November 28, 2026** 3 min read



Hermes Agent is an open-source agent by Nous Research. It learns as you use it: it creates skills from what it does and improves them over time. It works in the terminal and on channels such as Telegram, Discord, Slack and WhatsApp. It accepts any OpenAI-compatible address, so it can use All in One AI models with the same Interweb key, paid with AI credits or with the Coding Plan.

Before you start

1. Create a key in the **API Center** with the `ai:read` and `ai:write` permissions. Use a `sandbox_` key to check the setup without spending and a `prod_` key to work. The fields are explained in **creating an API key**.

2. Have AI credits or an active Coding Plan in **All in One AI**.

If you do not have Hermes Agent yet, the official install on Linux, macOS and WSL2 is one line:

```
curl -fsSL https://hermes-agent.nousresearch.com/install.sh | bash
```

On Windows, in PowerShell: `iex (irm https://hermes-agent.nousresearch.com/install.ps1)`. To keep it running all the time, put it on a VPS: see **why an agent needs a VPS**.

The simplest way: hermes model

```
hermes model
```

1. In the list, choose **Custom endpoint (self-hosted / VLLM / etc.)**.

2. For the API base URL, type `https://api.interweb.co.mz/v1`.

3. For the API key, paste your `sandbox_` or `prod_` key.

4. For the model name, type the catalogue id, for example `moonshotai/kimi-k2.7-code`.

5. When it asks for the context length, leave it blank so Hermes detects it, or type `262144`.

Hermes saves everything in `~/.hermes/config.yaml`. Done this way, the key ends up written in that file.

Keeping the key out of the file

Hermes keeps keys in `~/.hermes/.env` and settings in `~/.hermes/config.yaml`. Create the folder and open `.env`:

```
mkdir -p ~/.hermes
nano ~/.hermes/.env
```

Type this line, with your key, and save with Ctrl+O, Enter and Ctrl+X:

```
INTERWEB_API_KEY=prod_...
```

Then open the settings file with `nano ~/.hermes/config.yaml`. If it already has a `model:` block, replace only that block with this one. If the file is empty, paste it like this:

```
model:
  default: moonshotai/kimi-k2.7-code
  provider: custom
  base_url: https://api.interweb.co.mz/v1
  key_env: INTERWEB_API_KEY
```

The two spaces at the start of the lines matter: in YAML, they are what put each field inside `model`. `key_env` is the name of the variable that holds the key. On native Windows, the Hermes files live in `%LOCALAPPDATA%\hermes\` instead of `~/.hermes/`.

Checking it works

```
hermes doctor
hermes chat
```

`hermes doctor` warns about problems in the provider setup. When it opens, `hermes chat` shows the context limit it detected (Context limit). Say "hello". With a `sandbox_` key, the answer is always "Resposta de teste (sandbox).".

Which model to use

Hermes needs at least 64,000 tokens of context to work with tools, and a model that supports them. `moonshotai/kimi-k2.7-code` (262,144 tokens), `zai-org/glm-5.2`, `deepseek-ai/deepseek-v4-pro` and `deepseek-ai/deepseek-v4-flash` (1,048,576 tokens each) all work. To switch models, run `hermes model` again or change `default` in `config.yaml`.

Hermes works on its own, and every step is a paid request.
Start with the `sandbox_` key and check the real spending in the History tab of **All in One AI**. A key with only `ai:read` and `ai:write` limits the damage if it leaks.

RECOMMENDED

Want Hermes Agent to use our models? See All in One AI.

See All in One AI →

SEE ALSO

- OpenClaw, Hermes Agent, ZeroClaw, Nanobot, NanoClaw or PicoClaw: which agent to run on your VPS
- Using OpenClaw with All in One AI: your own provider, the key and the model
- Hosting an AI agent: why it needs a VPS and not shared hosting
- Testing All in One AI in the Sandbox at no cost: what each answer looks like
- All in One AI

interweb

Interweb Mozambique

RLX Holdings, S.A. · Group RLX Holdings

TAX ID (NUIT)	401 191 607
LEGAL ENTITY NO. (NUEL)	105 045 974
ECONOMIC ACTIVITY CODE (CAE)	62022
REGISTERED OFFICE	Bairro 7 de Setembro Vilankulo 1304 · CEP 0411-02, Inhambane, Mozambique
CONTACT	info@interweb.co.mz · +258 84 635 1111 · +258 86 635 1111 · www.interweb.co.mz
OPENING HOURS	Monday to Friday - 08 AM to 04 PM Saturday & Sunday - Closed

Online version of this document

RLX-MZ-KB-2926 · v2026.11.28 · Generated on 28/11/2026 15:47