interweb

Generated on 15/11/2026 00:16

KNOWLEDGE BASE · PUBLIC DOCUMENT

Running an AI model on your own computer with Ollama or LM Studio

How AI Works

CODE	VERSION	COUNTRY · LANGUAGE
RLX-AO-KB-1582	v2026.10.10	Angola · English

Interweb Angola
RLX Angola, Lda. · Group RLX Holdings
Tax ID (NIF) 500 260 5724 · Commercial registration no. 26464-25/250724

https://www.meu.interweb.ao/knowledgebase/1582

Running an AI model on your own computer with Ollama or LM Studio

How AI Works Created on **October 10, 2026** 3 min read



You can run an AI model **on your own computer, without sending anything over the internet**, with a program such as **Ollama** or **LM Studio**. You download the program, download an open model, and the conversation happens on your machine. It works well for experimenting, for confidential texts and for learning. It does not match the biggest cloud models, and the speed depends on your computer's **memory**.

This is different from running Ollama on a server, which is covered in **running models on your VPS**. Here you need no server, no domain and no outside connection.

Ollama or LM Studio

	Ollama	LM Studio
How you use it	From the command line, and through a local API for other programs	A windowed application: search for models, download and chat
Who prefers it	Those at ease with the terminal or wanting to connect other tools	Those who want to experiment without commands
Where to get it	The project's official website	The project's official website
Connecting other programs	By default it listens only on the computer itself (the port the documentation gives is 11434)	Has an optional local server; see the documentation

First steps, in short

1. **Download the program only from the official site.** There are fake copies with the same name.

2. **Pick a small model** to begin with. Each model states on its page the memory it needs; read it before downloading. The files are large: check your disk space.

3. **In Ollama,** the basic commands are `ollama pull model-name` to download, `ollama run model-name` to chat and `ollama list` to see what you have. The exact name comes from the project's model list.

4. **In LM Studio,** search for the model inside the application, download it and open a chat.

5. **Measure.** See how long it takes to answer and whether the result is enough for the task. Only then move to a bigger model, if memory allows.

What to expect

Aspect	Reality
Speed	Depends mostly on memory and on whether there is a graphics card. Without one, the bigger models get slow.
Quality	Small models make more mistakes and follow complicated instructions less well.
Privacy	The text does not leave the machine, as long as the program is not opened to the network.
Model licence	Each model has its own. Some restrict commercial use; read it before using the model for business.
Errors	They invent like the cloud ones do. See **how to check an answer**.

Do not open the program to the network
without knowing what you are doing. If a local service becomes reachable from other computers, anyone on the network can use your equipment. Keep it on localhost.

If the goal is to
protect data
, start by asking whether the task really needs a big model. Summarising, classifying and rewording short texts often goes well on small models.

RECOMMENDED

Need more memory than your computer has? Have a look at the VPS plans.

See VPS plans →

SEE ALSO

- Running AI models on your own VPS with Ollama: when it makes sense →
- What happens to what you type into an AI service: reading the data terms →
- AI hallucinations: how to check an answer before you use it →

interweb

Interweb Angola

RLX Angola, Lda. · Group RLX Holdings

TAX ID (NIF)	500 260 5724
COMMERCIAL REGISTRATION NO.	26464-25/250724
REGISTERED OFFICE	Rua Direita do Patriota, Bairro Dangereux Belas, Luanda, Angola
CONTACT	info@interweb.ao · +244 942 011 191 · www.interweb.ao
OPENING HOURS	Monday to Friday - 08 AM to 04 PM Saturday & Sunday - Closed

Online version of this document

RLX-AO-KB-1582 · v2026.10.10 · Generated on 15/11/2026 00:16