



Manage your local LLM models with style

Available Models

Scan Files

Mixtral 14B-8bit.gguf

Local



Port

Launch

Qwen 2.5 Coder 14B-8bit.gguf

Local



Port

Launch

Rocket 3B-8bit.gguf

Local



Port

Launch

Running Instances

Refresh

Stop All

PID	MODEL FILE	HOST	PORT	RAM USAGE	ACTIONS
-----	------------	------	------	-----------	---------

No models running

Settings

Public Access

Allow other devices on network to connect



GGUF Model Directory

Choose File

no file selected

/Applications/AI Models

Tip: If browser blocks path detection, paste the full path above manually.

Llama Server Binary

Use Internal (Llamatic Bundled)



Save & Restart Server