Me

Tell me, how long can i use this tool without any costs for free

AI Assistant

I do not have access to your personal account information, billing details, or the specific pricing structure of the platform you are using to interact with me.

Since you are using this model via **LMStudio**, the "cost" depends on how you have set up your environment:

1. **Local Execution:** If you are running the model locally on your own computer, there is typically no subscription fee. The "cost" would only be your hardware's electricity consumption and the initial investment in your computer's GPU/RAM.
2. **Cloud Providers:** If you are accessing this through a cloud-based API or a hosted service, those providers usually have their own pricing models (such as monthly subscriptions or per-token charges).

I recommend checking the documentation or the pricing page of the specific interface or provider you are using to get the most accurate information regarding limits and costs.

As your **AI Assistant**, I am here to help you with your software development tasks whenever you need!

Me

Tell me, where can i find local execution , i use lm studio

AI Assistant

Since you are using **LMStudio**, you are already using a tool designed specifically for **local execution**.

When you use LMStudio, "local execution" means the AI model runs directly on your own computer's hardware (your CPU, GPU, and RAM) rather than on a remote server or a cloud provider.

Here is a breakdown of how to ensure you are using it locally and what that entails:

1. How it works in LMStudio

By default, when you download a model within LMStudio and click the "Chat" or "Local Server" button, the software uses your computer's resources to generate text.

- **No Internet Required:** Once a model is downloaded, you can technically disconnect from the internet, and it will still work.
- **Privacy:** Your data stays on your machine and is not sent to an external company for processing.

2. Hardware Requirements

Because the "cost" is shifted from a subscription fee to your own hardware's power, your experience depends on your machine's specs:

- **GPU (Graphics Card):** This is the most important component. Models run much faster on NVIDIA (CUDA) or Apple Silicon (M1/M2/M3) GPUs.
- **RAM/VRAM:** You need enough memory to hold the model. For example, a 7B or 8B parameter model typically requires at least 8GB to 16GB of total system memory (or VRAM).
- **Disk Space:** You need enough storage to download the model files (often several gigabytes each).

3. How to verify you are running locally

You can usually see this in the LMStudio interface:

- **GPU Offloading:** In the "Load Model" settings on the right side, you will see options to offload layers to your GPU. If this is enabled and active, your graphics card is doing the heavy lifting.
- **System Monitor:** While you are chatting, you can open your computer's Task Manager (Windows) or Activity Monitor (Mac). You will see your CPU and GPU usage spike as the model generates text.

Summary: You don't need to "find" a local execution button—it is the primary way LMStudio functions. You are already using it! You only "pay" with the electricity your computer uses and the hardware you already own.

Me

So i use Ai completition, with lmstudio, now it seems, no costs so far, right?

AI Assistant

That is correct. Because you are using **LMStudio** to run the model locally on your own machine, there are **no ongoing costs** per message, per token, or per hour for the AI's responses.

Here is a quick summary of why there are no costs in your current setup:

1. No Subscription or Per-Token Fees

Unlike cloud services (like OpenAI, Google Gemini, or Anthropic), where you pay a monthly fee or a "pay-as-you-go" rate for every word the AI generates, your local hardware does the work. LMStudio is the interface that manages this for you.

2. No Data Outflow

Since the model is running on your hardware:

- Your prompts are not being sent to a remote server for processing.
- Your data remains private on your own machine.

3. What "Cost" actually exists?

While there is no software cost, there are two minor "hardware costs" associated with local execution:

- **Electricity:** Your computer (especially your GPU) will consume more power while the model is generating text.
- **Hardware Wear:** Running intensive tasks like AI inference uses your computer's components.

Summary Table: Local vs. Cloud

Feature	**Local (LMStudio)**	**Cloud APIs (e.g., OpenAI/Gemini)**
Cost per Request	**$0 (Free)**	Paid (Per token/request)
Privacy	High (Stays on your machine)	Lower (Sent to a provider)
Speed	Depends on your GPU/RAM	Depends on their servers
Internet Required	No (Once model is downloaded)	Yes

Conclusion: As long as you are running the model through LMStudio on your own machine, you can continue using it to help develop your **Java SDK (version 24)** project without worrying about any additional charges!