Beginner's Guide: Full Process Tutorial for Running DeepSeek-R1 Locally on AMD Graphics Cards (Extremely Detailed!)

Hello everyone!
I'm 九九吐, bringing you a super practical tutorial that will walk you through how to run DeepSeek-R1 locally on AMD graphics cards. If you've been looking for ways to make AMD cards work smoothly with large language models (LLMs), this guide is perfect for you! 🚀

Step 1: Environment Preparation (Alpha Phase)

1️⃣ Download Ollama for AMD

- Visit the official Ollama for AMD website.

- Navigate to the "Releases" page.

- Find and download the **.exe** file.

2️⃣ Download Runtime Libraries (ROCmLibs)

- Go to the ROCmLibs for AMD website.

- Download the shader version that matches your AMD graphics card.

- **Tip:** Unsure about your card's shader version? Check TechPowerUp for details.

Step 2: Installation and Configuration

1️⃣ Install Ollama

- Run the **ollama.exe** file you downloaded earlier to complete the installation.

2️⃣ Configure ROCm Runtime Libraries

- Extract the **.rar** files from the downloaded ROCmLibs.

- For example: **rocm.gfx1031.for.hip.sdk.6** (adjust according to your card).

- Copy the extracted files into the following directory:

C:\Program Files\Ocl-ICD\icd64

- Add the path to your system environment variables.

3️⃣ Run DeepSeek-R1 🚀

- Visit the official Ollama website and select the model you want.

- Open CMD and run the following command:

ollama run deepseek-r1

- If the conversation starts successfully, congratulations! Everything is working fine! 🎉

Step 3: Troubleshooting (If Necessary)

Common Issues:

1. **Ollama Still Running on CPU**

   - Solution: Download and install the AMD-compatible Ollama for CPU version.

   - Link: [AMD Ollama CPU Installer](#)

Step 4: Web UI Operations (More Convenient!)

Recommended Tools:

- **Page Assist (A W Extension):** A popular extension for enhancing web interactions with AI models.

Instructions:

1. Open the gear icon (⚙) in the top-right corner and set your preferred language.

2. Enter "RAG Settings," select the Ollama model, and save.

3. Enjoy a smooth web-based interaction! 🎮

Conclusion & Resources 🔗

- **Success!** You've successfully set up DeepSeek-R1 to run locally on your AMD graphics card!

- No need to rely on expensive NVIDIA cards—AMD users can now work with large language models too!

GitHub Resources:

- [DeepSeek-R1 Official Repository](#)

- [Detailed Performance Paper for DeepSeek-R1](#)