Yogesh Haribhau Kulkarni • You
AI Advisor (Helping organizations in their AI journeys) | PhD (Geometric …
now • 🌐

Wish to develop Generative AI apps locally? with Open Source LLMs? 🤨

As an AI advisor, I typically try the potential approaches myself first, on a smaller scale, using open source LLMs (Large Language Models), on my machine 🖥️. Although closed-source models are typically found to be superior, their open-source counterparts are not too far behind, with the obvious advantage of being free 📊.

So, how do I use open source LLMs (oLLms)? I have two approaches:

➡️ Offline mode: I use it via LM Studio 📁. It downloads the model locally and makes it easily callable from frameworks like LangChain for RAG (Retrieval Augmented Generation) or AutoGen for agents. You can explore different Hugging Face models! 🚀 However, it can be slow, depending on your hardware 🖥️.

➡️ Online mode: I use the Groq API, which provides popular open source LLMs like Llama (Meta Facebook), Gemma (Google), and Mixtral (Mistral AI) (and DeepSeek!) 🚀. It's fast, free (with rate limits), and easy to use with frameworks like LangChain via an OpenAI-like API 🤝.

How about you? Do you have a better alternative? Share your experiences in the comments! 💬

#AI #OpenSource #LLMs #GenerativeAI #LangChain

#HuggingFace #GroqAPI #MachineLearning #Innovation 🚀
#DataScience #NLP #GroqCloud #autogen #Python 🤩

GDG Cloud Pune GDG Pune #mvpbuzz #gde