

➤ **Title**

Run Your Own AI (But Private)

➤ **Source**

https://youtu.be/WxYC9-hBM_g?si=GOQklvsmUcmtJSQC

➤ **Introduction**

In this video he explains how to run an AI model on your localhost (Own System). Without depending on cloud platforms like OpenAI or Goggle. This is useful when you want your data to remain private, or want to experiment with AI in more hands-on and offline environment.

➤ **-Why we use Private AI**

- ✓ Public AI models (Like Chatgpt , Deepseek, Gemini) work online and store your data on external servers
- ✓ With private AI,your data stays secure on your own system-no risk of sharing sensitive information
- ✓ Its also great way to learn how AI models run in the background.

➤ **Tools used in video**

✓ **VMware Workstation**

This tool is used to create virtual machine on your own system. And the AI model will run inside this virtual machine.

✓ **Ubuntu**

It is installed inside the virtual machine as the base of operating System. It is lightweight and perfect running models.

✓ **LM Studio**

This app used to download and run open-source Large Language Models locally. Also, it supports models like LLaMA , Mistral and others also.

➤ **Step by step process explained in the video**

✓ **Installing VMware**

Download and Install VMware on your system after that create new virtual machine and install ubuntu Linux as the operating system.

✓ **Setting Up Ubuntu**

After that System allocates some Resources like RAM and Storage also Boot into ubuntu and complete basic setup.

✓ **Installation LM Studio**

Inside Ubuntu, download LM studio and app for running LLMs. Then Choose and Open Source LLM and download the model file.

✓ **Running AI model**

Once Installed then open LM studio and type prompts, then the response are generated on your own System with no internet required.

➤ **Benefits of this setup**

✓ Privacy: Everything runs on your own machine.

✓ Offline Access: After setup you don't need Internet to run your Own AI chatbot.

✓ Customization: You can test different models then settings and tools.

➤ **What I learned**

✓ I learned about how to create a private AI chatbot using few open-source tools.

✓ It's possible to use local hardware to run large AI models without depending on cloud services.

✓ This setup is useful for learning, exploring AI safety and maintaining privacy.

➤ Screenshot's:



INTRODUCING

VMware Private AI Foundation with NVIDIA

Integrated Gen-AI Platform:
Available Early 2024

VMware Private AI Foundation with NVIDIA

Triton Management Service | NVIDIA GPU and Network Operator | NVIDIA | NVIDIA NeMo Framework | NVIDIA TensorRT-LLM

Deep Learning VMs | Vector Database Postgres + PGVector | VMware | Catalog Setup Wizard | GPU Monitoring

VMware Cloud Foundation | NVIDIA AI Enterprise

Dell Technologies | Hewlett Packard Enterprise | Lenovo

fine tuning is cool and training an LLM on your own data is awesome,

vmware by Broadcom

NVIDIA

RAG "Retrieval-Augmented Generation"

LLM

RAG Data

Fast e-to

2