🔐 Dockerized AI Inference API with NGINX + Ollama

🧠 AI Agent Deployment — Project 2

🚀 Docker + NGINX + Ollama

👨‍💻 From Lab → Production

🔗 Watch Full Video

🛠️ What We Built

✅ **Multi-Model AI Agent** powered by:

- 🐳 Docker (containerized deployment)
- 🌐 NGINX (reverse proxy & load balancing)
- 🤖 Ollama (runs models like Mistral, Gemma, DeepSeek, LLaMA)
- 🧪 curl & Postman for testing

📦 Why This Matters

🔒 Production-grade architecture

⚡ Fast, isolated, and secure

📈 Learn to deploy AI models like the pros

🌍 Works across platforms — local or cloud

🧪 Demo in Action

📤 Send a prompt using curl/Postman

🧠 Ollama generates intelligent output

🌐 NGINX manages routing cleanly

📦 Everything running inside containers

💡 Key Skills You Learn

- ✅ API & model orchestration
- ✅ Serving multiple AI models
- ✅ Dockerfile best practices
- ✅ Reverse proxy with SSL-ready config
- ✅ Real-world DevOps meets AI

🧭 The Remoder Difference - remoder.com

🌐 **remoder.com**: We don’t just teach —

We **build production-ready AI systems** 🏗️

💥 Hands-on labs

💬 Mentorship

📂 GitHub-ready code

🚀 Real deployments

🎬 Watch the Full Project

- 📺 https://www.youtube.com/watch?v=QYtIB04AVWM
- 💥 Don’t just learn AI — build it.

#AIEngineer #LLM #Ollama #Docker #NGINX #FastAPI #Remoder #AIinProduction #TechEducation