Gen AI Engineering: Building RAG & Multi-Agent AI Systems

Date & Time: 9th March, 2025 (2:30 PM to 5 PM IST)

Hi Mastermind learners,

The world is shifting towards intelligent AI systems, in which AI agents can search, reason, and act on their own with specific domain knowledge. Let us take a step into understanding and building such an AI system.

In this **4-hour interactive session**, you'll gain insights into:

- Retrieval-Augmented Generation **(RAG)** and its applications in building smarter AI solutions
- Use **LLMs + FAISS** for fast, accurate data retrieval in RAG.
- Build **multi-agent AI** for various use cases in 2 different agentic AI framework
- Get hands-on experience with **real-world datasets**

Key Takeaways:

- Learn **FAISS-powered retrieval** for creating knowledge base for AI
- Build working system of **AI Agents** on multiple use cases
- Gain hands-on **OpenAI / Groq + CrewAI / PhiData** experience

💡 **Who should attend?**
💻 Software Developers | Data Analysts | 🧠 AI Enthusiasts | 💻 Tech leaders

This session is designed for developers and engineers who:

🛠️ Are comfortable coding in Python and eager to build and deploy real-world AI solutions.

Note - you're not hands-on with Python yet, you might find the practical exercises challenging, but you'll still gain a solid understanding of the concepts and frameworks!

If you're excited about building the future with AI, this session is your next step.

Agenda: What We'll Cover

🔹 **Retrieval-Augmented Generation (RAG) – Enhancing AI with Knowledge Retrieval**

- Understand how RAG improves **accuracy and relevance** in Generative AI by integrating expert/domain knowledge.
- Learn how **LLMs + FAISS** enable efficient, context-aware content generation.
- **Hands-on:** Build a RAG-based system to extract insights from reference documents and enhance them using LLMs.

🔹 **Agentic AI for Real-World Applications**

- Explore the evolution of AI towards **autonomous decision-making** with multi-agent systems.
- Learn how AI agents solve **industry-specific tasks** by leveraging specialized tools and knowledge.
- Discover the key **technologies and frameworks** that power agentic AI systems.

🔹 **Building Multi-Agent Systems with Crew AI & PhiData**

- **Design & implement** multi-agent AI solutions for use cases like **financial advisory, interactive learning, and data analytics.**
- Build an **AI-powered financial assistant** that analyzes real-world stock data to generate recommendations.
- Develop a **RAG-enabled agentic AI system** to create an interactive learning platform.
- Learn how agentic AI can process structured information to **derive actionable insights.**

A bit about the Mentors

1. Ramanathan - With over two decades of rich experience in Data Science and Machine Learning, Ramanathan has established himself as a trusted mentor and leader in the industry. He has a proven track record of delivering transformative solutions for global clients, leveraging his expertise in statistical modeling, predictive analytics, and advanced AI/ML techniques. He is actively exploring Agentic AI frameworks to shape the future of intelligent systems.

2. Vishnuvardhan B K M - A seasoned engineering software professional with over 2 decades of industrial experience spanning software development, certification, technical leadership, and program management. Extensive expertise in the aerospace and automotive industries, specializing in safety-critical systems. Adept at software architecture, design, and process definition, I am passionate about exploring and adopting emerging technologies. Currently a technical consultant, driving innovative solutions in software architecture and projects to develop solutions based on generative AI.

⚡ Prerequisites & Setup – Complete Before the Session begins!

🚀 **Get Ready for the Session!** Watch this **4-minute setup guide** to prepare your environment before we go hands-on.

🎥 **Watch Now:** Setup Your Environment

Make sure you're all set before the session so you can dive straight into building AI agents! 🔥

#1 Google Colab Setup

To ensure a smooth hands-on experience, we'll be using Google Colab—a free, browser-based platform to run Python code, including machine learning workflows.

Why Google Colab?

- No installations—run Python directly in your browser.
- Free access to GPUs/TPUs for accelerated computing.
- Work saved on Google Drive, accessible anytime.

How to Get Started

- Open: Google Colab - https://colab.research.google.com/
- Sign in with your Google Account (or create one).
- Once logged in you can,
  1. Open existing notebooks
  2. create a new notebook
  3. Import from Github or Google Drive

- Try creating simple "Hello World! script :
  print("Hello, Workshop Participants!")
  Press Shift + Enter to run it.

Tips: For using Google Collab

- Save your work regularly in Google Drive.
- To enable GPU/TPU: Go to Runtime > Change runtime type.
- Install libraries with !pip install <library_name>.

#2 Groq Setup

Obtain a GROQ API Key - For Agentic hands on session

To fully participate in masterclass, you'll also need a GROQ API Key to access the tools and workflows seamlessly. Here's how you can get it:

Steps to Get Your GROQ API Key

- Visit the https://console.groq.com/ or your organization's GROQ portal.
- Log in with your registered email and password. (New user? Click Sign Up and follow the prompts.)

- Go to API Key Settings under Settings or Account.
- Click Generate API Key, name it (e.g., “Workshop Key”), and set permissions if required.
- Copy the key immediately and save it securely—i**t’s displayed only once!**

Test Your API Key (Optional)
Use this Python snippet in Colab to verify:

```
!pip install groq
import os
from groq import Groq

client = Groq(api_key="generated GROQ_API_KEY")

chat_completion = client.chat.completions.create(
   messages=[
      { "role": "user","content": "Explain the importance of fast language models",}
   ],
   model="llama-3.3-70b-versatile",
)

print(chat_completion.choices[0].message.content)
```

Keep Your Key Secure

💡 Treat it like a password—don’t share or expose it in code repositories.

Steps to store it in Google Colab

- Visit your Google Colab page (https://colab.research.google.com/)
- After logging in, you will see ‘secret’ tab (key icon).
- Store the Groq API key with a name “GROQ_API_KEY” in a specific secret. Refer the video guide.

Note: Groq Free API keys have limitations in terms of tokens - you might face some difficulty while trying the hands on. Your Open AI API key (paid ones) will be very handy, if you have one.

#3 Hugging Face Setup

Hugging face is a platform that is provided with multiple pre-trained models that help in deep learning, Gen AI etc (https://huggingface.co/). We would be using some of the models from this platform in our Agentic AI hands on session. An API key for each user is required in this regard.

Steps to Get Your HF API Key

- Visit the https://huggingface.co/ and signup with your details (first time users).
    - This would require your basic details (mail id etc)
- Visit your user setting - Access token (https://huggingface.co/settings/tokens).
- Create a new token (with ‘read’ permission). Copy the token key and preserve it for your use. Remember this is personal and keep it safe.

Steps to store it in Google Colab

- Visit your Google Colab page (https://colab.research.google.com/)
- After logging in, you will see ‘secret’ tab (key icon).
- Store the hugging face API key / token with a name “HF_TOKEN” in a specific secret. Refer the video guide.