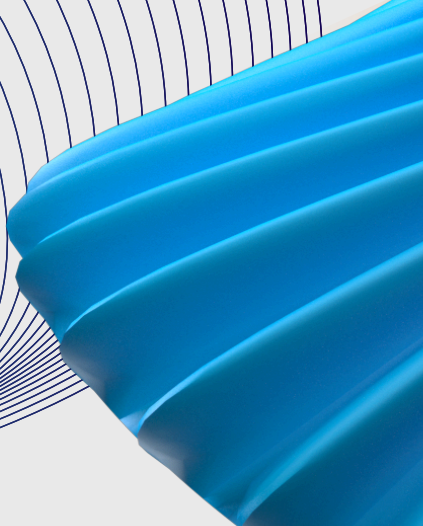
KRISH NAIK ACADEMY

Agentic AI & GenAI

With Cloud



aws

Google Cloud

This course is designed for AI enthusiasts and professionals eager to dive into the world of Agentic AI and Natural Language Processing (NLP). You will gain hands-on experience in building intelligent systems, mastering NLP techniques, developing agentic decision-making frameworks, and implementing state-of-the-art GenAI solutions. By the end, you'll be proficient in creating, deploying, and monitoring advanced NLP and agentic applications.

Learning Objectives

- Build a solid understanding of Natural Language Processing (NLP) techniques and their applications in AI systems.
- Implement and optimize NLP models for tasks like text classification, sentiment analysis, and named entity recognition using libraries like SpaCy and Hugging Face.
- Explore Generative AI (GenAI) techniques, including transformers and language models, for generating and completing text tasks.
- Develop agentic systems by learning multi-agent collaboration, decision-making frameworks, and feedback loops in complex environments.
- Master retrieval-augmented generation (RAG) for improving information retrieval and generation tasks in NLP applications.
- Apply state-of-the-art frameworks like LangChain and LangFlow to build scalable and efficient NLP workflows.
- Build and deploy conversational agents and chatbots using LangGraph, LangFlow, and integration with third-party tools and APIs.
- Understand and apply observability techniques for monitoring LLMs and agentic systems, leveraging tools like Langfuse and LangWatch.

Course Information

Prerequisites

Python : Proficiency in Python fundamentals, including variables, data types, control structures, and libraries like NumPy, Pandas, and Matplotlib for data manipulation and visualization.

NLP : Experience with Python libraries like NLTK, SpaCy, and TextBlob for text preprocessing, tokenization, stemming, lemmatization, and machine learning concepts like classification and vectorization.

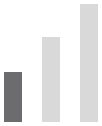
GenAI : Knowledge of Python, deep learning frameworks (TensorFlow, PyTorch), and Generative AI models like transformers, GANs, and VAEs, along with data manipulation skills using NumPy and Pandas.

The course is designed to be completed over a duration of approximately four to five months, providing an in-depth exploration of agentic AI and NLP concepts, with plenty of time for practical implementation and real-world applications.

Estimated Time

5 months 6hrs/week*

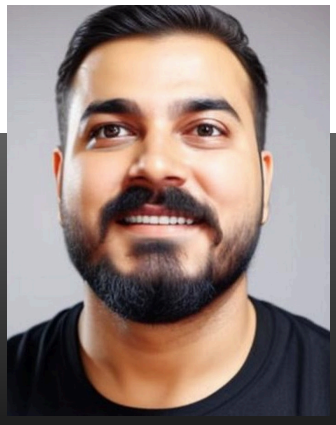
Required Skill Level



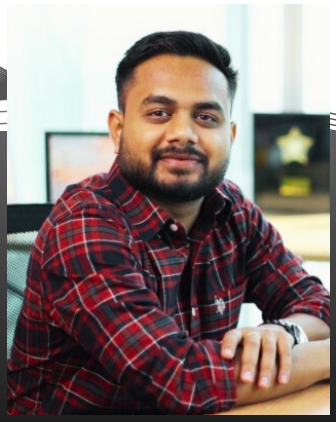
Begineer+

Course Instructors

Mayank Aggrawal
Senior ML Engineer
Linkedin



Krish Naik
Chief AI Engineer
Linkedin



Sunny Savita
GenAI Engineer
Linkedin

Sourangshu Pal
Senior Data Scientist
Linkedin

Module 1

Introduction to Agentic AI

In this module, we'll explore the fundamentals of Agentic AI, including what defines an agent, how it differs from traditional AI agents and generative AI, and the role of multi-agents in problem-solving. You'll learn about various frameworks used in Agentic AI, which enable the creation, management, and orchestration of intelligent agents. This foundational knowledge will set the stage for understanding the broader applications and evolution of agent-based systems in AI and how they can be leveraged for complex decision-making, automation, and problem-solving tasks.

Topics	
What is Agentic AI?	What are Agents?, Agentic AI vs AI Agents, Agentic AI vs Generative AI, What are Multi-Agents?
Agentic AI Frameworks	Overview of Agentic AI Frameworks

Module 2

Phi Data: Agentic AI Framework

This module introduces Phi Data as a powerful framework for building Agentic AI systems. You'll learn how to integrate agents with various models, tools, and knowledge sources. Topics covered include the essential concepts of chunking, vector databases, and embedding techniques that form the backbone of Agentic AI. The module also covers the design and execution of workflows, enabling you to create intelligent agents that process and retrieve data efficiently. We'll explore real-world use cases such as web search agents, financial agents, and retrieval-augmented generation (RAG) agents to understand how these concepts are applied in practice.

Topics	
Core Concepts	Agents in Phi Data, Models, Tools, Knowledge, Chunking
Data and Storage	Vector Databases (VectorDbs), Storage, Embeddings
Workflows	Workflow Design and Execution
Use Cases	Web Search Agents, Financial Agents, Retrieval-Augmented Generation (RAG) Agents

Module 3

LangChain

In this module, we dive into LangChain, a framework that simplifies the creation of complex AI applications using LLMs (Large Language Models). You will learn how to use LangChain's components for data ingestion, document loaders, and text splitting techniques to prepare data for processing. Additionally, we cover how to work with embeddings from various sources such as OpenAI, Ollama, and Hugging Face, and integrate them with vector storage systems like FAISS and ChromaDB. By the end of this module, you'll have a deep understanding of how to structure, process, and store data within LangChain for use in AI models and agents.

Topics	
Core Components and Data Handling	Introduction to Basic Components and Modules in LangChain, Data Ingestion with Document Loaders
Text Splitting Techniques	Recursive Character Text Splitter, Character Text Splitter, HTML Header Text Splitter, Recursive JSON Splitter
Embeddings and Vector Storage	OpenAI Embeddings, Ollama Embeddings, Hugging Face Embeddings, VectorStores: FAISS and ChromaDB, VectorStore and Retriever

Module 4

LCEL (LangChain Expression Language)

This module focuses on LangChain Expression Language (LCEL), which allows you to work with LLMs more effectively. You'll learn how to get started with open-source models using the Groq API and how to build and optimize language models (LLMs). The module also covers the creation of prompt and output chains with LCEL to create efficient workflows and decision-making processes for intelligent agents. Lastly, we will explore how to deploy LangServe runnables and chains as APIs, enabling you to create production-ready agentic AI solutions that can scale across various platforms and use cases.

Topics	
Getting Started	Open Source Models Using Groq API
Building and Deploying	Building LLMs, Prompt and Structured Output Chains with LCEL, Deploying LangServe Runnables and Chains as APIs

Module 5

LangServe for Efficient AI Deployment

In this module, we explore LangServe, a powerful framework that streamlines the deployment of AI models for production environments. You will learn how to deploy and scale AI applications efficiently using LangServe's robust features, ensuring smooth integrations with cloud platforms and optimization of resources for large-scale AI tasks.

Topics	
Overview and Setup	Overview of LangServe and Its Capabilities, Importance of Efficient AI Model Serving, Key Features and Benefits of LangServe, Setting Up the LangServe Environment, Installing LangServe and Initial Configuration, Configuring Environment Variables and Dependencies
Model Deployment	API-Driven Model Serving: How LangServe Bridges AI Models and Applications, Deploying Your Model with LangServe, Creating and Managing Custom Endpoints, Integrations with External Tools

Module 6

LangGraph

This module covers LangGraph, a framework that enables complex AI workflows. You will learn how to structure and manage state within LangGraph applications, handle deployments, and understand the integration of various components for building scalable AI systems, with an emphasis on state management and deployment strategies.

Topics	
Core Concepts	Introduction, Simple Graph, LangGraph Studio, Chain, Router
Agents	Agent, Agent with Memory, Intro to Deployment
State Concepts	State Schema, State Reducers, Multiple Schemas
Message Handling	Trim and Filter Messages
Deployment Concepts	Deployment Concepts, Creating and Connecting to Deployment

Module 7

UX and Human-in-the-Loop with LangGraph

This module focuses on creating human-in-the-loop workflows with LangGraph, where you will learn how to enhance user experiences in AI applications by integrating human feedback loops, optimizing AI outputs, and improving system interaction based on real-time user inputs.

Topics	
Interaction	Streaming, Breakpoints, Editing State and Human Feedback, Dynamic Breakpoints
Time Travel	Time Travel

Module 8

Agentic RAG

In this module, you will learn the principles of Agentic Retrieval-Augmented Generation (RAG), a technique that enhances AI agent capabilities by retrieving external data and combining it with generated content. The focus will be on creating agents that can autonomously retrieve and generate relevant information for improved performance.

Topics	
Adaptive RAG	Adaptive Rag, Adaptive Rag with Cohere, Adaptive rag in Local
RAG Variants	Agentic Rag, C-Rag, C-Rag in Local, Self Rag, Self Rag in Local, Self Rag with VectorDB

Module 9

Designing Multi-Agent Systems with LangGraph

This module delves into the design of multi-agent systems within LangGraph, where you will learn to build systems with multiple AI agents that can collaborate, share information, and solve complex problems together. You'll also explore agent communication and coordination mechanisms for efficient system performance.

Topics	
Agent Design	Building Agent Nodes in LangGraph, Agent Communication Protocols and Coordination, Defining Tasks and Roles for Agents
System Design	Creating Scalable Multi-Agent Systems in LangGraph, Building A Real-World Multi-Agent System

Module 10

CrewAI Platform

In this module, you will get an introduction to the CrewAI platform, a solution for creating and managing AI teams or agents. You will learn how to leverage CrewAI to coordinate and automate workflows involving multiple AI agents, optimizing collaborative tasks and decision-making.

Topics	
Overview	Definition and Overview, Key Features and Capabilities, Crew Collaboration Framework
Collaboration and Tools	AI-Agent Communication, Workflow Automation in CrewAI, Customizing CrewAI, Managing Data Across Agents, Role-playing, Memory, Tools, Focus, Guardrails, Cooperation, Using LangChain Tools

Module 11

LangFlow Overview and Setup

This module introduces LangFlow, a framework for creating and managing AI-driven flows. You will learn how to set up LangFlow for efficient workflow management, build AI-driven applications using its various components, and integrate different models seamlessly to automate tasks within a flow.

Topics	
Introduction and Setup	What is LangFlow? Overview and Use Cases, Key Features of LangFlow for LLM Applications, Setting Up Your LangFlow Environment
LangFlow UI and Terminologies	Understanding LangFlow UI and Workflows, Key Terminologies in LangFlow (Nodes, Chains, Prompts)
Quick Start	Quick Start: Creating Your First LangFlow Application
Core Concepts	Nodes and Chains: Core Concepts, Understanding LLMs and Their Integration with LangFlow, Pre-built vs. Custom Workflows

Module 11

LangFlow Overview and Setup

Topics	
LangChain and Prompt Engineering	Prompt Engineering Basics in LangFlow, LangChain Integration: Using LangFlow with LangChain
Commonly Used Nodes	Exploring Commonly Used LangFlow Nodes

Module 12

Integration with Third-Party Tools

In this module, we explore how to integrate LangChain and LangGraph with third-party tools and services. You will learn to extend AI workflows by connecting them to external APIs, databases, and other platforms, allowing for seamless data exchange and enhanced functionality in your AI projects.

Topics	
Data Integration	Connecting LangFlow with Data Sources (SQL, CSV, NoSQL), Using LangFlow with Vector Databases for Embeddings
API Integration	API Integration for External Services (REST, GraphQL), LangFlow with OpenAI and Hugging Face Models
Workflow Automation	Automating Workflows Using LangFlow, Building Chatbot Applications with LangFlow

Module 13

Langfuse for LLM Observability

This module covers Langfuse, a tool for tracking and monitoring large language model (LLM) performance. You will learn how to use Langfuse for observability, tracking model outputs, analyzing system performance, and identifying areas for optimization, ensuring that LLMs operate at their best.

Topics	
Langfuse Overview	What is Langfuse? Overview and Applications, Importance of Observability in LLMs, Key Features and Benefits of Langfuse, Understanding Langfuse's Integration Ecosystem
Integration and Monitoring	Step-by-Step Integration with Popular Frameworks (LangChain, OpenAI, etc.), Setting Up API Calls for Observability, Tracking Key Metrics: Response Times, Costs, and Errors, Monitoring Prompt Effectiveness and Token Usage

Module 14

Metrics and Monitoring in LangWatch

In this module, you will explore LangWatch, a monitoring tool for LangChain applications. You will learn how to track and analyze key metrics in real-time, gain insights into the health and performance of AI models, and ensure system reliability through effective monitoring strategies.

Topics	
LangWatch Overview	What is LangWatch? Overview and Use Cases, Key Features of LangWatch in Monitoring Language Models, Connecting LangWatch with LLMs
API Integration and Setup	API Integration: Sending Logs and Data to LangWatch, Setting Up Observability in AI Workflows
Using LangWatch with Frameworks	Using LangWatch with Popular Frameworks

Module 15

Langsmith

This module introduces Langsmith, a platform for enhancing and testing AI models. You will learn how to leverage Langsmith's features to refine and improve model outputs, test different model versions, and evaluate performance under varying conditions to ensure optimal AI model behavior.

Topics	
Langsmith Overview	What is LangSmith? Overview and Key Features, LangSmith in the AI Development Workflow
Setup and Configuration	Setting Up LangSmith: Installation and Configuration, Exploring the User Interface and Core Functionalities
Workflow Management	Understanding Workflow Pipelines in LangSmith, Creating and Managing AI Workflows, Data Integration in LangSmith, Preprocessing and Cleaning Data, Managing Data Streams and Sources

Module 16

Introduction to Autogen

This module introduces Autogen, an automated system for generating AI models. You will learn how Autogen simplifies model creation and tuning, allowing you to generate robust AI models faster and with minimal manual input, while maintaining high levels of performance and accuracy.

Topics	
Framework Overview	Overview,Key Concepts: Autonomy, Adaptability, and Inter-Agent Communication, Installation and Environment Setup
Agentic System Development	Introduction to Agents, Goals, Environments, and Actions, APIs, Libraries, and Tools Available Within the Autogen Framework, Designing and Developing Agentic Systems, Framework for Agentic Decision-Making
Agent Interaction and Learning	Interaction and Communication Between Agents, Implementing Feedback Loops, Handling Uncertainty and Constraints, Agent Learning and Adaptation, Multi-Agent Collaboration

Module 16

Introduction to Autogen

Topics	
Deployment and Monitoring	Deployment, Monitoring Agent Performance

Module 17

End to End Agentic AI Projects

In this module, we guide you through the process of creating end-to-end agentic AI projects, where you will learn how to build, deploy, and optimize autonomous AI agents that perform real-world tasks, incorporating data retrieval, processing, and decision-making within a unified AI system.

Topics	
Project-Based Learning	Agentic AI Projects

Module 18

AWS Cloud & Services for Generative AI

This module provides an introduction to AWS Cloud services, with a focus on deploying and managing generative AI models. You will learn how to use AWS services like EC2, S3, SageMaker, and Lambda for building, training, and scaling generative AI applications.

Topics	
Introduction to AWS Cloud	Detail introduction of AWS Cloud services, How to create an AWS account, How to create an IAM, Understanding Regions and Zones
AWS Compute and Container Services	AWS Elastic Container Registry, AWS Elastic Cloud Compute, AWS App Runner

Module 19

AWS Bedrock

In this module, we explore AWS Bedrock, a service for foundation models. You will learn how to use Bedrock for building and deploying large-scale AI models, with an emphasis on leveraging its inference capabilities, model types, and integration with other AWS services for seamless AI deployments.

Topics	
Introduction to AWS Bedrock	Amazon Bedrock - Introduction, Bedrock Console Walkthrough, Amazon Bedrock - Architecture
Bedrock Models and Use Cases	Bedrock Foundation Models, Bedrock Embeddings, Bedrock Chat Playgrounds
Bedrock Inference and Pricing	Amazon Bedrock - Inference Parameters, Bedrock Pricing

Module 20

AWS SageMaker

This module introduces AWS SageMaker, a comprehensive machine learning platform. You will learn how to use SageMaker for end-to-end ML development, including model training, deployment, and monitoring. Additionally, you'll explore SageMaker Studio for streamlined workflow management and optimization of AI models.

Topics	
Overview of AWS SageMaker	AWS SageMaker Overview, AWS SageMaker Walk-through, AWS SageMaker Studio Overview, AWS SageMaker Studio Walk-through
Model Deployment with SageMaker	Choose a Pre-trained Model, SageMaker Endpoint Creation, SageMaker Console Access, Create SageMaker Domain, Open SageMaker Studio, SageMaker Models Deployment

Module 21

AWS Lambda

This module focuses on AWS Lambda, a serverless compute service. You will learn how to create, deploy, and manage serverless functions using Lambda, as well as how to integrate it with other AWS services for automated workflows, reducing infrastructure management overhead.

Topics	
Overview of AWS Lambda	Overview of AWS Lambda, Lambda Console Walkthrough, Lambda Permissions Model

Module 22

AWS API Gateway

In this module, you will learn about AWS API Gateway, a service for building APIs. You will explore how to create RESTful and WebSocket APIs, integrate them with AWS Lambda, and use them for efficient AI application development, ensuring scalable and reliable API management.

Topics	
API Gateway Overview	AWS API Gateway, RESTful APIs, WebSocket APIs
Efficient API Development	Efficient API Development

Module 23

Text Summarization with AWS Services

This module guides you through creating text summarization applications using AWS Lambda and API Gateway, integrated with Bedrock services. You will learn how to set up Lambda functions, connect them to Bedrock, and expose a RESTful API for summarizing text efficiently.

Topics	
Integration of AWS Lambda with Bedrock and API Gateway	Creation of AWS Lambda function and Boto3 upgrade, Writing the AWS Lambda function to connect to Bedrock Service, Create REST API using AWS API Gateway and Lambda Integration

Module 24

Fine-Tuning Foundation Models on Custom Data

In this module, you will learn how to fine-tune pre-trained foundation models on your own custom data using AWS SageMaker. You will explore the steps involved in preparing data, configuring models, and performing training to optimize performance for specific use cases.

Topics	
Fine-Tuning Overview	Fine-Tuning of Foundation Model - Overview, Fine-Tuning of Foundation Model - Architecture
Hands-On with AWS SageMaker	Fine-Tuning of Foundation Models - Hands On AWS SageMaker

Module 25

Project : AWS

This module combines the creation of a Retrieval-Augmented Generation (RAG) system and a chatbot using Llama3, Langchain, and Streamlit. You will learn how to build a fully integrated project that combines data retrieval with natural language generation to provide intelligent conversational agents.

Topics	
Retrieval-Augmented Generation (RAG) in AWS	Overview, Setup, Data Transformation and Processing, LLM and Retrieval QA, Frontend and Backend Development
Building Chatbot with Llama3, Langchain & Streamlit	Overview, Setup, Data Handling and LLM Creation, Frontend and Final Demo

Module 26

GCP Basics & Introduction to Vertex AI

This module introduces Google Cloud's Vertex AI, a platform for building and deploying AI models. You will learn the basics of Vertex AI, including setup, model management, and integration with Google Cloud's infrastructure, enabling efficient AI solutions in the cloud.

Topics	
Introduction to Google Cloud and Vertex AI	What is Vertex AI?, Google AI Studio Introduction, Google Cloud Regions & Zones, Foundation Google Models
Vertex AI Setup and Installation	Vertex AI Installation, Google Cloud Setup for Production, Vertex AI Overview, Vertex AI Model Garden

Module 27

Gemini Models with Vertex AI and Google AI Studio

This module explores Google Gemini models and their integration with Vertex AI and Google AI Studio. You will learn how to leverage these models for advanced AI applications, utilizing the AI Studio's features for building, deploying, and managing Gemini-based AI projects.

Topics	
Introduction to Google Gemini	What is Google Gemini?, Google Gemini: Playing with Gemini, Gemini 1.5 Pro (Preview only), Gemini 1.0 Pro
Gemini Embeddings and Retrieval	Gemini Embeddings, Advanced Information Retrieval with Gemini
Working with Prompts	Working with Freeform & Structured prompts, Working with Text Chat prompt
Multimodal and Text-Based Use Cases	Generate Code, Unit test with Code Chat Bison model, Translate text with Translation LLM, Summarization, Classification
Multimodal Applications	Vision Model, Speech to Text & Text to Speech, Multimodal Prompts

Module 28

Project : GCP

This module combines the implementation of Retrieval-Augmented Generation (RAG) in GCP with building a chatbot using Gemini Pro, Langchain, and Streamlit. You will learn how to integrate RAG and build an intelligent conversational system using these technologies for enhanced interactivity.

Topics	
Retrieval-Augmented Generation (RAG) in GCP	Overview, Setup, Data Transformation and LLM Context, Frontend and Final Demo
Building Chatbot with Gemini Pro, Langchain & Streamlit in GCP	Overview, Setup, Data Transformation and LLM Creation, Frontend and Final Demo