Complete Generative AI: From Basics to Expert Level

Duration: 80 hours

Course Overview

This course offers a comprehensive, hands-on journey through the modern AI stack—from foundational machine learning and deep learning concepts to advanced applications using open-source tools like Hugging Face, LangChain, and Azure OpenAI. Learners will explore model training, quantization, unstructured data preprocessing, prompt engineering, semantic search, and vector databases. The program also covers chatbot development, agent orchestration, and LLMOps best practices for scalable deployment. With practical labs and real-world projects, participants will gain the skills to build efficient, responsible, and production-ready AI solutions across domains and modalities.

Pre-requisites

- Basic Python programming
- Foundational ML knowledge
- Experience with Jupyter Notebooks or Python IDEs

Course Agenda

Chapter 01: Foundations of Machine Learning & Deep Learning

- Core ML concepts: supervised, unsupervised, reinforcement learning.
- Deep learning architectures: CNNs, RNNs, Transformers.
- Types of Transformers Architecture
- Hands-on: Train Machine learning and Deep Learning models.
- Prerequisite: Basic Python

Chapter 02: Open Source Models with Hugging Face

- Explore Hugging Face Hub for open-source models.
- Use the `transformers` library for text, audio, image, and multimodal tasks.
- Share apps via Gradio and Hugging Face Spaces.
- Hands-on: Deploy your first Hugging Face AI app.

Chapter 03: Quantization Fundamentals with Hugging Face

- Why quantization matters for efficiency.
- Compress models using Hugging Face Transformers + quantization libraries.
- Explore linear quantization.
- Practice quantizing multimodal & language models.

Chapter 04: Preprocessing Unstructured Data for LLM Applications

- Extract data from PDFs, PowerPoints, Word, HTML, images, and tables.
- Add metadata for enhanced RAG.
- Apply layout detection, vision, and table transformers.
- Hands-on: Preprocess enterprise data for RAG.

Chapter 05: Prompt Engineering with Llama 2 & 3

- Best practices for prompting Llama 2 & 3.
- Work with Meta Llama 2 Chat, Code Llama, and Llama Guard.
- Design safe, responsible AI apps.
- Hands-on: Prompt design & evaluation.

Chapter 06: Building Applications with Vector Databases

- Applications: semantic search, RAG, anomaly detection, hybrid search, image similarity, recommender systems.
- Implement with Pinecone.
- Build hybrid multimodal search (text + images).
- Hands-on: Facial similarity ranking app.

Chapter 07: LangChain: Chat with Your Data

- Build chatbots for private data/documents.
- Use 80+ data loaders in LangChain.
- Hands-on: “Chat with Docs” project.

Chapter 08: Large Language Models with Semantic Search

- Improve search using LLMs + summarization.
- Keyword Search using BM25, Semantic Search using Transformers ,and Hybrid Search.
- Apply embeddings for dense retrieval.
- Evaluate and optimize search effectiveness.

Chapter 9: Finetuning Large Language Models

- Basics of finetuning vs. prompt engineering.
- Domain adaptation with real datasets.
- Hands-on: Fine-tune a Hugging Face model.
- Prerequisite: Basic Python

Chapter 10: Agents with Langchain and Semantic Kernel in Azure

- Introduction to AI agents and orchestration.
- Build and deploy agents using Azure OpenAI platform.
- Demo: Agent interacting with APIs and knowledge bases.
- Hands-on: Enterprise-grade agent demonstration in Azure.

Chapter 11: LLMOps

- Best practices for deploying & maintaining LLMs.
- Automate fine-tuning pipelines and API deployment.
- Version control for models and datasets.
- Responsible AI: safety scoring for harmful content.
- Hands-on: Build a complete MLOps → LLMOps workflow.