

# Generative AI Advanced with LLMOps

**Duration: 24 hours**

## Course Overview

This course provides a comprehensive journey into deploying and managing large language models in production. Beginning with the foundations of MLOps, you'll explore essential concepts such as versioning, interpretability, and containerized workflows. You'll dive into the world of prompt engineering, learning strategies for crafting effective prompts. The course guides you through packaging LLM applications, using tools like FastAPI, Docker, and Kubernetes for seamless deployment. You'll also master automation with GitHub Actions, ensuring efficient CI/CD pipelines tailored to LLM workflows. Monitoring and maintaining models in production become second nature with hands-on experience in tools like WhyLabs. Finally, you'll explore fine-tuning LLMs to build and deploy tailored AI applications, making you adept in end-to-end LLMOps practices.

## Course Contents

### Module 01: Introduction to LLMOps

- Introduction to MLOps
- Model versioning and data versioning(Mlflow etc )
- Model interpretability and explainability
- Containerized ML Workflow with Docker
- MLOps vs LLMOps
- Deploying ML via APIs and web applications

### Module 02: Prompt Engineering with LLMs

- Introduction to Prompt Engineering
- Architecture and Component of a Prompt
- Strategies and Techniques of Prompt Generation

### Module 03: Packaging the LLM Application

- Introduction to FastAPI
- Packaging the LLM Application
- Create Requirements.txt
- Deploy AI agent and RAG agent
- Introduction to Docker and installation
- Build Docker image for LLM application

### Module 4: Deploying Container App with Kubernetes

- Introducing Kubernetes
- Architecture of Kubernetes
- Running the Application on Kubernetes
- Implementing the Kubernetes for our containers

- Deploy LLM based app using Docker and Kubernetes

### **Module 5: Github Actions**

- Introduction to GitHub Actions
- Quick Demo on github actions YAML file
- Create github Actions from Scratch
- AKS, ACR and ACI for deployment
- Configure Workflow based on LLM use case
- Implement CI/ CD with Github Actions

### **Module 6: Monitoring of LLM Models in Production**

- Introduction to WhyLabs & LLM Monitoring
- Introduction to Langkit and WhyLabs Platform
- Hands On - Monitoring LLM in Production

### **Module 7: Fine Tuning and Deployment**

- Fine tune a LLM
- Build AI app with fine tuned LLM
- Deploy fine tuned model using Docker