

LLMOps: Docker, Kubernetes & GitHub Actions CI/CD

Learn to package, deploy and monitor LLM applications using FastAPI, Docker, Kubernetes and GitHub Actions CI/CD pipelines.

24 Hours

DURATION

7

MODULES

31

TOPICS

1

HANDS-ON ACTIVITIES

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.

WHO THIS COURSE IS FOR

MLOps Engineers

DevOps Engineers

AI Application Developers

Machine Learning Engineers

Cloud Engineers

LEARNING OBJECTIVES

- Explain MLOps concepts including versioning and model interpretability
- Compare MLOps and LLMOps workflows
- Apply prompt engineering strategies to craft effective prompts
- Package LLM applications with FastAPI and Docker
- Deploy containerised LLM applications using Kubernetes
- Implement CI/CD pipelines for LLM workflows with GitHub Actions
- Monitor deployed LLM models using WhyLabs and Langkit
- Fine-tune an LLM and deploy the tailored application

7 Modules, 31 Topics

01

Introduction to LLMOps

TOPICS

- 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

02

Prompt Engineering with LLMs

TOPICS

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

03

Packaging the LLM Application

TOPICS

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

04

Deploying Container App with Kubernetes

TOPICS

- Introducing Kubernetes
- Architecture of Kubernetes
- Running the Application on Kubernetes
- Implementing the Kubernetes for our containers
- Deploy LLM based app using Docker and Kubernetes

05

GitHub Actions

TOPICS

- 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

06

Monitoring of LLM Models in Production

TOPICS

Introduction to WhyLabs & LLM Monitoring · Introduction to Langkit and WhyLabs Platform

HANDS ON

- Monitoring LLM in Production

07

Fine Tuning and Deployment

TOPICS

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

WHAT IS INCLUDED

Everything You Need to Succeed

- Koenig course material for every module
- Delivery by certified instructors and subject matter experts
- Hands-on exercises throughout the course
- Course completion certificate from Koenig

ABOUT KOENIG

Who We Are

Established in 1993, Koenig Solutions is a globally recognised training provider with 35+ technology partnerships, 5,000+ training programmes and a network of certified instructors and subject matter experts, supporting individuals and enterprise teams worldwide.

FLEXIBLE DELIVERY

Learn the Way That Suits You

Classroom Training

Live Online Classes (ILO)

Fly-Me-A-Trainer (FMAT)

Flexi (Self-Paced)

1-on-1 Training

Ready to Get Started?

Talk to a learning advisor about schedules, group training and customised delivery.

Visit koenig-solutions.com

✉ info@koenig-solutions.com