KOENIG SOLUTIONS · COURSE OUTLINE

LLM Deployment: Quantization, Monitoring & Azure

Learn to preprocess, secure, optimise, monitor and deploy large language models, including quantization and deployment on Azure.

3 Days	6	16
DURATION	MODULES	TOPICS

COURSE OVERVIEW

This course covers data preprocessing for LLM training, including tokenization and text and image embedding, followed by responsible AI principles such as fairness, transparency and accountability. It examines security perspectives in generative AI, covering risk identification and mitigation measures, then addresses model optimization techniques including pruning, quantization and compression. It continues with model monitoring after deployment, covering performance monitoring, drift detection and issue resolution, and concludes with deployment on cloud platforms, focusing on Microsoft Azure infrastructure and scalability.

WHO THIS COURSE IS FOR

- **Machine Learning Engineers**
- **AI Engineers**
- **Data Scientists**
- **Cloud Infrastructure Engineers**
- **MLOps Engineers**

LEARNING OBJECTIVES

- Explain tokenization and text and image embedding for LLM inputs
- Describe fairness, transparency and accountability in responsible AI
- Identify security challenges and risks specific to generative AI systems
- Apply measures such as robustness training and privacy protection
- Implement optimization techniques such as pruning and quantization
- Analyse model performance and detect data drift post-deployment
- Diagnose and resolve issues in deployed models
- Deploy large language models on Microsoft Azure infrastructure

COURSE SYLLABUS

6 Modules, 16 Topics

01

Theory + Labs

Data Preprocessing for LLM Training

TOPICS

- **Tokenization:** Breaks text into smaller units like words or subwords for model input
- **Text Embedding:** Maps words or phrases to continuous vectors for semantic understanding
- **Image Embedding:** Represents images as vectors to incorporate visual information

02

Theory & Use Case

Responsible AI and GenAI

TOPICS

- **Fairness:** Mitigates bias in models to ensure equitable outcomes
- **Transparency:** Makes AI decision-making understandable for users
- **Accountability:** Establishes frameworks for AI system responsibility

03

Theory + lab

Security Perspectives in GenAI

TOPICS

- **Challenges:** Identifies unique security risks in AI systems
- **Measures:** Implements strategies like robustness training and privacy protection

04

Theory + Lab

Model Optimization Techniques for GenAI Models

TOPICS

- **Pruning:** Removes unnecessary parameters to reduce model size
- **Quantization:** Reduces numerical precision for faster inference
- **Compression:** Shrinks model size through techniques like weight sharing

05

Theory + Use Cases

Model Monitoring After Deployment

TOPICS

- **Monitoring Systems:** Continuously checks model performance post-deployment
- **Drift Detection:** Identifies changes in data distribution over time
- **Issue Resolution:** Diagnoses and resolves issues in deployed models

06

Theory + Use Cases

Deployment on Cloud Platforms

TOPICS

- **Azure Deployment:** Deploys models on Microsoft Azure cloud infrastructure
- Topics may include setting up infrastructure and managing scalability

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