
Google Generative AI in Production

Duration: 1 Day Course Code: GO8338 Delivery Method: Company Event

Overview:

Traditional MLOps is a set of practices to productionize traditional ML systems for enterprise applications. Generative AI raises new challenges in managing and productionizing applications at scale. The field of generative AI operations seeks to address these new challenges.

In this course, you learn about the challenges that arise when deploying and productionizing generative AI-powered applications. You learn how to secure your generative AI-powered applications. Finally, you will discuss best practices for logging and monitoring your generative AI-powered applications in production.

Products:

- Vertex AI

Updated August 2026

Company Events

These events can be delivered exclusively for your company at our locations or yours, specifically for your delegates and your needs. The Company Events can be tailored or standard course deliveries.

Target Audience:

Developers, DevOps engineers and machine learning engineers who wish to operationalize GenAI-based applications

Objectives:

■ After this course participants should be able to:

- | | |
|---|--|
| ■ Understand generative AI operations | ■ Maintain and update LLM models |
| ■ Compare traditional MLOps and GenAIOps | ■ Test and evaluate gen AI-powered apps |
| ■ Analyze the components of an LLM system | ■ Deploy CI/CD pipelines for gen AI-powered apps |
| ■ Define and compare RAG and ReAct | ■ Identify security challenges for gen AI applications |
| ■ Evaluate application deployment options | ■ Understand prompt security issues |
| ■ Deploy, package, and version apps | ■ Apply sensitive data protection and DLP API |
| | ■ Implement Model Armor |
-

Prerequisites:

Completion of the "Application Development with LLMs on Google Cloud" or equivalent knowledge.

Testing and Certification

None

Follow-on-Courses:

None

Content:

Module 01: Introduction to Generative AI in Production

- Generative AI operations
- Traditional MLOps vs. GenAIOps
- Components of an LLM system
- RAG/ReAct architecture

Module 02: Generative AI Application Deployment

- Application deployment options
- Deployment, packaging, and versioning
- Lab: Deploying an Agentic Application on Cloud Run

Module 03: Productionizing Generative AI

- Maintenance and updates
- Testing and evaluation
- CI/CD pipelines for gen AI-powered apps
- Lab: Tracking Versions of Generative AI Applications

Module 04: Securing Generative AI Applications

- Security challenges
- Prompt security
- Sensitive Data Protection and DLP API
- Model Armor
- Lab: Securing Generative AI-Powered Applications

Further Information:

For More information, or to book your course, please call us on Head Office 01189 123456 / Northern Office 0113 242 5931

info@globalknowledge.co.uk

www.globalknowledge.com/en-gb/

Global Knowledge, Mulberry Business Park, Fishponds Road, Wokingham Berkshire RG41 2GY UK