GenAIOps Enablement with Red Hat AI Enterprise

Course code: AI501

Experience the practices, culture, and tools that enable teams to reliably and efficiently build, deploy, and maintain GenAI applications in production.GenAIOps Enablement with Red Hat AI Enterprise (AI501) is a five-day immersive enablement, delivered the Red Hat Way, to build the skills that teams need to articulate and deliver on their AI vision. While many AI training programs focus on a particular framework or technology, this course covers how the tools fit together in a full Generative AI Operations workflow, treating the AI-enabled application, not just the model, as the unit of delivery.To achieve the learning objectives, participants should include multiple roles from across the organization. AI engineers, application developers, platform engineers, architects, and IT managers will gain experience working beyond their traditional silos. The daily routine simulates a real-world delivery team building an AI-powered application, where cross-functional teams learn how collaboration breeds innovation. Armed with shared experiences and best practices, the team can apply what it has learned to help the organization's culture and mission succeed in the pursuit of generative AI initiatives.This course is based on Red Hat AI Enterprise, including Red Hat OpenShift AI, as well as Red Hat OpenShift GitOps, Red Hat OpenShift Pipelines, and Generative AI models and open source libraries.

Who is the course for

This experience demonstrates how individuals across different roles must learn to share, collaborate, and work toward a common goal to achieve positive outcomes and drive generative AI innovation.

It is especially valuable for:

- AI Platform Users: AI engineers, application developers, data scientists, and data engineers building generative AI applications
- AI Platform Providers: ML/GenAIOps engineers and platform engineers deploying and managing AI infrastructure
- AI Platform Stakeholders: Architects and IT managers evaluating and overseeing generative AI adoption strategies

The scenario incorporates technical aspects of working with large language models and generative AI systems, offering practical insights into how these roles can align their efforts.

What we teach you

This course takes you on an end-to-end journey of an AI-enabled application, from prompt experimentation to production deployment, while bringing different personas together to collaborate on a single platform seamlessly.

- Understanding GenAI fundamentals, including tokens, context windows, and model behavior
- Experimenting with prompts and evaluating your first AI-enabled application
- Introducing an orchestration layer for standardized GenAI development
- Implementing Retrieval Augmented Generation (RAG) for knowledge-enhanced applications
- Building autonomous AI agents with tool-calling capabilities
- Deploying AI safety guardrails and implementing GenAI security practices
- Enabling observability with metrics, logging, and distributed tracing for GenAI systems
- Exploring small language models and multi-modal capabilities
- Optimizing models through quantization and compression techniques
- Implementing Models as a Service (MaaS) for scalable AI infrastructure

Required skills

- Take our free assessment
- to gauge whether this offering is the best fit for your skills.
- A chromium-based browser
- Containers, Kubernetes and Red Hat OpenShift Technical Overview
- (DO080) or basic understanding of OpenShift/Kubernetes and Containers is helpful
- Basic level understanding of AI or how your business can drive value from AI is beneficial

Course outline

GOPAS Praha
Na Strži 2097/63
140 00 Praha 4 - Krč
Tel.: +420 226 201 390
info@gopas.cz

GOPAS Brno
Nové sady 996/25
602 00 Brno
Tel.: +420 530 513 590
info@gopas.cz

GOPAS Bratislava
Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 902 903 132
info@gopas.sk

Copyright © 2026 GOPAS, a.s.,
All rights reserved

GenAIOps Enablement with Red Hat AI Enterprise

Core Foundations

GenAI Fundamentals

- Explore what GenAIOps is and how large language models work, including tokenization, context windows, and the factors that affect model behavior and performance.

Experimenting with Prompts

- Learn to craft effective prompts using system prompts and user prompts, configure temperature and output parameters, and optimize prompts for specific use cases.

Evaluating Your First AI-Enabled Application

- Implement prompt versioning, build evaluation pipelines, automate testing, and measure application quality systematically.

Introducing the Orchestration Layer

- Introduce an orchestration layer for building GenAI applications, deploy backend services, and implement GitOps practices for continuous deployment.

Advanced Topics

Integration and Orchestration

- Deploy vector databases, build RAG pipelines for knowledge-enhanced applications, implement tool calling, and create autonomous AI agents.

Safety and Observability

- Deploy AI safety guardrails, implement GenAI security practices, enable the three pillars of observability, metrics, logs, and traces.

Modeling Techniques

- Explore small language models for efficient deployment and multi-modal model capabilities for handling diverse input types.

Optimization and Deployment

- Apply quantization and compression techniques for improved performance, explore fine-tuning approaches, implement Models as a Service (MaaS), and bring it all together in a production deployment.

GOPAS Praha
Na Strži 2097/63
140 00 Praha 4 - Krč
Tel.: +420 226 201 390
info@gopas.cz

GOPAS Brno
Nové sady 996/25
602 00 Brno
Tel.: +420 530 513 590
info@gopas.cz

GOPAS Bratislava
Dr. Vladimíra Clementisa 10
Bratislava, 821 02
Tel.: +421 902 903 132
info@gopas.sk

Copyright © 2026 GOPAS, a.s.,
All rights reserved