GenAIOps Enablement with Red Hat AI Enterprise (AI501) Training

Duration: **1 Day**

BILGINÇ IT Academy
info@bilginc.com
bilginc.com

Overview

Experience the practices, culture, and platform capabilities required to build, deploy, operate, and maintain Generative AI applications reliably in production.

GenAIOps Enablement with Red Hat AI Enterprise (AI501) is a five-day immersive program designed to help teams develop the skills required to turn a Generative AI vision into reliable, production-ready applications.

While many AI training programs focus primarily on a specific framework, model, or technology, AI501 takes an end-to-end **Generative AI Operations** approach. The course treats the entire AI-enabled application—not only the underlying model—as the unit of delivery.

Participants work through the full lifecycle of a GenAI application, from prompt experimentation and evaluation to deployment, observability, continuous improvement, and Day-2 operations.

The experience is deliberately cross-functional. AI engineers, application developers, platform engineers, architects, and IT managers work beyond their traditional roles and collaborate as a shared delivery team.

The daily workflow reflects a real-world environment in which multiple disciplines contribute to a single AI-powered application. Participants experience how shared practices, common tooling, and collaboration can improve consistency and support innovation across Generative AI initiatives.

The course is based on **Red Hat AI Enterprise**, including **Red Hat OpenShift AI, Red Hat OpenShift GitOps, Red Hat OpenShift Pipelines**, Generative AI models, and open source libraries.

Benefits for Organizations

Organizations adopting Generative AI often encounter operational complexity caused by tool sprawl, prompt and configuration drift, quality regressions introduced through change, poorly governed grounding that increases hallucination risk, security concerns such as prompt injection and harmful content, and unpredictable latency or costs that make scaling difficult.

GenAIOps provides a structured approach to addressing these challenges.

The practices introduced in AI501 can help organizations:

- Manage prompts and configurations as code.
- Apply version control to GenAI application changes.
- Introduce continuous automated evaluation.
- Govern RAG implementations more consistently.
- Enforce safety guardrails through the platform.
- Establish end-to-end observability.
- Move GenAI applications from prototype to production in a more controlled manner.

The course provides practical exposure to the full lifecycle of an AI-enabled application, from prompt

BILGINÇ IT Academy

info@bilginc.com

bilginc.com

and configuration versioning to deployment, continuous evaluation, and Day-2 operations.

Benefits for Participants

After completing AI501, participants should have a stronger understanding of the GenAI platform landscape and how Red Hat AI Enterprise fits into the broader GenAIOps ecosystem.

Participants will gain practical experience to:

- Understand the end-to-end lifecycle of a GenAI-enabled application.
- Establish structured prompt development and evaluation processes.
- Build RAG-based AI applications.
- Develop AI agents with tool-calling capabilities.
- Apply security controls and safety guardrails.
- Use metrics, logs, and distributed tracing for GenAI observability.
- Evaluate model optimization approaches.
- Design scalable model delivery with Models as a Service.
- Move AI applications from prototype to production.
- Apply GenAIOps practices to improve reliability and maintainability at scale.

Recommended Next Courses

After AI501, participants may continue with:

- **MLOps Practices with Red Hat OpenShift AI (AI500)** – Recommended for teams also working with predictive AI and machine learning models,
- **Red Hat OpenShift Administration II: Configuring a Production Cluster (DO280)** – Recommended for platform engineers who want deeper OpenShift administration skills.

Prerequisites

Participants are recommended to have:

- Completed Red Hat's free skills assessment to determine whether the course matches their current experience level,
- Access to a Chromium-based browser,
- **Containers, Kubernetes and Red Hat OpenShift Technical Overview (DO080)** or a basic understanding of OpenShift, Kubernetes, and containers,
- A basic understanding of AI or how AI can create value for a business.

GenAIOps Enablement with Red Hat AI Enterprise (AI501) Training

Duration: **1 Day**

BILGINÇ IT Academy
info@bilginc.com
bilginc.com

Who Should Attend

AI501 is designed to bring together multiple roles involved in Generative AI delivery so that they can collaborate toward a shared outcome.

The course is particularly valuable for the following groups:

AI Platform Users

- AI engineers,
- Application developers,
- Data scientists,
- Data engineers.

These roles focus primarily on building Generative AI-enabled applications.

AI Platform Providers

- ML/GenAIOps engineers,
- Platform engineers.

These participants focus on deploying, operating, and managing the infrastructure that supports AI workloads.

AI Platform Stakeholders

- Architects,
- IT managers.

These roles are typically responsible for evaluating, guiding, and overseeing Generative AI adoption strategies.

The course scenarios include practical technical work with Large Language Models and Generative AI systems while demonstrating how these different roles can align their responsibilities.

GenAIOps Enablement with Red Hat AI Enterprise (AI501) Training

Duration: **1 Day**

BILGINÇ IT Academy
info@bilginc.com
bilginc.com

What You Will Learn

AI501 follows an AI-enabled application from early prompt experimentation through production deployment.

During the course, participants will:

- Understand GenAI fundamentals such as tokens, context windows, and model behavior.
- Experiment with prompts and evaluate an initial AI-enabled application.
- Introduce an orchestration layer for more standardized GenAI development.
- Implement Retrieval Augmented Generation (RAG) for knowledge-enhanced applications.
- Build autonomous AI agents with tool-calling capabilities.
- Deploy AI safety guardrails and apply GenAI security practices.
- Enable observability through metrics, logging, and distributed tracing.
- Explore Small Language Models and multimodal capabilities.
- Improve model efficiency with quantization and compression techniques.
- Implement Models as a Service (MaaS) for scalable AI infrastructure.

Outline

Core Foundations

GenAI Fundamentals

Explore the meaning of GenAIOps and understand the fundamentals of how Large Language Models operate.

Learn about tokenization, context windows, and key factors that influence model behavior and performance.

Experimenting with Prompts

Learn how to create more effective prompts using system prompts and user prompts.

Configure temperature and output parameters and refine prompts for specific use cases.

Evaluating Your First AI-Enabled Application

Implement prompt versioning and create evaluation pipelines to measure application quality.

Automate testing and establish more systematic ways to detect quality changes introduced by prompt or configuration updates.

Introducing the Orchestration Layer

Introduce an orchestration layer to support more standardized development of GenAI applications.

Deploy backend services and apply GitOps practices to enable continuous deployment.

GenAIOps Enablement with Red Hat AI Enterprise (AI501) Training

Duration: **1 Day**

BILGINÇ IT Academy
info@bilginc.com
bilginc.com

Advanced Topics

Integration and Orchestration

Deploy vector databases and build RAG pipelines for knowledge-enhanced AI applications.

Implement tool calling and create autonomous AI agents capable of interacting with external systems and services.

Safety and Observability

Deploy AI safety guardrails and apply security practices for Generative AI systems.

Enable the three core pillars of observability—**metrics, logs, and traces**—to gain visibility into application behavior and performance.

Modeling Techniques

Explore Small Language Models as an option for more efficient deployment scenarios.

Review multimodal capabilities for applications that need to process different types of input.

Optimization and Deployment

Apply quantization and compression techniques to improve model performance and resource efficiency.

Explore fine-tuning approaches, implement **Models as a Service (MaaS)**, and bring the course components together in a production deployment.