Course Overview

The Advanced Generative AI Development on AWS course prepares developers to build and deploy production-ready generative AI solutions. This advanced 3-day training covers the full generative AI stack—from foundation models to enterprise integration—along with vector databases, retrieval augmentation, prompt engineering, governance, and agentic AI systems. Participants also learn AI safety, performance optimization, cost management, monitoring, and testing aligned with AWS's proven model for moving from experimentation to production.

Start your AWS Essentials journey by accessing Official AWS e-Learning for FREE. Learn AWS Foundations: Getting Started with the AWS Cloud Essentials, Migrating to AWS: A high level introduction, Introduction to Amazon Elastic Compute Cloud (EC2) and more - GET STARTED

Level: Advanced

Duration: 3 Days/24 hours

Delivery Type: Instructor-Led Training

Course Objectives

- Build production-ready generative AI solutions on AWS that meet enterprise standards for security, scalability, and reliability.
- Evaluate and select the right foundation models for business use cases, including benchmarking and dynamic model selection.
- Design resilient foundation model systems with circuit breakers, cross-region deployments, and graceful degradation.
- Develop data processing pipelines for multi-modal inputs with validation and optimization workflows.
- Implement vector database solutions using Bedrock Knowledge Bases, OpenSearch, and hybrid retrieval augmentation.
- Create advanced prompt engineering frameworks, including chain-of-thought techniques and enterprise prompt governance.
- Build autonomous AI agents with Bedrock Agents, supporting complex reasoning and tool integrations.
- Apply AI safety and security controls such as content filtering, privacy protection, and adversarial testing.
- Optimize performance and cost with token efficiency, batching, and intelligent caching.
- Implement robust monitoring and observability for foundation model applications.Develop systematic testing and validation frameworks for continuous AI quality assurance.

Who Should Go For This Training?

- Software Developer

Pre-Requisites

Recommended

- AWS Technical Essentials
- Generative AI Essentials on AWS
- 2 or more years of experience building production grade applications on AWS or with open source technologies, general AI/ML or data engineering experience
- 1 year of hands-on experience implementing generative AI solutions

Course Outline

Foundation Model Selection and Configuration

- Enterprise foundation model evaluation framework
- Dynamic model selection architecture patterns
- Resilient foundation model system designs
- Cost optimization and economic modeling

Advanced Data Processing for Foundation Models

- Comprehensive data validation and quality assurance
- Multi-modal data processing pipelines
- Input optimization and performance enhancement

Vector Databases and Retrieval Augmentation

- Enterprise vector database architecture
- Advanced document processing and chunking strategies
- Sophisticated retrieval system implementation
- Hands-on Lab: Develop Retrieval Augmented Generation (RAG) Applications with Amazon
- Bedrock Knowledge Bases

Prompt Engineering and Governance

- Advanced prompt engineering frameworks
- Complex prompt orchestration systems
- Enterprise prompt governance and management

- Hands-on Lab: Develop conversation pattern with Amazon Bedrock APIs

Agentic AI and Tool Integration

- Agentic AI architecture and evolution
- Amazon Bedrock Agents implementation
- AWS Agentic AI service ecosystem
- Tool integration and production observability

AI Safety and Security

- Comprehensive content safety implementation
- Privacy-preserving AI architecture
- AI governance and compliance frameworks

Performance Optimization and Cost Management

- Token efficiency and cost optimization
- High-performance system architecture
- Intelligent caching systems implementation
- Hands-on Lab: Building Secure and Responsible Gen AI with Guardrails for Amazon Bedrock

Monitoring and Observability for Generative AI

- Foundation model monitoring systems
- Business impact and value management
- AI-specific troubleshooting and diagnostics

Testing, Validation, and Continuous Improvement

- Comprehensive AI evaluation frameworks
- Quality assurance and continuous improvement
- RAG system evaluation and optimization

Enterprise Integration Patterns

- Enterprise connectivity and integration architecture
- Secure access and identity management
- Cross-environment and hybrid deployments

Course wrap-up

- Next steps and additional resources
- Course summary