Oracle Cloud Infrastructure Generative AI Professional

Course Duration: 16 Hours **Course code: 1Z0-1127-25**

1. Course Overview

This course provides an in-depth understanding of Generative AI services in Oracle Cloud Infrastructure (OCI). Participants will learn how to leverage pretrained large language models (LLMs), fine-tune models, and integrate generative AI into enterprise applications. The training covers prompt engineering, embeddings, vector databases, multimodal AI, and model deployment on OCI AI Services. Learners will also gain hands-on experience with security, governance, monitoring, and cost optimization for Generative AI workloads in OCI. By the end of the course, participants will be able to design, implement, and manage end-to-end generative AI solutions on OCI.

2. What you'll learn?

By the end of this course, you should be able to:

- Describe OCI Generative AI architecture and services
- Understand the foundation of LLMs and embeddings
- Apply prompt engineering techniques for business scenarios
- Fine-tune generative AI models in OCI
- Deploy and integrate AI models with OCI Data Science, Functions, and Applications
- Store and retrieve embeddings with OCI Vector Databases
- Implement security, IAM policies, and governance for AI workloads
- Monitor, scale, and optimize generative AI solutions in the cloud
- Explore enterprise use cases for Generative AI in OCI

3. Target Audience

- AI/ML Engineers and Data Scientists
- Cloud Architects and Developers
- Enterprise IT Professionals adopting Generative AI on OCI

Phone: +91-999-911-1686 **Mail Us:** info@ssdntech.com

- Business and Technology Leaders evaluating AI-powered solutions

4. Pre-Requisites

Familiarity with:

- Oracle Cloud Infrastructure (OCI) fundamentals
- Basics of AI/ML concepts and neural networks
- Python programming (preferred)
- Understanding of REST APIs and cloud deployment

5. Course content

Module 1: Course Introduction

- Introduction
- Course contents

Module 2: Introduction to Generative AI in OCI

- What is Generative AI?
- Generative AI use cases in enterprises
- Overview of OCI AI Services and ecosystem

Module 3: OCI Generative AI Architecture

- Foundation models in OCI
- Key services and components
- Deployment models: public, private, hybrid

Module 4: Large Language Models (LLMs) in OCI

- Pretrained models and their capabilities
- Fine-tuning vs. prompt-tuning
- Model selection strategies

Module 5: Prompt Engineering

- Basics of prompt engineering

- Crafting effective prompts for accuracy
- Advanced prompting: chain-of-thought, role-based, and multi-step prompts

Module 6: Embeddings and Vector Databases

- What are embeddings?
- Storing and retrieving embeddings with OCI Vector Database
- Semantic search and retrieval-augmented generation (RAG)

Module 7: Deploying Generative AI Models on OCI

- Using OCI Data Science for deployment
- Serverless integration with OCI Functions
- APIs and SDKs for application integration

Module 8: Security and Governance in Generative AI

- IAM policies for AI workloads
- Data privacy and compliance
- Responsible AI principles in OCI

Module 9: Monitoring and Scaling AI Workloads

- Logging, metrics, and diagnostics
- Autoscaling models and workloads
- Performance optimization techniques

Module 10: Integrating Generative AI with OCI Services

- Integration with Oracle Digital Assistant
- Using Generative AI with databases and analytics
- AI-driven automation in enterprise apps

Module 11: Enterprise Use Cases of Generative AI

- Customer service automation
- Knowledge management and semantic search
- Document summarization and content generation

- Industry-specific scenarios (finance, healthcare, retail)

Module 12: Wrap-Up and Best Practices

- Best practices for OCI Generative AI deployments
- Cost optimization strategies

Phone: +91-999-911-1686 **Mail Us:** info@ssdntech.com