Oracle Cloud Infrastructure Generative AI Professional (2025): Hands-on Workshop

Oracle Cloud Infrastructure

DURATION	MODULES	COURSE CODE
3 Days	**10 Lectures**	**—**

Course Overview

Please make sure to subscribe to one of these regions to access the OCI Generative AI service.

What You Will Learn

Module 01: Course Overview & Preparation

- For whom is this course intended?
- Course Outline
- #1: Fundamentals of Large Language Models (LLMs)
- #2: Deep Dive on OCI Generative AI Service
- #3: Implement RAG using OCI Gen AI service + Oracle Database 23ai + LangChain
- #4: Deep Dive on OCI Generative AI Agents Service
- Meet your instructors
- Generative AI Labs
- Measuring Your Progress: Skill Checks
- Ask Your Instructor Form / OU Community
- Learning and Exam Tips

Module 02: Fundamentals of Large Language Models (LLMs)

- Introduction to LLMs
- What is a Large Language Model?
- This Module Overview
- LLM Architectures
- Encoders and Decoders
- Model Ontology
- Encoders
- Decoders

- Encoders-Decoders
- Architectures at a Glance
- Prompting and Prompt Engineering
- Affecting the Distribution over Vocabulary
- Prompting Basics
- Prompt Engineering Techniques
- In-context Learning and Few-shot Prompting
- Example Prompts
- Advanced Prompting Strategies
- Issues with Prompting (Prompt Injection, Memorization)
- Training LLMs
- Training Overview
- Hardware Costs
- Decoding
- Greedy Decoding
- Non-Deterministic Decoding
- Temperature
- LLM Challenges
- Hallucination
- Groundedness and Attributability
- LLM Applications
- Retrieval Augmented Generation (RAG)
- Code Models
- Multi-Modal Models
- Language Agents

Module 03: OCI Generative AI Service

- OCI Generative AI Introduction
- How OCI Generative AI Service Works
- Pretrained Foundational Models
- Fine-tuning
- Dedicated AI Clusters
- Hands-on Demos:
- Generative AI Service Walkthrough
- Chat Models & Parameters
- Tokens
- Pretrained Chat Models
- Preamble Override
- Temperature, Top-k, Top-p
- Frequency & Presence Penalties
- OCI Generative AI Inference API
- OCI Config Setup for Generative AI API

Module 04: Embedding Models

- Embeddings Overview

- Word Embeddings
- Sentence Embeddings
- Semantic Similarity
- Embeddings Use Cases
- Hands-on Demos:
- Embedding Models in Generative AI
- Demo: Embedding Model

Module 05: Prompt Engineering & Customization

- Prompt Engineering in LLMs
- LLMs as Next Word Predictors
- Aligning LLMs to Follow Instructions
- In-context Learning / Few-shot Prompting
- Prompt Formats & Advanced Strategies
- Customize LLMs with Your Data
- Training LLMs from Scratch vs Fine-tuning Pretrained Models
- Fine-tuning Benefits
- RAG Overview
- Fine-tuning and Inference Workflows in OCI Generative AI
- T-Few Fine-tuning Process
- Reducing Inference Costs

Module 06: Dedicated AI Clusters & Fine-tuning

- Dedicated AI Clusters Overview
- Sizing and Pricing
- Cluster Unit Types and Sizing
- Pricing Examples
- Hands-on Demos:
- Dedicated AI Clusters Setup
- Generative AI Fine-tuning Configuration
- Fine-tuning Parameters (T-Few)
- Understanding Fine-tuning Results (Accuracy, Loss)
- Hands-on Demos:
- Fine-tuning & Custom Models
- Inference Using Endpoint

Module 07: OCI Generative AI Security

- Dedicated GPU and RDMA Network
- Model Endpoints
- Customer Data and Model Isolation
- Leveraging OCI Security Services

Module 08: OCI Generative AI Integrations

- OCI Generative AI and LangChain Integration

- LangChain Components
- LangChain Models
- LangChain Prompt Templates
- LangChain Chains
- Prompt, Model, and Chain Interaction
- LangChain Memory
- OCI Generative AI and Oracle 23ai Integration

Module 09: Retrieval Augmented Generation (RAG)

- RAG Overview and Pipeline
- Document Processing
- Document Loading, Chunking, Chunk Size, Overlap
- Splitting Methods
- Embed and Store Documents
- Embeddings Capture Semantic Relationships
- Vector Embeddings & Data Types
- Connect to Vector Database
- Add Metadata to Chunks & Create Documents
- Retrieve Documents & Generate Response
- Comparing Vectors (Embedding Distance)
- Vector Indexes for Faster Search
- Augmented Generation Workflow
- Conversational RAG (Chatbot Integration)

Module 10: OCI Generative AI Agents

- OCI Generative AI Agents Overview
- Agents Architecture
- Agent Concepts
- Guidelines for Knowledge Base
- Object Storage
- Oracle Database 23ai
- Agent Components
- Endpoint
- Chat
- Agent Limitations