Generate Smarter Generative AI Outputs on Google Cloud

Course Duration: 40 Hours **Course code: GSGAOGC**

1. Course Overview

This course is designed to help learners generate high-quality, accurate, and context-aware outputs using Generative AI on Google Cloud. It focuses on advanced prompt engineering, grounding techniques, and optimization strategies using Vertex AI and foundation models. Learners will explore how to reduce hallucinations, improve relevance, and build intelligent AI-powered applications for real-world business use cases.

2. What you'll learn?

By the end of the course, you will be able to:

- Understand how Generative AI models generate responses
- Design effective prompts for high-quality outputs
- Apply advanced prompt engineering techniques
- Use grounding and Retrieval-Augmented Generation (RAG)
- Improve accuracy and reduce hallucinations
- Optimize outputs for business and technical use cases
- Integrate generative AI into applications using Vertex AI
- Evaluate and refine AI-generated responses

3. Target Audience

- AI/ML Engineers
- Developers and Cloud Engineers
- Data Scientists
- Business Analysts and Product Managers

- Anyone working with Generative AI tools

4. Pre-Requisites

Before taking this course, you should have:

- Basic understanding of AI/ML concepts
- Familiarity with Google Cloud Platform (GCP)
- Basic programming knowledge (Python recommended)
- Understanding of APIs and data handling

5. Course content

Module 1: Course Introduction

- Course overview and objectives
- Introduction to Generative AI on Google Cloud
- Business use cases and applications

Module 2: Foundations of Generative AI

- Overview of Large Language Models (LLMs)
- How generative models work
- Tokenization and context windows
- Limitations and challenges

Module 3: Google Cloud Generative AI Ecosystem

- Vertex AI overview
- Generative AI Studio
- Foundation models and APIs
- Integration with GCP services

Module 4: Prompt Engineering Fundamentals

- Structure of effective prompts
- Role, context, and instruction design

- Output formatting techniques
- Common prompting mistakes

Module 5: Advanced Prompt Engineering

- Zero-shot, few-shot, and chain-of-thought prompting
- Role-based and scenario-based prompts
- Iterative prompting techniques
- Prompt templates for different use cases

Module 6: Improving Output Quality

- Controlling tone, style, and format
- Ensuring consistency and accuracy
- Handling ambiguous queries
- Output validation techniques

Module 7: Grounding and RAG (Retrieval-Augmented Generation)

- Concept of grounding
- Implementing RAG with enterprise data
- Using embeddings and vector search
- Improving factual correctness

Module 8: Reducing Hallucinations

- Causes of hallucinations
- Detection techniques
- Mitigation strategies
- Designing safe prompts

Module 9: Integrating Generative AI into Applications

- Using Vertex AI APIs
- Building AI-powered applications

- Backend integration
- Real-time AI interactions

Module 10: Evaluation and Testing

- Evaluating AI-generated outputs
- Metrics for quality and relevance
- A/B testing prompts
- Continuous improvement

Module 11: Personalization and Context Management

- Managing conversation context
- Personalizing outputs
- Session handling
- Multi-turn interactions

Module 12: Security and Responsible AI

- Data privacy and protection
- Responsible AI practices
- Ethical considerations
- Compliance and governance

Module 13: Automation and Workflow Optimization

- Automating tasks using GenAI
- Workflow design with AI
- Integration with other tools
- Event-driven AI workflows

Module 14: Advanced Use Cases

- Content generation (marketing, blogs)
- Code generation and debugging

- Customer support automation
- Knowledge assistants

Module 15: Real-World Projects and Capstone

- Building a GenAI-powered application
- Implementing RAG-based solution
- Best practices and optimization
- Final project and evaluation