Application Development with LLMs on Google Cloud

In this course, you'll dive into the details of using Large Language Models (LLMs) in your applications. You'll start by exploring the core principles that underpin prompting LLMs. Next, you will focus on Google's latest family of models, Gemini. You'll explore the various Gemini models and their multimodal capabilities. This includes a deep dive into effective prompt design and engineering within the Vertex AI Studio environment. Then, the course moves to application development frameworks and how to implement these concepts into your applications.

DURATION

1 day

LEVEL

Intermediate

FORMAT

Instructor-led

What you'll learn

- Explore the different options available for using generative AI on Google Cloud.
- Use Vertex AI Studio to test prompts for large language models.
- Develop LLM-powered applications using generative AI
- Apply advanced prompt engineering techniques to improve the output from LLMs
- Build a multi-turn chat application using the Gemini API and LangChain

Overview	5 modules · 4 labs

Who this course is for	• Customers, Partners

Products	• Vertex AI
	• Vertex AI Studio
	• Vertex AI Workbench

Prerequisites	• Completion of "Introduction to Developer Efficiency on Google Cloud" or equivalent knowledge.

Not covered	• This course does have a deep dive on topics around developing agents using Agent Development Kit and deployment of agents on Google Cloud using solutions such as Agent Engine.

Module 01	Introduction to Generative AI on Google Cloud

Topics	• What is generative AI
	• Vertex AI on Google Cloud
	• Generative AI options on Google Cloud
	• Introduction to course use case

Module 02	Vertex AI Studio

Topics	• Introduction to Vertex AI Studio
	• Designing and testing prompts
	• Data governance in Vertex AI Studio

Activities	• Lab: Getting Started with the Vertex AI Studio User Interface

Module 03   Generative AI Fundamentals

Topics
- Introduction to grounding
- Integrating the Vertex AI Gemini APIs
- Chat, memory and grounding
- Search principles

Activities
- Lab: Getting Started with LangChain + Vertex AI Gemini API

Module 04   Prompt Engineering

Topics
- Review of few-shot prompting
- Chain-of-thought prompting and thinking budgets
- Meta prompting, multi-step, and panel prompts
- RAG and ReAct

Activities
- Lab: Advanced Prompt Architectures

Module 05   Creating Custom Chat Applications with Vertex AI Gemini API

Topics
- LangChain for chatbots
- ADK for chatbots
- Chat retrieval

Activities
- Lab: Implementing RAG Using LangChain