Google Cloud

Deep Dive: Advanced Generative AI

Generative AI is a rapidly evolving field of artificial intelligence with a wide range of applications across all verticals. Generative AI models typically use deep learning techniques to process vast amounts of data and identify underlying relationships to generate original content that is relevant and engaging. This course will equip participants with practical expertise to design, tune, and deploy production Generative AI solutions.

DURATION
Deep dive with capstone projects: 10 full days or 15 half days

Deep dive, content only: 5 full days or 7 half days

LEVEL
Advanced

FORMAT
Instructor-led

What you'll learn

- Overview of generative AI
- Google Cloud generative AI offerings
- Vertex AI tools (Gemini API, Model Garden) to build LLM solutions
- Transformer-based LLMs
- Image generation models

Overview	4 core modules, 3 elective modules
Who this course is for	This course is designed for technical teams with a foundational understanding of machine learning (ML) concepts who are comfortable programming in Python. Participants aspire to work on generative AI projects and want to develop practical skills using Google Cloud and open source tools for building applications, fine-tuning LLMs, or creating image generation models.

Module 01   Generative AI and Prompt Design on Vertex AI

Topics
- Introduction to generative AI
- Large language model overview
- Overview of Vertex AI and generative AI tools
- Prompt design, PaLM, and Gemini API

Module 02   Building Advanced Generative AI Systems

Topics
- LangChain and Vertex AI Vector Search
- Retrieval-augmented generation (RAG) using LangChain and Vertex AI Agent Builder
- Function calling
- Fine-tuning LLMs on Generative AI Studio on Vertex AI
- Distilling student models from larger foundation models

Module 03   Guided Project

Objectives
- Apply everything learned from previous models to build an end-to-end solution using a structured guided project.

Deep-dive-only format where no capstone project is undertaken.

Module 04   Immersive Capstone Project (Additional Week)

Objectives

Working alongside Google Advanced Solutions Lab (ASL) ML solution engineers, participants roll up their sleeves and utilize generative AI to tackle real world business challenges. This option requires a 10-full-day, or 15-half-day, format and supplants Module 3. A typical virtual cohort will organize into 3-4 project teams,

each working to develop end-to-end POC solutions for their chosen Gen AI business use case (1-2 project teams per cohort is more typical for in-person learning at Google). At the conclusion of the training, the teams present results and learnings to ASL instructors, colleagues, and senior stakeholders.

Elective modules

Where topics align with customer objectives, chosen class format, and available time.

Module 05   **Large Language Models (LLMs) and Text Generation with Keras**

Topics
- Transformer encoder and BERT
- Transformer decoder and text generation
- Overview of advanced topics: caveats of LLMs and evaluation techniques

Objectives
Learn the intuition behind large language models (LLMs) and how to use Vertex AI to build custom generative AI models with Keras.

Module 06   **Modern Image Generation**

Topics
- Overview of image generation
- Generative adversarial networks (GANs)
- Diffusion models
- Fine-tuning stable diffusion

Objectives
Explore in detail the diffusion model, a state-of-the-art image generation algorithm, and learn how to fine-tune foundational models on Vertex AI.

Module 07   **Orchestrating Systems with Vertex AI Pipelines**

Topics
- Orchestrate model training and deployment with KFP and Vertex AI.
- Use custom containers with Vertex AI training.
- Build and push a training container.
- Train, tune, serve, and query a model.
- Describe a KFP pipeline using KF domain-specific language (DSL).
- Build pipelines using pre-built and lightweight components.

Objectives
Learn how to orchestrate and automate AI project pipelines using Kubeflow Pipelines (KFP), Vertex AI, and Vertex AI Pipelines.