Generative AI and Responsible AI Practices with Google Cloud

Duration: 05 days (40 hours)

Courseware: Unofficial

Module 1: Introduction to Image Generation

This module introduces diffusion models, a family of machine learning models that recently showed promise in the image generation space. Diffusion models draw inspiration from physics, specifically thermodynamics. Within the last few years, diffusion models became popular in both research and industry. Diffusion models underpin many state-of-the-art image generation models and tools on Google Cloud. This module introduces you to the theory behind diffusion models and how to train and deploy them on Vertex AI.

Module 2: Attention Mechanism

This module will introduce you to the attention mechanism, a powerful technique that allows neural networks to focus on specific parts of an input sequence. You will learn how attention works, and how it can be used to improve the performance of a variety of machine learning tasks, including machine translation, text summarization, and question answering.

Module 3: Encoder-Decoder Architecture

This module gives you a synopsis of the encoder-decoder architecture, which is a powerful and prevalent machine learning architecture for sequence-to-sequence tasks such as machine translation, text summarization, and question answering. You learn about the main components of the encoder-decoder architecture and how to train and serve these models. In the corresponding lab walkthrough, you'll code in TensorFlow a simple implementation of the encoder-decoder architecture for poetry generation from the beginning.

Module 4: Transformer Models and BERT Model

This module introduces you to the Transformer architecture and the Bidirectional Encoder Representations from Transformers (BERT) model. You learn about the main components of the Transformer architecture, such as the self-attention mechanism, and how it is used to build the BERT model. You also learn about the different tasks that BERT can be used for, such as text classification, question answering, and natural language inference.

Module 5: Create Image Captioning Models

This module teaches you how to create an image captioning model by using deep learning. You learn about the different components of an image captioning model, such as the encoder and decoder, and how to train and evaluate your model. By the end of this module, you will be able to create your own image captioning models and use them to generate captions for images

Module 6: Introduction to Vertex AI Studio

This module introduces Vertex AI Studio, a tool for prototyping and customizing generative AI models. Through immersive lessons, engaging demos, and a hands-on lab, you'll explore the generative AI workflow and learn how to leverage Vertex AI Studio for Gemini multimodal applications, prompt design, and model tuning. The aim is to enable you to unlock the potentials of these models in your projects with Vertex AI Studio.

Module 7: Vector Search and Embeddings

This module introduces Vertex AI Vector Search and describes how it can be used to build a search application with large language model (LLM) APIs for embeddings. The course consists of conceptual lessons on vector search and text embeddings, practical demos on how to build vector search on Vertex AI, and a hands-on lab.

Module 8: Inspect Rich Documents with Gemini Multimodality and Multimodal RAG

Complete the intermediate **Inspect Rich Documents with Gemini Multimodality and Multimodal RAG** skill badge to demonstrate skills in the following: using multimodal prompts to extract information from text and visual data, generating a video description, and retrieving extra information beyond the video using multimodality with Gemini; building metadata of documents containing text and images, getting all relevant text chunks, and printing citations by using Multimodal Retrieval Augmented Generation (RAG) with Gemini.

Module 9: Responsible AI for Developers: Fairness & Bias

This module introduces concepts of responsible AI and AI principles. It covers techniques to practically identify fairness and bias and mitigate bias in AI/ML practices. It explores practical methods and tools to implement Responsible AI best practices using Google Cloud products and open-source tools.

Module 10: Responsible AI for Developers: Interpretability & Transparency

This module introduces concepts of AI interpretability and transparency. It discusses the importance of AI transparency for developers and engineers. It explores practical methods and tools to help achieve interpretability and transparency in both data and AI models.

Module 11: Responsible AI for Developers: Privacy & Safety

This module introduces important topics of AI privacy and safety. It explores practical methods and tools to implement AI privacy and safety recommended practices through the use of Google Cloud products and open-source tools.

Module 12: Machine Learning Operations (MLOps) for Generative AI

This module is dedicated to equipping you with the knowledge and tools needed to uncover the unique challenges faced by MLOps teams when deploying and managing Generative AI models, and exploring how Vertex AI empowers AI teams to streamline MLOps processes and achieve success in Generative AI projects.