# [Sf=ir] Institute

Google Cloud | GCP200DEVLLM

# Application Development with LLMs on Google Cloud

Intégrer des LLM et l'IA Générative dans vos applications

7h

# Objectifs pédagogiques

- Explorez les différentes options disponibles pour utiliser l'IA générative sur Google Cloud.
- Utilisez Vertex Al Studio pour tester les prompts pour les Large Language Models.
- Développer des applications basées sur LLM à l'aide de modèles LangChain et LLM sur Vertex Al.
- Appliquer des techniques de prompt engineering pour améliorer les résultats des LLM.
- Créez une application de chat multi-tours à l'aide de l'API PaLM et de LangChain.

Modalités d'évaluation : Les objectifs pédagogiques sont évalués à travers la réalisation des parties pratiques (labs dirigés) sous la supervision du formateur délivrant la session de formation.

### Public cible

 Toute personne intéressée par intégrer l'IA Générative et les LLMs dans une application Google Cloud

# Prérequis

Avoir suivi le cours '<u>Introduction to Developer Efficiency on Google Cloud</u>' ou avoir des connaissances équivalentes

## Programme

#### Module 01: Introduction à l'IA générative sur Google Cloud

#### Sujets

- Vertex AI sur Google Cloud
- Options d'IA générative sur Google Cloud
- Introduction au cas d'utilisation du cours

#### Objectifs

 Découvrez les différentes options disponibles pour utiliser l'IA générative sur Google Cloud.

#### Module 02: Vertex Al Studio

#### Sujets

- Présentation de Vertex Al Studio
- Modèles disponibles et cas d'utilisation
- Conception et test de prompts dans la console Google Cloud
- Gouvernance des données dans Vertex Al Studio

#### Objectifs

- Utilisez Vertex Al Studio pour tester les prompts pour les Large Language Models.
- Comprendre comment Vertex Al Studio assure la sécurité de vos données

#### Activités

• Lab: Explorer Vertex Al Studio

#### Module 03: Fondamentaux de LangChain

#### Sujets

- Introduction à LangChain
- Concepts et composants LangChain
- Intégration des API Vertex AI PaLM

• Chaîne de questions/réponses utilisant l'API PaLM

#### Objectifs

- Comprendre les concepts et composants de base de LangChain
- Développer des applications basées sur LLM à l'aide de modèles LangChain et LLM sur Vertex Al.

#### Activités

• Lab: Premiers pas avec l'API LangChain + Vertex Al PaLM

#### Module 04: Prompt Engineering

#### Sujets

- Examen des prompts few-shot
- Approche Chain-of-thought
- Retrieval augmented generation (RAG)
- ReAct

#### Objectifs

- Appliquer des techniques de prompt engineering pour améliorer les résultats des LLM.
- Implémenter une architecture RAG pour ancrer les modèles LLM.

#### Activités

• Atelier: Techniques de Prompt Engineering

#### Module 05 : Création d'applications de chat personnalisées avec l'API Vertex Al PaLM

#### Sujets

- LangChain pour les chatbots
- Mémoire pour les chat multitours
- Récupération de chat

#### Objectifs

• Comprendre le concept de mémoire pour les applications de chat multi-tours. Créez une application de chat multi-tours en utilisant l'API PaLM et LangChain.

#### Activités

• Lab : Implémentation de RAG à l'aide de LangChain