

DSCI-273

# Generative AI on Cloudera



## Course Overview

**Course Type**

Instructor-Led Training

**Level**

Intermediate

**Duration**

2 days

**Platform**

Cloudera on cloud

**Topics Covered**

- Large Language Models (LLMs)
- Prompt Engineering
- Retrieval Augmented Generation (RAG)
- Fine Tuning
- AI Agents

## About This Training

Generative AI (GenAI) and Large Language Models (LLMs) are extremely powerful new tools that are changing every industry. To fully take advantage of GenAI and LLMs, these new capabilities need to be combined with your existing enterprise data. This two-day course teaches how to use Cloudera AI to train, augment, fine tune, and host LLMs to create powerful enterprise AI solutions.

## What Skills You Will Gain

Through lecture and Hands-On exercises, you will learn how to:

- Select the right LLM model for a use case
- Configure a Prompt for an LLM
- Use Retrieval Augmented Generation (RAG)
- Fine Tune an LLM Model with Enterprise Data
- Use the AI Model Registry and host an LLM
- Create an AI Agent with Crew AI

## Who Should Take This Course?

This course is designed for data scientists and machine learning engineers who need to understand how to utilize Cloudera AI to leverage the full power of their enterprise data, generative AI, and Large Language Models and deliver powerful business solutions.

## Other Training that Might Interest You

- Introducing Python
- Introducing Git

DSCI-273

# Generative AI on Cloudera

## Introduction to LLMs

- History of LLMs
- How Transformers Work
- Different Types of LLMs
- Limitations of LLMs

## LLM Model Selection

- How LLMs are Evaluated
- Model Selection by Use Case
- Hugging Face Model Hub
- Hands-On Exercise: Open LLM Leaderboard
- Hands-On Exercise: Can you run it? LLM Version

## Model Registries and Inference Services

- Cloudera AI Model Registry
- AI Inference Service
- Hands-On Exercise: Text Summarization with Amazon Bedrock

## Prompt Engineering

- Components of a Prompt
- Shot Prompting
- Hands-On Exercise: Prompt Engineering with Mistral

## Retrieval Augmented Generation

- Retrieval Augmented Generation (RAG)
- RAG Use Cases
- Hands-On Exercise: LLM Chatbot Augmented with Enterprise Data
- Hands-On Exercise: Intelligent QA Chatbot with NiFi, Pinecode, and Llama2

## Fine Tuning

- Motivation for Fine Tuning
- Principles of Fine Tuning

## Parameter Efficient Tuning

- Limitations of Fine Tuning
- Principles of Parameter Efficient Tuning

## Fine Tuning a Foundation Model

- Quantization
- Low Rank Adaptation
- Hands-On Exercise: Fine Tuning a Foundation Model for Multiple Tasks (with QLoRA)

## AI Agents

- Introduction to AI Agents
- AI Agent Architecture
- Hands-On Exercise: All Your Agents with Crew AI