

**Intermediate Generative AI for Developers**

**Course Number:** AI-122WA  
**Duration:** 2 days

**Overview**

Accelebrate's Intermediate Generative AI (Gen AI) training teaches developers advanced techniques like fine-tuning LLMs, Retrieval Augmented Generation (RAG), and Vector Embeddings. Attendees also learn how to integrate LLMs into development pipelines.

**Prerequisites**

* Practical experience in Python (at least 6 months):
  + Data Structures, Functions, Control Structures
  + Exception Handling, File I/O, async, concurrency (recommended)
* Practical experience with these Python libraries: Pandas, NumPy, and scikit-learn
  + Understanding of Machine Learning concepts - regression, clustering, classification
  + ML Algorithms: Gradient Descent, Linear Regression
* Loss Functions and evaluation metrics

**Materials**

All Generative AI training students receive comprehensive courseware.

**Software Needed on Each Student PC**

All attendees must have a modern web browser and an Internet connection.

**Objectives**

* Master vector embeddings and LLM-powered applications to transform unstructured data into actionable insights
* Optimize data ingestion and retrieval augmented generation (RAG) techniques for enhanced LLM performance and efficiency
* Develop enterprise-grade RAG pipelines that leverage semantic search and LLM chains for tailored solutions
* Build and deploy robust, production-ready LLM applications focusing on scalability, security, and privacy
* Confidently navigate LLM deployment and LLMOps for successful implementation in your organization

**Outline**

* Introduction
* Building LLM-powered Applications
  + Vector Embeddings
  + Ingesting Private Data with LlamaIndex
  + Types of Indexing and Chunking for Data Ingestion
  + Introduction to Retrieval Augmented Generation (RAG)
  + Semantic Search for Code libraries
* LangChain Integration and Advanced RAG
  + LLM Chains and Prompt Templates
  + The LangChain “Tools” Library
  + Enterprise-grade RAG Pipelines
  + RAG Pipeline Optimization and Performance Monitoring
* Enterprise API Applications
  + Generative AI Tech Stack
  + Scalable and Efficient Architectures
  + Privacy/Security Considerations with Enterprise Data
  + Conversational Agents in Enterprise
  + Best Practices for production-ready LLM Applications
  + Enterprise Application Pipelines
  + Choosing the right foundation model
  + Cost and ROI Evaluation Strategy
* LLM Deployment for Developers
  + LLM Deployment Frameworks
  + Introduction to LLMOps for Developers
  + LLM Security Considerations
  + Enterprise Privacy
  + Cloud Deployment vs Local (Private) Serving
* Conclusion