

**Introduction to Generative AI for Developers**

**Course Number:** AI-120WA  
**Duration:** 3 days

**Overview**

This Introduction Generative AI (GenAI) training teaches developers how to build the next generation of intelligent applications. Through hands-on learning and real-world examples, participants master the core concepts of GenAI and large language models (LLMs), gaining the practical skills to develop scalable and innovative software solutions.

**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**

* Define and describe the evolution of artificial intelligence, including traditional AI, machine learning, deep learning, and the emergence of generative AI.
* Master the core concepts of large language models (LLMs), including their architecture, training processes, and applications
* Use prompt engineering techniques to communicate with LLMs and elicit desired outputs effectively
* Access and utilize LLMs through APIs, integrating them into various applications and workflows
* Implement Retrieval Augmented Generation (RAG) to connect LLMs with external knowledge sources for enhanced performance
* Deploy LLMs securely and efficiently, considering factors like privacy, data governance, and cost optimization
* Evaluate the capabilities and limitations of LLMs and their potential impact across industries

**Outline**

* Introduction to Generative AI
  + What is Intelligence?
  + Mechanisms of Intelligence
  + What is Artificial Intelligence?
  + How does AI work?
  + Evolution of AI
  + Applications of AI
  + Traditional AI and Machine Learning
  + Deep Learning: Neural Networks
  + Emergence of Generative AI
  + Transformer Models
* Understanding Large Language Models
  + What is a Language Model?
  + Natural Language Processing (NLP)
  + What are Large Language Models (LLMs)?
  + Tokenization and Word Embeddings
  + Case Study: GPT-3 Model
  + Types and Applications of LLMs
  + Training and Architectures of LLMs
  + Power and Limitations of LLMs
* Introduction to Prompt Engineering
  + Basics of Prompting and Engineering
  + Communicating with Large Language Models
  + Effective Prompting Strategies
  + Advanced Prompt Engineering Concepts
  + Case-Study: GPT-o1
  + Ensembling and Iterative Refinement
  + Resources for Mastering Prompt Engineering
* Accessing LLMs through API
  + Closed Source vs. Open Source LLMs
  + Accessing LLMs via API
  + Utilizing LLM APIs in Applications
  + Prompt Templates and Chaining
  + Producing Structured Outputs
  + Function Calling and LangChain
* Introduction to Retrieval Augmented Generation
  + Introduction and Advantages of RAG
  + RAG Phases – Indexing, Retrieval, and Generation
  + Integrating RAG with Knowledge Bases
  + Vector Stores and Embeddings in RAG
  + Using RAG for Document Summarization
  + Managing Large Documents and Knowledge Bases
  + Evaluating and Scaling RAG Systems
* LLM Deployment
  + Security and Privacy in LLM Deployment
  + Data Governance and Compliance
  + Deploying LLMs for Edge Computing
  + Deploying LLMs with Auto-Scaling Capabilities
  + Cost Management in LLM Deployment
* Conclusion