

Developing Generative AI Applications on AWS

Duration: 2 days

Audience: Professionnel TI

Who is this training for?

Software developers interested in using LLMs without fine-tuning

Training objectives

- Describe generative AI and its alignment with machine learning
- Define the importance of generative AI and explain its potential risks and benefits
- Identify the business value of generative AI use cases
- Discuss the technical basics and key terminology of generative AI
- Explain the planning steps for a generative AI project
- Identify some of the risks and mitigations when using generative AI
- Understand how Amazon Bedrock works
- Familiarize yourself with the basic concepts of Amazon Bedrock
- Recognize the benefits of Amazon Bedrock
- List typical Amazon Bedrock use cases
- Describe the typical architecture associated with an Amazon Bedrock solution
- Understand the cost structure of Amazon Bedrock
- Implement a demo of Amazon Bedrock in the AWS Management Console
- Define prompt engineering and apply best practices
- Identify the types of basic prompt techniques, including zero-shot and few-shot learning
- Apply advanced prompt techniques when necessary for your use case
- Identify prompt techniques that are best suited to specific models
- Identify potential misuses of prompts
- Analyze potential biases in FMs' responses and design prompts that mitigate these biases
- Identify the components of a generative AI application and how to customize an FM
- Describe Amazon Bedrock's core models, inference settings, and core APIs

- Identify Amazon Web Services (AWS) offerings that help monitor, secure, and govern your Amazon Bedrock applications
- Describe how to integrate LangChain with LLMs, Prompt models, strings, chat models, text embedding models, document loaders, retrievals, and agents for Amazon Bedrock
- Describe the architecture patterns that you can implement with Amazon Bedrock to build generative AI applications
- Apply the concepts to build and test best use cases using the various Amazon Bedrock models. LangChain, and the Augmented Generation by Recovery (RAG) approach

Summary

This course is designed to introduce generative artificial intelligence (AI) to people developing software interested in using large language models (LLMs) without fine-tuning. The course provides an overview of generative AI, planning a generative AI project, the basics of Amazon Bedrock, the fundamentals of prompt engineering, and architecture patterns for building generative AI applications using Amazon Bedrock and LangChain.

Course outline

Module 1: Introduction to Generative AI - The Art of Possibility

- Machine Learning Overview
- Generative AI Basics
- Generative AI Use Cases
- Generative AI in Practice
- Risks and Benefits

Module 2: Planning a Generative AI Project

- Fundamentals of Generative AI
- Generative AI in Practice
- Context of Generative AI

- Steps to Planning a Generative AI Project
- Risks and Mitigations

Module 3: Getting Started with Amazon Bedrock

- Introduction to Amazon Bedrock
- Architecture and Use Cases
- How to Use Amazon Bedrock
- Demo: Setting Up Bedrock Access and Using Playgrounds
- Basic Model Basics
- Prompt Engineering Fundamentals
- Basic Prompt Techniques
- Advanced Prompt Techniques
- Model-Specific Prompt Techniques
- Demo: Fine-tuning a Basic Text Prompt
- Addressing Misuse of prompts
- Mitigating Bias
- Demo: Mitigating Image Bias

Module 4: Fundamentals of Prompt Engineering

- Basic Model Basics
- Prompt Engineering Fundamentals
- Basic Prompt Techniques
- Advanced Prompt Techniques
- Model-Specific Prompt Techniques
- Demo: Fine-tuning a Basic Textual Prompt
- Addressing Prompt Misuse
- Mitigating Bias
- Demo: Mitigating Image Bias

Module 5: Amazon Bedrock Application Components

- Generative AI Application Components Overview
- Basic Models and FM Interface
- Working with Datasets and Embeddings
- Demo: Word Embeddings
- Additional Application Components
- Augmented Generation by Recovery (RAG)
- Model Fine-tuning
- Securing Generative AI Applications
- Generative AI Application Architecture

Module 6: Amazon Bedrock Basic Models

- Introduction to Amazon Bedrock Basic Models
- Using Amazon Bedrock FMs for Inference
- Amazon Bedrock Methods
- Data Protection and Auditability
- Demo: Invoking the Bedrock Model for Text Generation Using a Zero-Shot Prompt

Module 7: LangChain

- Optimizing LLM Performance
- Using Models with LangChain
- Building Prompts
- Demo: Bedrock with LangChain using a prompt including context
- Structuring documents with indexes
- Storing and retrieving data with memory
- Using strings to sequence components • Managing external resources with LangChain agents

Module 8: Architecture Patterns

-

Introduction to Architecture Patterns

- Text Summarizing
- Demo: Summarizing Small File Text with Anthropic Claude
- Demo: Abstract Text Summary with Amazon Titan Using LangChain
- Answering Questions
- Demo: Using Amazon Bedrock to Answer Questions
- Chatbot
- Demo: Conversational Interface - Chatbot with AI21 LLM
- Code Generation
- Demo: Using Amazon Bedrock Templates for Code Generation
- LangChain and Agents for Amazon Bedrock
- Demo: Integrating Amazon Bedrock Templates with LangChain Agents

Approach and methodology

Practical and structured approach combining focused theory and guided workshops. Participants gradually develop generative AI applications on AWS through real-world exercises inspired by professional scenarios, promoting immediate application of learning. They learn how to integrate generative AI services, design intelligent solutions, and leverage the capabilities of the AWS Cloud to build innovative applications. Led by a certified trainer, the training focuses on interactivity and the development of directly transferable technical skills to design and deploy AI solutions in a professional context.

Prerequisites

- Have completed AWS Technical Essentials or equivalent content
- Intermediate proficiency in Python

Recommendations

- Solid foundation in programming (Python recommended)
-

Understanding of artificial intelligence and machine learning concepts

- Familiarity with APIs, cloud services and application integration
- Knowledge of AWS services (EC2, S3, SageMaker, Lambda – strong asset)
- Notions of generative AI and language models (prompts, embeddings – asset)