AWS Certified Generative AI Developer - Professional (AIP-C01)

This expert-level AWS AI training is designed to help you pass the AWS Certified Generative AI Developer - Professional exam (AIP-C01). The course focuses on building, deploying, and operating production-ready generative AI applications using AWS AI services. You'll work with practical architectures like RAG and agentic AI, manage prompts, and learn how to optimize GenAI systems for cost, performance, and responsible use. Although it's not required, you might want to start with the AWS Certified Data Engineer or Machine Learning Engineer Associate certifications before you attempt this AWS Certified Generative AI Developer - Professional course.

CBT Nuggets course material →

STUDY PLAN

WEEK 1

Getting Started

Certification Overview

Learning Goals

Learning Methods

First Things First

Setup a Budget Alert

Setup IAM Identity Center

Validation

Architecting GenAI Solutions

Skill Introduction

AWS Well-Architected Framework

AWS Well-Architected Tool Walkthrough

Our First Project Design

Gathering Our Data

Validation

Data Enrichment with Amazon Bedrock

Skill Introduction

Our Data Enrichment Progression

Setting Up SageMaker Studio

Configuring a Data Catalog

Audio Transcription

Augmenting Data with Amazon Bedrock

Adding Transcriptions to Our Data Catalog

Validation

Document Chunking Strategies

Skill Introduction

What is Chunking?

Deciding on a Chunking Strategy

Chunking Our Data

AWS Glue Visual ETL

Reviewing Our Chunked Data

Validation

Embedding Generation and Model Selection

Skill Introduction

How Does Embedding Work?

How Do Embeddings Enable Similarity?

Embeddings in Action

Comparing Embedding Models

Creating Embeddings for Our Data

Testing Similarity

Validation

Vector Databases with pgvector

Skill Introduction

What's Our Vector, Victor?

Fun with S3 Vector Stores

Setting Up Our RDS Instance

Creating a PostgreSQL pgvector Store

Bulk Loading Our Data

Some Basic Vector Searches

Validation

OpenSearch Vector Store Setup

Skill Introduction

Similarity Algorithms

Search Algorithms for Vector Databases

How Much Will This Thing Cost?

OpenSearch Vector Store Creation

Configuring our Index

Loading Data into OpenSearch

Similarity Searches

Manual Snapshot and Restore

Validation

BM25 and Hybrid Searches

Skill Introduction

Anatomy of a Hybrid Search

BM25 Keyword Search

BM25 Optimization

Reciprocal Rank Fusion

Validation

Query Processing and Reranking

Skill Introduction

OpenSearch Dev Tools

Query Enhancement and Expansion

Cohere Rerank Integration

Complete Retrieval Pipeline

Retrieval Optimization

Validation

RAG Prompt Engineering

Skill Introduction

RAG Fundamentals and CREATE Framework

Key Interface Parameters

Completing the RAG System (Almost)

Inference Configuration

Nova Model Comparison

Validation

RAG System Evaluation

Skill Introduction

RAG Evaluation Methods

RAGAS in Action

Bedrock Console Evaluations

LLM-as-a-Judge

RAGAS for Optimization

Set up a CloudWatch Alarm for RAGAS Metrics

Validation

Amazon Bedrock Knowledge Bases

Skill Introduction

Creating Our Knowledge Base

Testing our Knowledge Base RAG

Custom Generation Prompts

RAG Guardrails

Bedrock RAG Evaluation

Conversations with Our Documents

Teardown

Validation

Getting Started with Agentic AI

Skill Introduction

AWS Perspective on Agentic AI

Gathering Our Data

Building the Spell Lookup Tool

Creating the Bedrock Agent

ReAct Reasoning in Action

Session Memory with DynamoDB

Validation

Prompt Engineering for Agents

Skill Introduction

Bedrock Agent Memory

Prompt Flows

Advanced Agent Prompting

Pre and Post Processing Prompts

Validation

Strands Agents and Multi-Agent Orchestration

Skill Introduction

Adding a Monster Lookup Tool

Building with Strands Agents

Prompt Engineering with Verbose Mode

Multi-Agent Orchestration

AgentSquad

Validation

Amazon Bedrock AgentCore

Skill Introduction

Adding a Dice Roller

Adding a Damage Calculator

Creating Our AgentCore Project

Testing Locally

Deploying to the Cloud

AgentCore Browser Tool

Validation

Model Context Protocol (MCP) Integration

Skill Introduction

Model Context Protocol Interfaces

Build the Pantheon MCP Server

Consume the Pantheon in a Strands Agent

Refactor Pantheon as a Lambda Function

Create an AgentCore Gateway

Consuming the MCP from the Gateway

Validation

Circuit Breaker and Human-in-the-Loop Patterns

Skill Introduction

Redeploying our AgentCore Agent

Circuit Breaker Pattern using Step Functions

Testing our Circuit Breaker

Human-in-the-Loop Pattern in Step Functions

Testing Our Human-in-the-Loop Workflow

Validation

Agentic AI Guardrails

Skill Introduction

Build the Guardrail

Denied Topics, Word Filters, and Prompt Attacks

Protecting PII and Sensitive Data

Grounding and Versioning

Inline Agent Guardrail Testing

Validation

Model and Prompt Versioning

Skill Introduction

S3 Versioning and Replication

Preparing the Discogs Dataset

Lambda Weighted Aliases

Bedrock Prompt Management

Rollback Strategy

Validation

Batch Inferences and Intelligent Routing

Skill Introduction

WEEK 2

Batch Embedding

S3 Vector Setup

Batch Vector Store Loading

Lambda and Prompt Management

Intelligent Routing

Validation

Conversational AI and AI-Assisted Search

Skill Introduction

Framing Conversational AI Options

Rebuilding our Catalog Vector Search

Amazon Lex for Catalog Lookup

Lex Fulfillment Lambda

Testing Our Lex Album Bot

Amazon Q Business

Validation

Operational Monitoring and Cost Optimization

Skill Introduction

EventBridge and the Pipeline Pattern

Cost Savings Opportunities

X-Ray Tracing

Model Invocation Logs

Validation

Production API and Security

Skill Introduction

AWS Networking Overview

What We're Building

VPC, Subnets, and Endpoints

Attaching Our Lambda to Our VPC

API Gateway, API Keys, and Usage Plans

Web Application Firewall

Teardown

Validation

Auditing and Hardening a GenAI Application

Skill Introduction

Setting Up GearBot

Kicking the Tires

PII Discovery with Macie

A Quick Security Evaluation

Bias Detection

Remediation

Teardown

Validation

Last updated 6.10.2026