



**10,000+**  
Students Trained

**15,000+**  
Students Placed

**1000+**  
Placement  
Companies

**15Years**  
of Student Trust

# GEN-AI

DEVELOPER

## COURSE CURRICULUM

# Gen AI Developer Program

COURSE DURATION  
3 MONTHS

SESSION HOURS  
200 HRS

CASE STUDIES  
& PROJECTS

### Who Is This Program For?

- ⇒ Software & Backend Engineers
  - ▶ Test Automation Engineers ▶ Data ▶ Cloud / Solution / Enterprise Architects
- ⇒ DevOps & Platform Engineers
- ⇒ Technical Leads transitioning into AI
- ⇒ Professionals serious about Generative AI as a long-term career

### Who Is This Program For?

- ⇒ We start from first principles and build up to expert-level system design.

### Why This Program Is Different (Critical)

Typical GenAI Courses	This Program
Prompt tricks	System engineering
Typical GenAI Courses	This Program
Chatbots	Enterprise platforms
No evaluation	Evaluation-driven development (EDD)
Toy agents	Controlled Agentic AI (LangGraph)
No governance	Security, HITL & compliance
Vendor-locked	Cloud-agnostic + MCP

GEN-AI

## Who Program Philosophy (How Real AI Is Built) Is This Program For?

Modern GenAI systems fail not because of models — they fail because of poor architecture, lack of evaluation, no cost control, and no governance.

### This program trains you to

- ⇒ Think in systems, not prompts
- ⇒ Design for failure and recovery
- ⇒ Prove AI quality using metrics
- ⇒ Control cost, risk, and scale
- ⇒ Communicate clearly in system design interviews

## Terminology & Mental Model Bootcamp (Naive Onboarding)

- ⇒ Before building anything, everyone aligns on language
- ⇒ Generative AI vs ML vs Rule-based systems
- ⇒ Large Language Models (LLMs): what they are & what they are NOT
- ⇒ Tokens, context windows, latency, cost
- ⇒ Prompts, hallucinations, grounding
- ⇒ RAG (Retrieval-Augmented Generation)
- ⇒ Agentic AI (LLM + tools + memory + control loop)
- ⇒ HITL (Human-in-the-Loop)
- ⇒ MCP (Model Context Protocol)
- ⇒ Evaluation, guardrails, LLMOps

This removes fear for beginners and ambiguity for architects

## Core Curriculum (Enterprise-Grade)

### Core Curriculum (Enterprise-Grade)

- ⇒ How LLMs reason (probabilistic, not deterministic)
- ⇒ Where GenAI should and should NOT be used
- ⇒ LLMs as stateless reasoning engines
- ⇒ Separation of concerns: reasoning, data, execution, control

GEN-AI

## LLM Application Engineering

- ⇒ Prompt engineering as software engineering
- ⇒ Prompt templates & versioning
- ⇒ Structured outputs (JSON, schemas)
- ⇒ Tool / function calling
- ⇒ Retry, fallback, and safety patterns

## Retrieval-Augmented Generation (RAG)

- ⇒ Why RAG beats fine-tuning in most enterprises
- ⇒ Embeddings & similarity search
- ⇒ Chunking strategies (fixed, semantic, layout-aware)
- ⇒ Vector databases & metadata filtering
- ⇒ Hybrid search & reranking
- ⇒ Citations, grounding, freshness control
- ⇒ Multi-tenant RAG design

## Multimodal AI (Text + Images + Tables)

- ⇒ Chat with PDFs, images, diagrams
- ⇒ OCR limitations
- ⇒ Table reasoning
- ⇒ Evidence-backed answers

## Agentic AI (Core Differentiator)

**Agentic AI = LLM + tools + memory + control**

- ⇒ Tool-driven agents
- ⇒ Planner-executor patterns
- ⇒ Multi-agent systems
- ⇒ Explicit failure modes:
  - ▶ Infinite loops
  - ▶ Tool hallucination
  - ▶ Cost explosions
  - ▶ Context poisoning

GEN-AI

## **LangGraph (Production-Grade Agent Framework)**

- ⇒ Why LangGraph (determinism, safety)
- ⇒ Nodes, edges, state
- ⇒ Conditional routing
- ⇒ Interrupt & resume
- ⇒ Human approval checkpoints
- ⇒ Supervisor–worker agents

## **Memory in AI Systems**

- ⇒ Short-term vs long-term memory
- ⇒ Episodic vs semantic memory
- ⇒ Retrieval strategies
- ⇒ Privacy & eviction policies

## **Human-in-the-Loop (HITL)**

- ⇒ Why enterprises require HITL
- ⇒ Approval, review, escalation
- ⇒ Audit trails
- ⇒ Feedback ingestion
- ⇒ LangGraph-based HITL workflows

## **Model Context Protocol (MCP)**

- ⇒ Why MCP exists
- ⇒ MCP vs REST vs plugins
- ⇒ Tools, resources, prompts
- ⇒ Sessions & transports
- ⇒ Enterprise MCP gateways
- ⇒ Governance & security boundaries

GEN-AI

## What MOST Courses Miss (We Don't)

### Evaluation & Benchmarking (EDD)

How do you know your AI system works?"

- ⇒ Ground truth creation
- ⇒ Offline & online evaluation
- ⇒ Precision@k / Recall@k
- ⇒ Faithfulness & relevance
- ⇒ Human scoring rubrics
- ⇒ Regression detection
- ⇒ Cost vs quality curves

### Failure Engineering & Guardrails

What happens when AI fails?"

- ⇒ Agent runaway detection
- ⇒ Circuit breakers & kill switches
- ⇒ Budget guards
- ⇒ Safe degradation (AI rules)
- ⇒ Timeout & partial failure handling

### Data Engineering for GenAI

- ⇒ Document ingestion pipelines
- ⇒ Incremental re-indexing
- ⇒ Change data capture (CDC)
- ⇒ Corpus versioning
- ⇒ Data lineage for AI

### AI Cost Engineering

- ⇒ Token budgeting
- ⇒ Model tier routing
- ⇒ Cached responses
- ⇒ Retrieval vs generation trade-offs
- ⇒ Preventing cost explosions

GEN-AI

## Enterprise AI Platform Architecture

- ⇒ AI gateways
- ⇒ Prompt registries
- ⇒ Evaluation services
- ⇒ MCP governance
- ⇒ Observability & cost control
- ⇒ Internal AI marketplaces

## LLMOps / Cloud-Native Deployment

- ⇒ Prompt & model lifecycle management
- ⇒ Observability (latency, cost, drift)
- ⇒ CI/CD for AI systems
- ⇒ Docker & Kubernetes patterns
- ⇒ AWS / Azure / GCP architectures
- ⇒ Security & secrets management

## Security, Privacy & Responsible AI

- ⇒ Prompt injection threats
- ⇒ Tool abuse
- ⇒ Data exfiltration risks
- ⇒ Compliance (GDPR, audits)
- ⇒ Explainability & transparency

## GenAI System Design Interview Readiness

You will practice:

- ⇒ Designing enterprise RAG systems
- ⇒ Designing safe autonomous agents
- ⇒ Designing MCP-based platforms
- ⇒ Explaining trade-offs under pressure
- ⇒ Handling failure & cost questions

GEN-AI

## Capstone Projects (Mandatory)

### Developer Capstones

- ⇒ RAG-based knowledge copilot
- ⇒ LangGraph agent automation
- ⇒ MCP server ecosystem

### Architect Capstones

- ⇒ Enterprise AI platform design
- ⇒ Governance & cost model
- ⇒ Migration strategy (non-AI → AI)

### Joint Capstone

- ⇒ Architects design
- ⇒ Developers implement  
(simulates real organizations)

### Career Outcomes

Graduates are ready for global roles:

- ⇒ Generative AI Engineer
- ⇒ LLM Engineer
- ⇒ Agent Engineer
- ⇒ AI Platform Engineer
- ⇒ AI Architect

You graduate with:

- ⇒ A portfolio
- ⇒ System design confidence
- ⇒ Interview-ready explanations
- ⇒ Enterprise credibility

## Final Takeaway

It teaches you: ▶ How real GenAI systems are built ▶ How they fail  
▶ How they are evaluated ▶ How they are governed ▶ How they are scaled globally

GEN-AI

## Python Foundations

### 1. Python Basics

- ⇒ Data types (strings, lists, tuples, dicts), control flow, functions
- ⇒ OOP concepts (optional overview)

### 2. Environment Setup

- ⇒ Virtual environments (venv, conda)
- ⇒ Basic Git & GitHub for version control

### 3. Data Handling

- ⇒ File I/O (CSV, JSON)
- ⇒ Simple reading/writing with Pandas

### Hands-On Lab

- ⇒ Lab 1: Build a Python script to parse a CSV and produce basic summary stats.
- ⇒ Push code to GitHub.

**Outcome:** Students establish a solid Python base, environment management, and Git.

## Python Data Structures & ML Basics

### 1. Advanced Python Data Structures

- ⇒ List/dict comprehensions, sets, decorators (optional)

## FastAPI Essentials

### 1. FastAPI Introduction

- ⇒ RESTful API concept, asynchronous I/O
- ⇒ Core features (routers, Pydantic models)

### 2. Basic CRUD

- ⇒ Handling GET/POST/PUT/DELETE
- ⇒ Path & query parameters

### 3. API Security

- ⇒ Oauth2 basics, JWT tokens (high-level overview)

### Hands-On Lab

- ⇒ **Lab 3:** Create a basic FastAPI app with CRUD endpoints for a mock resource (e.g., “employees”).
- ⇒ Test endpoints using Postman or curl.

**Outcome:** Ability to scaffold a REST API in Python using FastAPI.

## FastAPI Deep Dive & Unit Testing

### 1. Pydantic Models & Validation

⇒ Request/response schemas, data validation

### 2. Unit Testing

⇒ Pytest basics, mocking & fixtures

⇒ Testing FastAPI endpoints

### 3. Deployment Packaging

⇒ Requirements.txt or Pipfile

⇒ Dockerfile basics (intro only)

### Hands-On Lab

⇒ **Lab 4:** Extend the Fast API app with validated data models and write pytest unit tests.

⇒ Optionally create a Docker file for local container testing.

**Outcome:** Students can build a tested, container-ready API with validated schemas

## Deploying Python APIs on AWS

### 1. AWS Overview

⇒ **Key services:** AWS ECS, AWS Elastic Beanstalk, AWS Lambda + API Gateway

⇒ IAM basics (roles, policies)

### 2. Docker & AWS Deployment

⇒ Building Docker images

⇒ Pushing to AWS ECR (Elastic Container Registry)

⇒ Running containers on ECS or Elastic Beanstalk

### 3. Serverless (Optional)

⇒ Brief mention of AWS Lambda

### Hands-On Lab

⇒ **Lab 5:** Containerize the FastAPI application and deploy to AWS ECS.

⇒ Validate endpoints via public URL.

**Outcome:** Real-world exposure to container-based deployment on AWS.

## Deploying Python APIs on Azure

### 1. Azure Basics

- ⇒ Resource groups, Azure Container Registry (ACR)
- ⇒ Azure Web App for Containers or Azure Container Instances (ACI)

### 2. Deployment Pipeline

- ⇒ Docker image push/pull from ACR
- ⇒ Setting environment variables & app settings

### 3. Monitoring & Logging

- ⇒ Azure Monitor, App Insights for logs and metrics

### Hands-On Lab

- ⇒ **Lab 6:** Deploy the same Dockerized FastAPI app to Azure Web App for Containers.
- ⇒ Review logs and basic metrics in Azure portal.

**Outcome:** Students learn how to replicate a container deployment process on Azure.

## Deploying Python APIs on GCP

### 1. GCP Overview

- ⇒ Services: Cloud Run, GKE, App Engine

### 2. Cloud Run Deployment

- ⇒ Containerizing (Docker)
- ⇒ Submitting images to Google Container Registry
- ⇒ Configuring environment variables and concurrency

### 3. CI/CD

- ⇒ GitHub Actions or GCP Cloud Build integration

### Hands-On Lab

- ⇒ **Lab 7:** Deploy the same FastAPI container to GCP Cloud Run.
- ⇒ Validate auto-scaling behavior by sending multiple requests.
- ⇒ **Outcome:** Students see how to deploy container-based Python services on GCP.

## Deploying the ML Model on AWS, Azure & GCP

### 1. Model Serving Approaches

- ⇒ Containerizing an ML model + FastAPI into one container
- ⇒ Dealing with environment & memory constraints (esp. for large models)

### 2. Cloud-Specific Details

- ⇒ AWS: ECS or SageMaker (intro only)
- ⇒ Azure: Web App for Containers + model storage
- ⇒ GCP: Cloud Run or Vertex AI (intro only)

### Hands-On Lab

- ⇒ **Lab 9:** Deploy the ML + FastAPI container to each cloud (or pick your favorite).
- ⇒ Validate that your inference endpoint is accessible, stable, and logs performance metrics.

**Outcome:** Students learn to host a simple ML model on AWS, Azure, and GCP, reinforcing multi-cloud concepts with a working example.

## Summary of Curriculum

- ⇒ **Python & FastAPI:** Students learn Python basics and build robust REST APIs.
- ⇒ **Testing & Containerization:** Master unit testing (pytest) and Docker packaging.
- ⇒ **Cloud Deployments:** Deploy the same containerized application to AWS, Azure, and GCP.



## Our Students Are Placed In

						
						
						
						
						
						
						
						
						
						