AI-Native CI/CD, MLOps & LLMOps on GPU Platforms

Build secure pipelines for software and AI models, then serve LLMs on GPU-backed Kubernetes with full observability.

DURATION	FORMAT	LEVEL	LABS
120 hours	Onsite, online or hybrid	Intermediate	Cloud sandbox included

About this programme

A 120-hour hands-on programme that takes engineers from Git and GitHub Actions to production platforms for AI workloads: DevSecOps quality gates, secrets and policy as code, progressive delivery on Kubernetes, GPU model serving, LLM observability and evaluation-driven MLOps. GenAI is used across the delivery lifecycle rather than saved for a final module, and every module ends with a lab and a take-home assignment. Teams finish with a capstone, including an optional GPU serving track.

Who it is for: Software, DevOps and platform engineers moving into AI delivery, and ML engineers who need production pipelines

What your team will be able to do

- Design GitHub Actions pipelines with reusable workflows, quality gates and supply-chain controls such as SBOMs and signed artefacts
- Secure pipelines and AI applications with Vault, OPA policies and guardrails against prompt injection and data leakage
- Release safely with blue-green, canary and GitOps rollouts on Kubernetes
- Serve open models on GPU nodes with vLLM, KServe or NVIDIA NIM, and schedule GPUs efficiently
- Trace, cost and evaluate LLM applications with OpenTelemetry, Langfuse, Ragas and DeepEval
- Run versioned model and prompt lifecycles with MLflow and automated retraining

Before you start

- Working knowledge of one programming language such as Java, Python, C# or JavaScript
- A basic grasp of the software development lifecycle and reading and debugging code

Course outline

We send the module-by-module outline with your proposal, adapted to your team's tools and goals.

Assessment and certification

How learning is checked

- A short knowledge check on day one and again at the end, so you can see each person's progress
- Graded hands-on labs in every module, reviewed by the trainer
- A final exercise that applies the programme to a problem like your team's own

Certificate and exams

- Participants who complete the programme receive a Cognitive Networks AI certificate of completion.

Formats and cohort size

- Onsite: At your office or a venue you choose, in full days.
- Online, live: Instructor-led over video, usually in half-day sessions so people can keep up with work.
- Hybrid: Some days in a room together, the rest online. Useful when a team works across cities.

Cohort size: 15 to 25 people per trainer. Larger teams run as parallel cohorts on the same outline, so everyone learns the same thing.

Pricing: We quote per cohort, based on duration, format and number of cohorts. There is no per-seat price list: tell us your team and we send a proposal within one business day.

Bring this programme to your team

Request a proposal at cognitivenetworks.ai/contact?programme=ai-cicd-mlops-gpu

reachus@cognitivenetworks.ai · Bangalore +91 72043 05303 · Dallas +1 (303) 827-7415