Rathinam Trainers & Consultants Private Limited

The AI Engineering Series · Program 02 · 2026

Master Agentic AI

Master agentic AI, RAG pipelines, multi-agent orchestration, fine-tuning, guardrails, and cloud deployment — in 16 weeks.

4	4	16	9
MODULES	CLOUD PLATFORMS	WEEKS	SUBSYSTEMS

THE PROBLEM

Who Builds the AI Systems Everyone Uses?

AI tools are everywhere — but the engineers who can design, deploy, fine-tune, evaluate, and monitor production AI systems are critically scarce. Every company wants a RAG pipeline, an agent platform, an AI assistant. The tools exist. The cloud platforms exist. The engineers who can operate them end-to-end do not.

> **The real skill gap isn't using AI tools. It's building and running the AI platforms that power them.**
> This program closes that gap with hands-on experience building a complete AI platform across four cloud environments.

WHAT YOU WILL BE ABLE TO DO

Learning Outcomes

- ✓ Design and implement RAG pipelines with retrieval, reranking, and grounded generation
- ✓ Fine-tune open-source models using LoRA/QLoRA with tracked experiments
- ✓ Implement guardrails for PII detection, content filtering, and hallucination check
- ✓ Track ML experiments with MLflow and implement CI/CD for models
- ✓ Route queries between SLMs and LLMs for cost-optimized inference
- ✓ Build multi-agent systems with tool use, memory, and orchestration
- ✓ Deploy AI applications on all four major cloud AI platforms
- ✓ Build evaluation pipelines with RAGAS metrics and A/B testing
- ✓ Monitor AI systems with latency tracking, drift detection, and auto-scaling
- ✓ Make informed cloud AI platform recommendations based on hands-on experience

WHY THIS PROGRAM

Five Things No Other Program Does

01

All Four Cloud Platforms

Most programs teach one cloud. You deploy on Vertex AI, Bedrock, Azure AI Studio, and OpenShift AI — and understand the trade-offs between managed, serverless, and self-hosted.

02

End-to-End, Not Just Models

From document ingestion to production monitoring. RAG, agents, fine-tuning, guardrails, evaluation, MLOps, AIOps — the complete agentic AI lifecycle.

03

One Real Project Throughout

SageDesk grows with each module. Every concept has a concrete, working implementation — not isolated notebooks that go nowhere.

04

Production Practices from Day One

Cost tracking, tenant isolation, model versioning, guardrails, drift detection — production concerns are woven into every module, not bolted on at the end.

05

Open-Source First

No vendor lock-in. Local development with Ollama and open-source models. MLflow and Langfuse are self-hosted. You own every piece of the stack. Cloud platforms are deployment targets, not dependencies.

MULTI-CLOUD

Four Cloud Platforms. Complete Deployment.

Every student deploys SageDesk on all four platforms — not just one.

GCP Vertex AI

Managed AI on Google Cloud. Model Garden, Vertex AI Endpoints, Vector Search, Vertex AI Training. Best for: teams already on GCP wanting Gemini models.

AWS Bedrock

Serverless AI on AWS. Foundation Models, Knowledge Bases, Bedrock Agents, Bedrock Guardrails. Best for: broadest model selection, zero infrastructure.

Azure AI Studio

Enterprise AI on Azure. Azure OpenAI, AI Search, Content Safety, Prompt Flow. Best for: GPT-5 with enterprise SLAs.

OpenShift AI

Self-managed AI on Kubernetes. KServe, vLLM, OpenShift Pipelines, Prometheus + Grafana. Best for: full control, air-gapped, open-source only.

WHO SHOULD ATTEND

Built for Agentic AI Engineers

From ML engineers who want production skills to platform engineers adding AI infrastructure — this program delivers hands-on value at every level.

ML Engineers

Move beyond model training — learn to deploy, serve, evaluate, and monitor in production.

Platform Engineers

Add AI infrastructure to your toolkit — model serving, GPU management, auto-scaling.

Backend Developers

Transition into AI engineering — build RAG pipelines, agents, and LLM applications.

Solutions Architects

Evaluate cloud AI platforms with hands-on deployment experience across all four.

DevOps Engineers

Extend CI/CD to ML models, implement LLMOps and AIOps production practices.

Engineering Managers

Understand what it takes to build and operate AI platforms — make informed team decisions.

Prerequisites: Intermediate Python. Familiarity with SQL, Git, Docker, and command line. Completion of Master AI Tools (Program 01) or equivalent experience recommended.

CURRICULUM

Four Modules. 16 Weeks. One Complete Platform.

Each module builds a new subsystem of SageDesk — from GenAI foundations to production deployment.

MODULE 1

Weeks 1-4

GenAI Foundations & RAG

LLMs, prompt engineering, embeddings, vector databases, RAG pipelines, reranking, RAGAS evaluation. Build SageDesk's chat interface and full RAG pipeline.

MODULE 2

Weeks 5-8

Agentic AI & Orchestration

LangChain, LlamaIndex, AutoGen, CrewAI, tool calling, multi-agent systems, agent memory and safety. Build SageDesk's agent orchestration.

MODULE 3

Weeks 9-12

Fine-Tuning & Guardrails

LoRA/QLoRA fine-tuning, model router, PII detection, content filtering, hallucination detection, RAGAS evaluation, Langfuse. Build guardrails and evaluation.

MODULE 4

Weeks 13-16

Cloud Deployment & Production

Vertex AI, Bedrock, Azure AI Studio, OpenShift AI, Prometheus, Grafana, monitoring, capstone integration. Full production deployment.

THE FULL TECH STACK

75+ Tools, Frameworks & Platforms

Every tool listed below is used hands-on during the program. All core tools are open source or have verified free tiers.

Application Core

Tool	Version	What It Does
Python	3.14	Primary language — backend, services, ML pipelines
FastAPI	0.115+	Async REST API framework with automatic OpenAPI docs
SQLAlchemy	2.0	Async ORM with mapped_column syntax
Alembic	1.14+	Database schema migrations with version tracking
PostgreSQL	18	Multi-tenant relational database
pgvector	0.3+	Vector similarity search extension for PostgreSQL
pydantic	v2	Request/response validation and serialization
uvicorn	0.34+	High-performance ASGI server
React	19	Chat UI and admin dashboards
TypeScript	5.8	Type-safe frontend codebase
Tailwind CSS	4	Utility-first styling
Vite	7	Build tool and dev server with fast HMR

AI/ML Frameworks

Framework	What It Does	Module
LangChain	RAG chains, retrieval pipelines, agent executors, tool calling	2, 3
LlamaIndex	Query engines, sub-question decomposition, document indexing	2, 3
AutoGen	Multi-agent conversations, handoff patterns, escalation	3
CrewAI	Task-based multi-agent orchestration for sequential workflows	3
Google ADK	Code-first agent framework — Gemini-optimized, model-agnostic	3
Sentence Transformers	Local embedding generation for semantic search	2
HF Transformers	Load, run, and fine-tune open-source models	4
PEFT	LoRA/QLoRA — parameter-efficient fine-tuning	4
bitsandbytes	4-bit/8-bit quantization for fine-tuning on consumer GPUs	4
scikit-learn	Query complexity classifier for SLM/LLM routing	4

LLM Providers

Provider / Model	What Students Learn	Free Tier
OpenAI GPT-5	General-purpose LLM — baseline, function calling, streaming	Trial credits
Anthropic Claude	Complex reasoning, long-context, structured outputs	Trial credits
Google Gemini	Multi-modal, long-context, Vertex AI integration	✓ free tier
Llama 3 (8B/70B)	Open LLM — fine-tuning base, local serving via Ollama	✓ free
Mistral (7B)	Efficient inference, fine-tuning base	✓ free
Phi-3 (3.8B)	SLM for cost-efficient routing, local development	✓ free
Qwen 2.5 (7B)	Multilingual support, fine-tuning experiments	✓ free

Local-first: Ollama runs open-source models on personal laptops, eliminating API costs during labs.

MLOps / LLMOps

Tool	What It Does	Free Tier
MLflow	Experiment tracking, model registry, model serving	✓ open source
Langfuse	LLM observability — traces, latency, cost, prompt management	✓ open source
RAGAS	RAG evaluation — faithfulness, relevancy, precision, recall	✓ open source
NeMo Guardrails	Programmable guardrails for LLM input/output safety	✓ open source
Presidio	PII detection and redaction (names, emails, SSNs)	✓ open source
BERTopic	Topic clustering on query data for theme identification	✓ open source
Weights & Biases	Fine-tuning experiment visualization (optional)	✓ free tier

Infrastructure, CI/CD & Monitoring

Tool	What It Does	Free Tier
Docker	Application and model serving containers	✓ open source
Kubernetes	Container orchestration for model serving and scaling	✓ open source
Helm	Kubernetes package management — templated deployments	✓ open source
Terraform	Infrastructure-as-code for multi-cloud provisioning	✓ open source
GitHub Actions	CI/CD for code, tests, and ML model retraining	✓ free (public)
KServe	Kubernetes-native model serving with autoscaling	✓ open source
vLLM	High-throughput LLM serving engine	✓ open source
Prometheus	Metrics collection — latency, error rate, GPU utilization	✓ open source
Grafana	Real-time dashboards — cost, routing, performance	✓ open source

Testing

Tool	What It Does	Free Tier
pytest + pytest-asyncio	Async unit and integration test runner	✓ open source
httpx	Async HTTP test client for FastAPI endpoints	✓ open source
pytest-cov	Code coverage with fail-under thresholds	✓ open source
RAGAS	RAG pipeline evaluation as a testing tool	✓ open source
Playwright	End-to-end browser testing for React frontend	✓ open source

THE PROJECT

Build SageDesk — An Enterprise AI Assistant

You build **SageDesk**, a multi-tenant Enterprise AI Assistant that answers employee questions from company knowledge bases, orchestrates actions through agents, routes queries between SLMs and LLMs for cost efficiency, and runs reliably across four cloud AI platforms.

> **Your final deliverable:** A deployed, monitored AI platform with RAG, agents, fine-tuning, guardrails, evaluation, ML pipelines, and multi-cloud deployment — a portfolio project that demonstrates real agentic AI engineering capability.

Subsystem	What It Does	Curriculum Area
Knowledge Ingestion	Upload, chunk, embed, and store documents	RAG
RAG Pipeline	Retrieve, rerank, generate grounded answers	RAG
Chat Interface	Streaming chat with conversation history	GenAI
Agent Orchestration	Search, action, and escalation agents	Agentic AI
Model Router	SLM/LLM routing for cost optimization	SLM/LLM
Fine-Tuning Pipeline	LoRA/QLoRA training, model versioning	Fine-Tuning
Guardrails Engine	PII detection, content filtering, hallucination check	LLMOps
Evaluation Framework	RAGAS metrics, A/B testing	LLMOps
ML Analytics	Usage patterns, topic clustering, MLflow	MLOps
Platform Monitoring	Latency, drift, auto-scaling, alerting	AIOps
Multi-Cloud Deploy	Vertex AI, Bedrock, Azure AI, OpenShift AI	Cloud

PROGRAM DETAILS

Format & Structure

This is a deep, immersive program. Each week produces a working component of SageDesk. Concept coverage (30 min) is followed by hands-on lab work (90 min) building real infrastructure.

FORMAT	Instructor-led with hands-on labs
DELIVERY	Virtual (live, instructor-led)
DURATION	16 weeks (one 2-hour session per week)
SESSION	Every Sunday, 7:00 PM – 9:00 PM IST
LAB SETUP	Personal laptop (Ollama for local models) or cloud-based lab instances
LANGUAGES	Python 3.14 (backend) · TypeScript 5.8 / React 19 (frontend)
TAKEAWAYS	All lab code, deployment configs & reference guide

Session Time by Timezone

Region	Timezone	Session Time
India	IST (UTC+5:30)	**Sunday 7:00 PM – 9:00 PM**
USA (East Coast)	EST (UTC-5)	Sunday 8:30 AM – 10:30 AM
UK / Europe	GMT (UTC+0)	Sunday 1:30 PM – 3:30 PM
UAE / Middle East	GST (UTC+4)	Sunday 5:30 PM – 7:30 PM
Singapore / East Asia	SGT (UTC+8)	Sunday 9:30 PM – 11:30 PM

IMPORTANT

Note on Cloud Credits & Free Tiers

Cloud credits and free tiers may change. All cloud platforms used in this programme have verified free tiers or education credits as of February 2026. However, cloud providers may modify, reduce, or remove free tiers at any time without notice. **Students are responsible for all costs incurred for AI tools, cloud platforms, GPU usage, and API calls.** Students must create their own free-tier accounts for labs. If a platform's free tier is no longer available, students must bear the cost of a basic subscription.

Ready to Build Production AI Systems?

Contact us for batch schedules, corporate training, and custom program design.

Rathinam Trainers & Consultants Private Limited

We train engineers, not just tool users.