

SURENDIRAN SELVAM

AI Platform | Multi-Agent Systems | LLM Ops | Agentic AI | RAG

+919900768678 @contact@surendiran.ai Portfolio: surendiran.ai

LinkedIn: linkedin.com/in/surendiran-selvam1 Bengaluru, India

SUMMARY

Senior AI Platform Engineer / Technical Lead with 10+ years building scalable engineering systems and the last 2+ years delivering enterprise GenAI platforms in regulated financial services. Lead 3–5 engineer cross-functional teams to ship RAG ingestion pipelines and Python-based agent workflows with multi-agent orchestration (Lang Graph), hybrid retrieval (Qdrant), and production observability (Lang Smith + logs/metrics). Design governance-ready AI systems with guardrails, PII redaction, bias mitigation, evaluation (RAGAS/LLM-as-judge), and audit-friendly controls—translating platform capabilities into roadmap features and measurable business outcomes.

EXPERIENCE

Senior Engineer

01/2024 - Present

EPAM Systems

Bengaluru,India

- Led Confluence-to-RAG ingestion for Fortune client, spearheaded a strategy that facilitated near real-time KB updates through LangChain, influencing knowledge management.
- Designed extraction and normalization standards that improved knowledge base quality and cut ingestion runtime by 35–45%.
- Managed Elite Agent workflows in Python, achieving 90% accuracy and supporting hundreds of users daily with policy-aware responses.
- Enhanced Qdrant hybrid retrieval by 25–35%, reducing false-positives by 20–30% and improving query precision.
- Implemented AI governance with PII redaction, bias mitigation, and reliability controls to ensure compliance and improve query quality.
- Lowered P95 latency by 20–25% and triage time by 50–60% using trace/log-driven troubleshooting with Lang Smith.
- Guided AI vendor selection, achieving a 40% cost reduction while ensuring quality, influencing decisions on vector databases and retrieval architecture.

Lead Automation Engineer (Automation Platform)

08/2021 - 01/2024

Encora Innovation Labs

Bengaluru

- Company Description
- Maintained a 99%+ success rate for daily runs by developing a platform for 8 teams, significantly enhancing operational efficiency.
 - Decreased manual tasks and re-runs by 50–70% through Azure DevOps pipeline optimization, facilitating continuous validation.
 - Decreased environment setup time from hours to under an hour using Kubernetes-based tooling via Helm for shared services deployments.
 - Ingested tens of GB of logs daily with centralized logging through ELK and Kibana dashboards, reducing diagnosis and debugging time by 40–50%.
 - Enforced Prometheus and Grafana for monitoring and alerting, cutting infrastructure bottlenecks and failures by 30–40%.
 - Scaled Selenium Grid to 50+ concurrent sessions, reducing end-to-end suite runtime by 50–65% and improving system stability under load.
 - Automated failure-to-triage workflows, decreasing triage time and ownership handoffs by approximately 40–45%.

Software Engineer (Automation & Platform)

08/2019 - 2021

ASG Technologies

Chennai, India

- Hired by ASG while at Wipro (Wipro: June 2018 – July 2019)
- Led 3 engineers, updated stakeholders; ensured on-time releases. Enabled Docker/Kubernetes-based environments, cut provisioning time by 60%. Built Jenkins pipelines for WebLogic/Tomcat, facilitating systematized, zero-downtime deployments.
 - Enabled container-based environments using Docker/Docker Compose and Kubernetes deployments via Helm charts, reducing environment provisioning time by 60%.
 - Developed CI/CD pipelines in Jenkins (Groovy + shell + Docker) for WebLogic/Tomcat deployments and continuous validation, enabling systematized deployments with zero-downtime rollbacks.

Software Engineer

06/2015 - 02/2018

Finatel Technologies

Chennai, India

- Hired by OASYS while at Finatel (OASYS: March 2018 – May 2018)
- Created test suites and identified automation candidates; maintained RTM and functional coverage mapping, achieving 70% automation coverage.
 - Developed automation scripts using Java, Selenium, TestNG, Maven; implemented stable locator strategies and page object design, reducing script maintenance effort by 40%.

KEY ACHIEVEMENTS

- Improved Query Precision
- Enhanced query precision by 35% through optimized hybrid retrieval using Qdrant.
- Streamlined Ingestion Process
- Decreased ingestion runtime by 45%, enhancing knowledge base quality through new standards.
- Optimized Agent Workflows
- Achieved 90% workflow accuracy, supporting hundreds of policy-aware daily user interactions.
- Effective Troubleshooting Implementation
- Reduced triage time by 60% through effective trace/log-driven troubleshooting with Lang Smith.

PORTFOLIO PROJECT - LIVE

- Multi-Agent Product Intelligence Platform
- Deployed Demo RAG to Agentic RAG to Multi-Agent System
- GitHub Repo: [GitHub](#) | Live Demo: [Live Demo](#) | Architecture: [Architecture](#)
- Enterprise GenAI | Agentic Workflows | Multi-Agent Orchestration | Stateful Agents | Tool Calling | Hybrid RAG | Deployed (GCP)
- Built & shipped a multi-agent RAG platform on GCP using Lang Graph + Qdrant hybrid retrieval (BM25+embeddings, RRF), tool calling, structured outputs, persistence, and CI/CD eval gates; **median latency** on live demo traffic (500 products / 45,948 reviews) — more on [GitHub](#).

SKILLS

- Core
- Python, FastAPI, Lang Graph, LangChain, Lang Smith, Qdrant, Hybrid Search, BM25, Semantic, RRF, Embeddings, Reranking, Cohere, RAG Ingestion, RAGAS, LLM-as-Judge, OpenAI, GPT-4, GPT-4.1, GPT-4 mini, Gemini, Groq, Lite LLM, Instructor, Pydantic, PostgreSQL, SQL, Elasticsearch, Docker, Docker Compose, Kubernetes, Helm, ECK, CI/CD, Azure Pipelines, Jenkins, Groovy, GitHub Actions, SonarQube, ELK Stack, Elastic, Kibana, Prometheus, Grafana

CERTIFICATIONS

Google Cloud Professional Machine Learning Engineer	Certified Kubernetes Administrator (CKA)
2025	2025
Microsoft AI-900: Azure AI Fundamentals	
2025	


PUBLICATIONS

Medium Blogs
https://medium.com/@surendiran-ai/the-modern-ai-product-lifecycle-for-enterprise-platforms-e631f13a7bbc
Personal Blogs
https://www.surendiran.ai/blog

EDUCATION

Karpagam University	2011 - 2014
Bachelor Of Engineering	

SKILLS

Platform Operations
Observability, Tracing, Dashboards, Monitoring, CI/CD Quality Gates, Release Gates, Automation, Pre-commit Hooks, Repo Standardization, Reliability Controls, Guardrails, Fallbacks, Regression Checks, Drift Checks, Reverse Proxy, Caddy, Nginx, Rate Limiting, TLS Termination, Secrets Management
Cloud & Security
GCP, Compute Engine, Secret Manager, IAM, VPC/Firewall Rules, Basic Auth, API Key Management, Audit Logging
Responsible AI & Governance
AI Guardrails, PII Redaction, Bias Mitigation, Safety Controls, Evaluation Frameworks, Hallucination Auditing, Human-in-the-Loop Orchestration
Dev Productivity:
Selenium WebDriver, Selenium Grid, Rest Assured, Postman, JMeter, TestNG, JUnit4, Maven, Spring Boot, SoapUI Pro, Swagger
AWARDS
 Super Encorian 2024