

SARAVANAMUTHU S

Senior Software Engineer • AI Platform & Backend Engineer

+91 9894957042 | saravanamuthuise@gmail.com | linkedin.com/in/saravanamuthu-s

PROFILE SUMMARY

Software Engineer with 4 years of experience building production-grade backend systems, autonomous multi-agent AI platforms, and multi-cloud deployment solutions used by thousands of users. Deep expertise in real-time distributed architectures, LLM agent orchestration, context engineering, and self-improving learning systems. Track record of owning critical systems end-to-end — from architecture design to performance optimization — while collaborating closely with product and engineering stakeholders to ship high-impact products.

WORK EXPERIENCE

Senior Software Engineer | Adya.ai

Aug 2023 – Present

Autonomous Multi-Agent Platform (SAI)

- Architected the **supervisor core of SAI**, an autonomous multi-agent delivery system that turns a plain-language request into delivered artifacts — classifying intent, generating multi-phase plans, delegating to specialist agents, and executing independent steps in parallel via a **DAG engine** with dependency resolution, retries, and cascade-failure handling.
- Re-architected the skill system into a **central skill registry** — skill metadata indexed in a **vector database**, raw definitions in **MongoDB** — equipping 1,400+ skills on demand at runtime and **auto-installing** new capabilities from an online skill marketplace when no local match exists.
- Engineered a **3-layer context-management system (Auto Compact)** across SAI — cross-request history compaction, per-call recovery, and per-tool-call micro-compaction — enabling unbounded-length agent-building sessions without hitting LLM token limits.
- Built **HAA (Human Assistant Agent)** backed by a per-workspace VM context layer (Fly.io runtime) that stores and retrieves workflow context, memory, session summaries, and agent configuration — letting the assistant auto-answer user questions on their behalf.
- Integrated **human-in-the-loop control** (plan approve / modify / reject, structured mid-run questions) with **multi-channel engagement** — WhatsApp, Email, In-App, and Web Push — so users can answer agent questions, check run status, and steer long autonomous runs from any channel.

AI Workflow Automation (AIA)

- Re-architected AIA's natural-language-to-workflow builder into a **master-LLM recursive tool loop** that dynamically orchestrates specialist sub-agents (requirement gathering, component prediction, flow connection, validation & repair), grounded in a **750+ component catalog** (including MCPs) via vector retrieval.
- Bootstrapped AIA's knowledge base by running real use-cases through **AIA's own synthesis engine** to generate **~2,000 domain-segmented reference workflows**, fed back as grounding data to strengthen the component ontology, retrieval accuracy, and cold-start learning.
- Streamed every node and edge mutation **live to the canvas** on both create and update paths, with structural validation and one-shot automated repair of malformed flows.

Self-Improving Learning Layer (ESLL)

- Architected **ESLL**, a self-improving, **event-sourced learning layer** for generative AI systems: it observes each decision and the context that influenced it, extracts durable knowledge from behavioral telemetry, injects that knowledge into future decisions **under a token budget**, and attributes outcomes back to the knowledge that shaped them — promoting what works and demoting what does not, **with no gradient training**.
- Deployed ESLL on the AIA workflow builder to improve NL-to-workflow synthesis — using **materialized projections** for credit assignment and reinforcement, and **demand-signal extraction** to surface capability gaps as a product-roadmap input.

Backend, Data & Platform Engineering

- Designed a **no-code ETL platform** spanning major databases, cloud storage, and SaaS sources, then evolved it into a **fully agentic data pipeline** where users trigger extract-transform-load workflows through natural-language conversation.

- Built real-time collaboration and streaming with **WebSockets** and **Server-Sent Events**, enabling live concurrent editing and low-latency updates across platform modules.
- Optimized the production backend through targeted indexing, **Redis caching**, and **MongoDB connection-pool tuning** — cutting RAM usage ~30% under peak load — and handled burst traffic with **Go routines and channels** at stable latency.
- Integrated multiple **MCP (Model Context Protocol)** servers and extensible components for modular, configurable agent workflows at scale.

Cloud, Infrastructure & Leadership

- Architected **Cloud Studio**, a unified PaaS/SaaS deployment platform for AI agents; managed multi-cloud infrastructure hands-on — **VMs and AKS on Azure**, Compute VMs on GCP, and Bedrock LLM access on AWS — with **Azure AI Foundry** for model and key management.
- Maintained Dockerized service pipelines, Jenkins deployments, and Nginx configuration for consistent environment management.
- Mentored engineers and drove **architectural decision-making**, raising team standards on backend design and Go best practices; represented Adya.ai at platform launch events and technical workshops, and served as **Lead Jury** at the SECE hackathon.

Software Engineer Intern | Eunimart

Aug 2022 – Aug 2023

Backend Engineering — ONDC & E-Commerce

- Engineered **ONDC-compliant** backend services for a centralized e-commerce platform in **Go and Node.js**, and built the **ONEST platform** delivering course and job provider-seeker marketplaces on a unified backend.
- Developed a **Reconciliation Service Provider (RSP)** system automating transaction reconciliation and financial settlement, integrating banking partners **HSBC, IDFC, and NPCI** into secure automated pipelines.
- Integrated marketplaces, webstores, and logistics providers to enable omnichannel seller operations, owning production backend services end-to-end.

TECHNICAL SKILLS

Languages	Go (Golang), Node.js, JavaScript, Python (working knowledge)
Backend	REST API Design, Microservices, WebSockets, Server-Sent Events (SSE), Kafka
AI / LLMs	Agentic AI, Multi-Agent Orchestration, LLM Integration, Model Context Protocol (MCP), Context Engineering, RAG / Vector Retrieval, Self-Improving Learning Systems, Browser-Use
Data	MongoDB, Redis, PostgreSQL, Vector DB (Milvus / Zilliz), ETL Pipelines, Connection-Pool Optimization
Cloud & DevOps	Azure (VM, AKS, AI Foundry), GCP (Compute VM), AWS (Bedrock), Docker, Jenkins, Nginx, PaaS/SaaS Deployment
Concepts	Distributed Systems, Real-Time Architectures, System Design, Performance Optimization, Concurrency, DAG Orchestration, Event Sourcing

EDUCATION

B.E. in Information Science & Engineering

Bannari Amman Institute of Technology, Sathyamangalam

Graduated 2023

INTERESTS

System Design & Optimization • Problem Solving • Badminton • Photography • Traveling