Aravind Anchala

Senior Software Engineer, AI Systems & ML Infrastructure

San Francisco Bay Area · aravindanchala5@gmail.com · aravindanchala.com · GitHub · LinkedIn

SUMMARY

Software engineer with 8+ years building the systems that turn models into reliable products: model-serving and inference workflows, agentic systems, evaluation loops, data pipelines, observability, authorization, and cost controls. My background is in software that has to work in production, across large-scale distributed systems, cloud-native infrastructure on AWS and Kubernetes, applied optimization, and reliability engineering. I care as much about latency, cost, safety, and evaluation as I do about model quality, and I hold 5 U.S. patents in automated network optimization.

EXPERIENCE

Software Engineer · MeeruAI

Feb 2025 - Present

San Francisco Bay Area

Building production software for an AI-powered enterprise SaaS platform that automates complex business workflows with LLM-enabled agents, secure data access, observability, and full-stack product experiences.

- Build and ship LLM-enabled agentic workflows that automate complex enterprise processes, with structured task execution, controlled backend access, prompt guardrails, and orchestration patterns that improve reliability and operational safety.
- Design natural-language-to-structured-data workflows that translate user intent into governed data access, analytics, and business-process execution across frontend UX, backend APIs, and authorization.
- Harden production maturity with policy-driven authorization, observability, and debugging workflows, improving security, incident investigation, and platform maintainability.
- Contribute across the stack, from backend services and API integrations to data workflows, access control, and production readiness.

Senior Network Engineer · Samsung Electronics America

Apr 2022 - Jan 2025

Reston, VA

Owned production troubleshooting, reliability analysis, and data-driven optimization for large-scale distributed systems, with a focus on incident resolution, performance debugging, and cross-functional execution.

- Resolved 100+ Tier-3 production performance and reliability incidents, owning root-cause and failure-mode analysis and driving measurable gains in system stability.
- Built Python analysis pipelines over large operational and performance datasets for root-cause analysis, anomaly and trend detection, and optimization recommendations.
- Turned complex system behavior into actionable engineering insight with SQL, Tableau, and operational analytics, and supported live operational dashboards.
- Ran configuration audits to surface recurring failure patterns and reduce operational risk. This data-intensive troubleshooting maps directly to AI/ML platform reliability, model-serving observability, and AI SRE.

Network Engineer · DISH Network Technologies

Dec 2018 - Apr 2022

Herndon, VA

Contributed to cloud-native platform architecture, automation, and performance analysis across a large-scale, virtualized and containerized distributed system.

- Helped design and evaluate a cloud-native, service-oriented platform on AWS with virtualized and containerized components, built for scalability and performance.
- Built Python, Bash, and SQL automation and data pipelines for operational reporting, model tuning, capacity analysis, and acceptance workflows, cutting repetitive manual work across design, deployment, and optimization.
- Developed automated optimization and predictive model-tuning algorithms. This cross-functional work led to 5 U.S. patents in automated network optimization and interference management.
- Worked across AWS, Linux, IP networking, and REST-API automation, building the platform, data-pipeline, and reliability foundations that carry directly into AI/ML infrastructure.

Engineering Intern, Python Programming · Caterpillar Inc. May 2015 - Aug 2015

Chennai, India

Built Python-based numerical optimization workflows for experimental data fitting and algorithmic modeling.

- Developed a hybrid Nelder-Mead + Genetic Algorithm optimizer to fit experimental damping curves, reducing error by ~10 dB.
- Built an N-dimensional curve-fitting workflow for experimental data analysis and model optimization.

SELECTED PROJECTS

LLM Inference Gateway · Inference & Reliability In progress

A multi-provider LLM gateway with a reliability control plane in front of it.

Post-Training Orchestrator · Training & Post-Training Planned

Reproducible, gated distributed fine-tuning and post-training runs.

Agentic RAG Copilot · Agents & Retrieval Planned

A permission-aware agentic copilot that assists incident response.

Eval & Data Flywheel · Evaluation & Data Planned

Close the loop: evaluations and production signals feed a governed feature store.

Trust & Safety Platform · Trust & Safety Planned

Auditable, policy-driven moderation across text and images.

Real-Time Risk Intelligence · Streaming & Real-Time Planned

Streaming feature computation and scoring from edge to cloud.

SKILLS

AI / ML systems	LLMOps · MLOps · RAG · Agentic AI · Model serving · Model evaluation · Real-time ML
Platform & infra	Kubernetes · AWS · Terraform · Linux · Cloud-native systems · Cloudflare
Backend & data	Python · Go · Rust · SQL · REST APIs · Distributed systems · Data pipelines
Frontend	React · Next.js · TypeScript · Tailwind CSS
Reliability	Observability · AI SRE · Incident response · Performance analysis · Cost engineering

EDUCATION

Brown University

Master's degree

Indian Institute of Technology, Madras

Bachelor's degree

PATENTS, CERTIFICATIONS & RECOGNITION

- AWS Certified Solutions Architect - Associate
- Google IT Automation with Python
- VMware Cloud on AWS - Trained Professional
- Named inventor on 5 U.S. patents (automated optimization, interference management, predictive model tuning)
- Peer-reviewed publication (quantum-dot photodetectors)