KRISH ARYAN

+91 90365 77433 • krish.aryan.2003@gmail.com • linkedin.com/in/krish-aryan • Bengaluru, India

SUMMARY

AI Engineer building production-grade agentic systems that operate in real, high-stakes environments — not just demos. Designed and shipped TraceLens, a 40+ tool autonomous incident investigation agent (12+ data sources, native ReAct loop, no LLM frameworks) compressing triage from ~2 hours to 15 seconds (480×), live in production on AWS EKS. Equally fluent in the infrastructure layer that agents actually run on: AWS, Kubernetes, observability at scale, on-call engineering. Seeking founding-stage or AI-first roles where the mandate is agents that work in production.

SKILLS

AI & Agentic Systems: ReAct loop design, Native LLM tool-calling (no frameworks), Multi-source agent orchestration, Tool schema design, Prompt engineering, SSE streaming, Council of Models (production grade implementation)

LLM & AI Backends: Azure AI Foundry, Anthropic Claude, Google Gemini, Model Context Protocol (Anthropic-certified), OpenAI-SDK, AWS Bedrock

Backend Engineering: Python, FastAPI, Shell Scripting, Node.js/Express, Java, gRPC, REST APIs, WebSockets, SSE, Async event-driven architectures

Frontend: React 18, Vite, Tailwind CSS, real-time SSE-driven streaming UIs, data visualization dashboards

Cloud & Infrastructure: AWS (Bedrock, EKS, EC2, S3, Lambda, SQS, Kinesis, DynamoDB, MWAA, CloudWatch), Kubernetes, Docker, Nginx, Jenkins, Git

Observability & Data: Elasticsearch, Prometheus, Grafana, OpenTelemetry, Tempo, Sentry, PagerDuty, ClickHouse, Redis, Valkey, PostgreSQL, DynamoDB

EXPERIENCE

NETRADYNE | Bengaluru, India **Feb 2025 – Present**

Software Engineer – Site Reliability Engineering & AI Tooling

- Designed and deployed **TraceLens**, a fully autonomous incident investigation agent with 37 specialized tools across 12+ data sources (Elasticsearch, Prometheus, AWS, GitHub, Jira, Sentry, MWAA, PagerDuty) — native ReAct loop, no LLM frameworks; compressed triage time **from ~2 hours to ~15 seconds (480× improvement)**, live in production on AWS EKS.
- Architected the full agent system end-to-end: multi-source handler/controller where the agent self-generates queries and selects optimal data source per metric; FastAPI + SSE backend; React 18 + Vite + Tailwind dark terminal frontend; PostgreSQL-backed history; JWT auth; auto-generated Jira incident tickets.
- Engineered a **ReAct Log Analysis Agent** running natively on the S3 mount filesystem — treats bash tools (grep, awk, zcat, zgrep) as its toolset; handles 200k–30M log lines per service per day with zero ingestion overhead, no vector database; @service routing, timeline flags, rich terminal UI.
- Built an **Exception Clustering system** from scratch — an org-centric Sentry alternative that groups, deduplicates, and ranks exception patterns across 215+ production services tuned to internal service topology; fully replaced a paid SaaS observability tool.
- Designed and deployed **Anomaly Detection systems** for API latency and traffic rate deviations; built SLO/SLI automation with early-warning alerting pipelines; built AlertFlow for real-time alert dependency graph visualization across 185+ services.
- Served as on-call SRE across 185+ production services; prevented **3 Sev-1/P0 incidents** through proactive signal correlation; led live incident investigation across logs, metrics, and distributed traces.
- Sole engineer on **PostgreSQL → DynamoDB migration** (schema redesign, pipeline, zero-downtime cutover); served as **core architect of multi-cluster EKS enablement** across the organization.

PROJECTS

TraceLens — Autonomous AI Incident Investigation Agent

Production-deployed agentic RCA system — agents.sre.netradyne.info/incident-analysis-agent

- Built a native ReAct tool-calling loop (no LangChain or agent frameworks) with 37 tools; agent self-selects optimal data source per query across 12+ integrated sources, delivers structured confidence-scored hypotheses with evidence chains, and auto-generates Jira tickets.
- Stack: Python 3.11, FastAPI, SSE streaming, React 18, Vite, Tailwind CSS, Azure OpenAI GPT-4o, PostgreSQL, S3, AWS EKS with Nginx reverse proxy; full production deployment with JWT authentication.

Log Analysis Agent — S3 ReAct CLI + Browser UI
Filesystem-native agentic log intelligence tool — agents.sre.netradyne.info/log-analysis-agent

- Designed a ReAct agent that treats bash tools (grep, awk, zcat, zgrep) as its hands — runs directly on the production S3 mount with zero ingestion infrastructure, no vector store, no preprocessing overhead.
- Handles real log-native scale (200k–30M lines/service/day); supports @service-name pinning, --date/--range timeline flags, /export and /clear commands, and a rich terminal UI; backed by Azure AI Foundry via OpenAI-compatible SDK.

CERTIFICATIONS

Anthropic: Introduction to Model Context Protocol • Model Context Protocol – Advanced Topics • Claude 101
Amazon Web Services: Cloud Practitioner Essentials • Getting Started with DevOps on AWS
ClickHouse: ClickHouse Database Associate • chDB Professional

RESEARCH PUBLICATIONS

- Real-Time Biometric Facial Recognition Attendance System — ACT 2024; Scopus Indexed, GRENZE International Journal of Engineering and Technology.
- Privacy-Preserving Border Surveillance System — I-SMAC 2024; Published in IEEE Xplore.

EDUCATION

New Horizon College of Engineering, Bengaluru **2021 - 2025**
Bachelor of Engineering, Artificial Intelligence & Machine Learning | GPA: 9.24 / 10