Rohit Jain

Founding Engineer · Applied AI & Agent Infrastructure

Bengaluru, India | +91 97522 71891 | contact2rohitjain@gmail.com | linkedin.com/in/rohitjainlink | github.com/xRME-18

SUMMARY

Founding engineer who took an LLM agent from feasibility prototype to a production platform resolving e-commerce support tickets across many brands. Works end to end across architecture, safety, evals, and the runtime; also ships independent products, like Canvas, a live macOS AI app. Named inventor on 3 pending patents in LLM-agent planning and evaluation.

EXPERIENCE

Kim.cc **Jun 2024 – Present**

Founding Engineer *Bengaluru, India*

Second engineer; joined post-seed at ~7 people. Original author and primary contributor to Kim's AI customer-support agent runtime.

- **Owned the agent runtime end to end,** from agent architecture through production infrastructure, powering **275K+ customer turns across ~50 Shopify brands YTD 2026** and ~160K tool calls/month across 90+ commerce actions.
- **Built the planner–executor core:** a policy engine compiles merchant SOPs into typed plans, executed by a bounded ReAct agent against live commerce APIs—separating business policy from LLM reasoning.
- **Built a resumable multi-loop agent runtime:** composed planning, execution, human review, and customer interaction into one stateful workflow where components can yield and later resume from prior state—supporting supervisor approvals, iterative refinement, replanning, and customer clarification. Routed **7K+ high-risk writes** including cancellations, refunds, and subscription changes through human gates.
- **Built promoted plans** that bypass full planning for validated workflows, **cutting LLM cost 50–60% and median latency** ~44% on requests they serve.
- **Built merchant workflow-generation infrastructure:** helped build the original SOP-to-workflow pipeline and **designed SOP Execution Graphs (SEGs) and their compiler**, turning approved support procedures into executable workflows with deterministic control flow, commerce tools, and bounded LLM judgment.
- **Built an AI root-cause investigator** that automates the trace triage engineers did by hand: it traces flagged production failures across knowledge, retrieval, tools, and reasoning, and clusters recurring defects for prioritization.
- **Mentored 2 new-grad engineers** and served as the primary reviewer for runtime changes.

Spiralyze **Jan 2024 – Jun 2024**

Software Engineer Intern *Atlanta, US (Remote)*

- Built a serverless crawler (AWS Lambda, SAM) that detects site changes likely to break clients' A/B tests, including pages behind token, form, and OAuth logins.
- Migrated it from Elastic Beanstalk to Lambda, **cutting operating cost by 30%** and making it easy for other teams to deploy.

Atlassian **Jun 2023 – Jul 2023**

Software Development Engineer Intern *Bengaluru, India*

- Built Kotlin/Spring Boot monitors tracking execution time and state at each stage of Opsgenie's data-residency migration service.
- Built live New Relic dashboards (P95/P99 processing time, throughput, callback delays) and used CloudWatch and Splunk to find and fix a log-leakage issue across the service's SNS/SQS/DynamoDB microservices. **Received a pre-placement offer.**

PATENTS & AWARDS

- **3 pending patent applications** (named inventor): execution-plan memory for LLM agents, root-cause evaluation for multi-step agents, and tiered plan caching.
- **1st place**, AI Agents Hackathon (Airtribe × Lyzr AI), 2025.

SELECTED PROJECTS

- **Canvas** — Live macOS app that turns natural language into working desktop widgets; built and launched solo.
- **PageCue** — Open-source WebMCP library letting a visitor's browser agent read page state and suggest guidance.

EDUCATION

BITS Pilani **2024**

B.E. (Hons.) Computer Science + M.Sc. (Hons.) Mathematics · CGPA 8.5/10 *India*

- Teaching Assistant for Discrete Mathematics and Mathematics 2

SKILLS

AI systems: LLM agents, planner–executor, ReAct, RAG, tool use, structured outputs, evals, tracing

Languages: Python, TypeScript, SQL, Kotlin

Infrastructure: asyncio, FastAPI, RabbitMQ, gRPC, Redis, MongoDB, PostgreSQL, Pinecone, AWS, Docker, Langfuse, OpenTelemetry, Grafana