Nikhil Navali

Autonomous AI Agents & Evaluation Infrastructure

Seattle, WA · nikhilnavali@gmail.com · nikhilnavali.com · linkedin.com/in/nikhilnavali

SUMMARY

Eleven years at Amazon, most recently building long-horizon autonomous AI agents: systems that complete work previously measured in human days. Depth in agent orchestration, LLM evaluation, and connecting agents to the systems, rules and institutional context that make their output trustworthy enough to act on. Shipped a multi-tenant agent platform to 100+ users and removed human audit from the release path of production ML decisioning.

CORE COMPETENCIES

Agentic AI: Agent runtime design · Long-horizon autonomy · Planner/executor orchestration · Tool use & function calling · Context engineering & compaction · RAG over domain knowledge bases

Evaluation & Trust: LLM-as-judge design · Eval frameworks & golden sets · Independent verification · Citation grounding

Data & Distributed Systems: Spark · EMR Serverless · Entity resolution at billion-row scale · Inference cost & throughput optimization

Cloud & Languages: AWS (Bedrock, Lambda, Step Functions, DynamoDB) · Python · Scala · TypeScript

EXPERIENCE

Amazon · Senior Software Engineer

Apr 2022 – Present

- Built and own a **multi-tenant agentic AI platform**: agent runtime, 25+ tools, RAG over domain knowledge bases, MCP tool stack, run observability. Serves **100+ users across 5 engineering disciplines** and **50+ autonomous investigations daily**.
- Built the **autonomous execution engine** that turns an intent into senior-engineer-grade work, whether the deliverable is an investigation, an analysis, a design document or working code. It decomposes the goal, chooses its own tools, reaches every system a senior engineer can reach, and cites a source for every claim. A **separate verifier re-checks the work against those sources**, with per-deliverable gates flagging anything unsupported before it ships. Each run deposits durable lessons the next one inherits, so competence **compounds across runs**. **~2,000 runs**; a day of engineer work now returns in under an hour (accepted for **oral presentation at IAAI-27**).

- Replaced procedural scaffolding with **behavioral specification**, which outperformed the staged plan/execute/verify pipeline. The same design let one generic agent cover new domains, so a second team stood up its own domain agent **in under a day** by contributing only a knowledge base.
- **Removed human audit from the release path** for production ML decisioning: agents hold the ambiguous precision/recall tail across tens of thousands of decisions a day at **95%+ accuracy under independent audit**.
- Authored the **system-specification standard** the team now writes against and retrofitted 10 production systems to it, enforced by a review-time rule that fires on every change regardless of author, human or agent. Those specs are what agents load for system context, which makes spec accuracy an operational property of every downstream agent run.
- Named **sole technical owner** for the organization-wide agent platform; 4 recorded organization-wide talks.
- Earlier in the role (2022 to 2024), architected distributed services for **automated entity validation over multimodal signals**, product imagery, web content and registry data, with enrichment pipelines fusing heterogeneous sources: **60% less human audit effort at 98% accuracy**.

Amazon · Software Engineer I & II

Jul 2015 – Mar 2022

- Co-led distributed entity-matching systems processing **1.5B+ item pairs per 12-hour window at 93%+ precision** (Scala, Spark, EMR).
- Built a self-service model-probing tool adopted by **hundreds of auditors**: debugging cycles days to hours, model precision **+5 percentage points**.

EDUCATION, PUBLICATIONS & PATENTS

M.Eng. Computer Science, Cornell University · **B.E. Computer Science**, VTU Bengaluru

First author, “A Knowledge-Grounded Autonomous Agent for Deep-Dive Analytics over Enterprise Data Lakes.” IAAI-27 (co-located with AAAI-27), oral presentation, 2027 · **IAAI Deployed Application Award**

3 granted patents in large-scale ML systems: US11514321B1 (transfer learning for intra-cluster analysis) · US11675766B1 (scalable hierarchical clustering) · US20250173627A1 (distributed model training)