Aditya Shaji

Software developer — agent systems, evaluation, and developer tooling

Mumbai, India · adityalpha16@gmail.com · github.com/ps-aditya · linkedin.com/in/adityashaji · x.com/psdevelop

SUMMARY

I build agent systems, and the harnesses that check them. Author of **AnyBridge**, an MCP server that compiles a model plan once into a human-approved, engine-verified workflow and then replays it with no model calls, and **AgentTrace**, agent guardrails that recover only inside the buyer's original mandate or abort. Before that I shipped developer tooling to real users: Redis State Explorer at 2,000+ installs, Docker Vitals at 500+. Invited Reviewer, ISMIR 2026.

SELECTED WORK

AnyBridge

8/8 measured tasks · 8/8 red-team blocked

MCP server for software with no integration — plan, approve, verify, replay.

- **Planner.** The model chooses IDs from a numbered capability menu and submits a typed JSON workflow; it never writes shell strings, selectors, URLs or code.
- **Control.** Approval is out-of-band, single-use, hash-bound and expires in 10 minutes; policy is re-checked on the resolved step at execution time. Destructive actions are denied by default.
- **Verification.** Success is engine-computed evidence — exit codes, DOM text, file hashes — never the model's claim. Verification returns pass, fail or inconclusive.
- **Replay.** Verified workflows are stored hash-pinned and versioned; replay re-runs the saved checks and fails on drift, at zero token cost.
- **Evidence.** Red-team gate over 8 covered attack classes blocked 8/8 (prompt injection in observed content, path-jail escape, unapproved execution); 46 unit tests pass.

TypeScript · MCP · Zod · Playwright · Node — built during the Meta Global AI Developer Hackathon

AgentTrace

INR 25,834 value preserved · 8 runs

Agent guardrails for checkout drift at the moment of payment.

- Captures a Recovery Mandate at decision time — functional, fulfilment and economic consent — and re-verifies it after price or stock drift.
- Unknown provider data is an abort condition, not a guess: a similarity score can rank candidates, it cannot authorise them. Approval is withheld if a substitution would breach the buyer's actual ceiling.
- Batch harness of 8 designed and 100 randomised transactions, each writing a full per-transaction JSON trace so every aggregate number has a receipt behind it. Recovered 6 of 7 orders and correctly aborted the one with no viable recovery.

TypeScript · Gemini · Claude · Razorpay API — real test-mode orders, payments and refunds

Low End LLM

In progress

An adaptive runtime for quantized LLMs on 4GB-RAM laptops.

- Real-time RAM, CPU and battery profiling with a per-turn safety floor that halts inference before memory exhaustion instead of letting the OS swap.
- GGUF / Q4_K_M loading via llama.cpp with mmap weights, behind a pre-flight capacity check that refuses a model the machine cannot hold.

Python · llama.cpp · GGUF · Q4_K_M

Claude Code Java

From scratch

A coding agent written by hand — the loop, not a wrapper.

- Send prompt plus tool specs, execute the requested tool call, feed the result back, repeat until the model answers with no tool calls. Read, Write and Bash tools with tool-call detection and JSON argument parsing written directly against the OpenAI-compatible API.

Java · OpenRouter · openai-java · Jackson

Redis State Explorer

2,000+ installs

See exactly which Redis keys changed, and when, inside VS Code.

- Save-diff timeline: snapshot Redis state before and after a file save, then show exactly which keys the save caused to change. A 200+ command terminal, live key explorer and a code detective that maps keys back to the lines that reference them.
- Connects over the wire with ioredis — no backend, no extra service, sub-second updates. MIT licensed and actively maintained; Docker Vitals (live Docker monitoring in VS Code) has 500+ installs.

TypeScript · Redis · ioredis · VS Code Extension API

Pulsar classification

ROC-AUC 0.97

Calibrated Random Forest on HTRU2 (~17,900 candidates, ~9% positive class).

- Probability calibration (Platt / isotonic) validated with reliability diagrams and Brier score; the decision threshold comes from an explicit cost model rather than a default 0.5 cutoff. Recall ~0.90 at precision ~0.90 on held-out test data.

Python · scikit-learn · pandas

EXPERIENCE

Founder — Low End LLM

Jul 2026 – Present · Remote

- Building and shipping an adaptive runtime that runs quantized LLMs on 4GB-RAM laptops, for developers without high-end GPUs.
- Own the runtime end to end: profiling, memory budgeting, quantized model loading and the capacity checks that refuse a model the machine cannot hold.

Creator & Maintainer — Redis State Explorer

Jun 2026 – Present · Remote

- Designed, built and maintain a VS Code extension used by 2,000+ developers, from the Redis protocol layer to the diff timeline.
- Invited Reviewer, ISMIR 2026 (International Society for Music Information Retrieval).

EVALUATION PRACTICE

- **Harnesses with receipts.** Scenario batches, per-run JSON traces, and aggregates that can be audited back to an individual transaction.
- **Explicit success criteria.** Pass / fail / inconclusive for agent outcomes; an order is not a payment and an abort without a valid recovery is a pass, not a miss.
- **Red-team gates.** Attack-class suites where the gate is "was it blocked", not "did it behave".
- **Calibration.** Brier score, reliability diagrams and cost-model thresholds before any number is quoted.

SKILLS

AGENTS	MCP servers over stdio, tool/function calling, planner and execution loops, structured output (Zod), human-in-the-loop approval, deterministic replay, prompt-injection defence, audit traces
MODELS	OpenRouter, Gemini, Claude, llama.cpp / GGUF, Q4_K_M quantization, mmap weight loading, memory budgeting, local inference
EVALUATION	harness design, scenario batches, per-run traces, red-team gates, probability calibration, Brier score, cost-model thresholds
LANGUAGES	Python, Java, TypeScript, JavaScript, Go, SQL
SYSTEMS	Redis, Docker Engine API, Playwright, Node.js, Git / GitHub Actions, VS Code Extension API, Chrome Extension APIs, Oracle SQL, Azure / Azure ML

EDUCATION

Atlas SkillTech University

Aug 2025 – Aug 2029 · Mumbai

B.Tech Computer Science — Artificial Intelligence & Machine Learning

All work above is public: github.com/ps-aditya

Portfolio: ps-aditya.github.io/portfolio