

ANUJ OJHA

AI-Native Product Engineer | Agent tooling, evals, LLM observability, MCP

Bhopal, India | ojhaanuj224@gmail.com | +91 98930 37152

anujojha.com | github.com/Anuj7411 | linkedin.com/in/anujojha18

SUMMARY

Ships AI agent tooling and products solo and measures whether they work: a Claude Code plugin with a locked, reproducible benchmark, a signed guardrail for AI shopping agents, and **25 merged PRs** into LLM observability platforms. Seeking a remote internship or early-stage engineering/product role.

EDUCATION

B.Tech, Computer Science and Business Systems, OIST Bhopal | Jun 2027 (Expected) | **CGPA 8.19/10**

IEEE CODE-2K26 (Apr 2026): lead author and presenter, AI-driven circular subscription fashion (now CIRQLE)

PROJECTS

Sipcode Claude Code plugin, MCP server and CLI | github.com/Anuj7411/sipcode | npmjs.com/package/sipcode

- PreToolUse hook, 4 skills and 16 MCP tools that cap verbose tool output and dedupe same-session re-reads; zero network calls, enforced by a build-failing test; **2,200+ tests**; listed on the official MCP Registry.
- **62.6% median tool-output savings** on a locked 20-task benchmark anyone can rerun; **10,000+ npm downloads**, **#12 Product of the Day** on Product Hunt (143 points), 46 GitHub stars.

Pakka guardrail for AI shopping agents | github.com/Anuj7411/pakka | pakka-l78z.onrender.com

- Checks an agent's cart against the human's instruction before the Razorpay order exists; deterministic checks can **block**, the LLM judge can only escalate; decisions are Ed25519-signed and hash-chained.
- A real shopping agent diverged from its instruction **7.1% of the time (n=70)**; the deterministic layer had **0 false positives on 813 conforming carts**; prompt-injection demo; **556 tests** plus mutation testing.

Answerfox AI-readiness audit toolkit | github.com/Anuj7411/answerfox | npmjs.com/org/answerfox

- Audits what AI crawlers receive from a site: **39 live checks**, incl. agent readiness (MCP Server Card, A2A agent card, WebMCP); each issue comes back as a fix to commit.
- TypeScript monorepo of **7 published npm packages (6,600+ combined downloads)**: CLI, Cheerio audit engine, JSON-LD and sitemap helpers; network-free test seams and CI-gated releases.

Also shipped: Judix (Rust real-time eval engine for agents and RAG, live demo), Sotto (anonymous Secret Santa web app), Epley Coach (Kotlin Android app, 212 tests).

OPEN SOURCE CONTRIBUTIONS

- **25 merged PRs** into Comet Opik (21, Java backend, 22k+ stars), BerriAI LiteLLM (2, 61k+ stars) and Arize Phoenix (2, 11k+ stars), Jun - Sep 2026; 1,100+ GitHub contributions in the past year.
- Found LLM cost-tracking bugs that **mis-billed flagship models 2x-16x** (cached, audio, reasoning tokens, tiered pricing); fixes in Java and Python, plus provider fixes that load **500+ dropped model prices**.
- Every fix ships a regression test that fails on the old code; merged into Opik v2.0.58, v2.0.63 and v2.0.64; thanked publicly and invited to keep contributing by Opik's lead maintainer.

SKILLS

AI and Agents: Claude Code (plugins, skills, hooks), Codex, multi-agent orchestration, Model Context Protocol (MCP), LLM evals and observability (Opik, LiteLLM, Phoenix), RAG, prompt-injection defense, OpenTelemetry

Languages: TypeScript, Python, Rust, Java, Kotlin, SQL

Frameworks and Infrastructure: Next.js, React, Node.js, Supabase / PostgreSQL (RLS), Drizzle, Inngest, Cloudflare, Docker, GitHub Actions, Vitest, Stryker, pytest, JUnit, Cheerio, OpenAI / Gemini / Anthropic APIs

WORK EXPERIENCE

Data Scientist Intern, Hackveda Solutions (Remote) | Jul 2025 - Oct 2025

- Built ML pipelines in Python and Scikit-learn; picked models on precision, recall and F1 for stakeholders.

CERTIFICATIONS

Anthropic: Claude Code in Action, Intro to MCP | Oracle: OCI GenAI Professional | McKinsey Forward