

NIKUNJ AGRAWAL

nikunj-agrawal.com | nikunjag24@gmail.com | linkedin.com/in/nikunj3 | github.com/nikunj474 | No Sponsorship Needed

SUMMARY

Software engineer building agentic systems end to end: multi-agent orchestration in LangGraph, sandboxed execution and eval harnesses, and full-stack delivery in Python (asyncio, FastAPI) and TypeScript (Node, React). Shipped LLM systems serving 10K+ daily queries at sub-2s p99 behind eval loops, guardrails, and human-in-the-loop checkpoints; available to start immediately.

EDUCATION

University of Pennsylvania

Philadelphia, PA, USA

Masters in Computer Science

Aug 2024 - May 2026

- Teaching Assistant, Algorithms; NSF-funded Research Assistant. Coursework: Distributed Systems, ML, Databases

Shri Ram College of Commerce (SRCC), University of Delhi

B.Com. (Honors), Finance & Economics (Quantitative Methods); GPA 8.69/10 (Top 5%)

July 2018 - May 2021

EXPERIENCE

University of Pennsylvania, Provost's Office

May 2025 - Aug 2025

ML Engineering Intern

- Rewrote a batch inference path with asyncio concurrency for a 6x throughput improvement at identical held-out accuracy, and wired it to a React frontend without changing a single line of the frontend data contract
- Shipped a versioned TypeScript REST API with Zod validation, rate limiting, and full OpenAPI contract documentation, serving ML enrollment insights to 3 downstream teams across 12 academic schools
- Set up GitHub Actions CI/CD enforcing type-safety, lint, unit test, and integration test coverage gates on every pull request; built an MLflow pipeline with a model registry and automated retraining triggers

Darwinbox (HR SaaS Unicorn)

Dec 2021 - Jan 2023

AI Product Engineer (Founder's Office)

- Architected a production RAG support agent (LangChain orchestration, FAISS retrieval, GraphQL data layer) over enterprise knowledge bases, cutting average resolution time by 40% and lifting user adoption by 25%
- Built the LLM evaluation harness the team shipped against: RAGAS and BERTScore scoring, A/B harnesses, and nightly regression suites gating releases, lifting intent accuracy 32% over base across 10K+ daily queries at sub-2s p99
- Owned observability for production LLM deployments: drift detection, latency alerting, and a WebSocket event stream feeding TypeScript dashboards tracking usage and SLA metrics across 50+ enterprise tenants

GD Partners (Tech and Govt. Consulting)

Jan 2023 - Feb 2024

Technical Consultant (Software Engineering)

- Built FastAPI services and a React and TypeScript frontend with role-based access control for a civic platform serving 150K+ users, fed by asyncio-parallelized pipelines from 50+ sources with schema validation, retries, and dead-letter queues

PROJECTS

Autonomous Debugging Agent, Self-Healing Code Pipeline | *LangGraph, Code RAG, Docker, LangSmith*

- Built a multi-agent system where Investigator, Coder, and Tester agents share one typed Pydantic state object, over a Code RAG index of AST-aware chunks in Chroma that the Investigator tool-calls to localize faults
- Constrained patch generation to a strict JSON schema and ran every candidate under pytest in an isolated Docker sandbox, routing failing tracebacks back to the Coder through a conditional edge under a retry ceiling, then pausing at a human-in-the-loop interrupt for approve or reject; LangSmith traces attributed token spend, retries, and latency per agent

Agentic Financial Intelligence MCP Server | *FastMCP, LangGraph, FastAPI* github.com/nikunj474/financial-mcp-server

- Open-sourced an MCP server unifying 3 data sources (SEC EDGAR, FRED, Alpha Vantage) into 12+ typed tools, with LangGraph orchestrating research chains up to 25 tool calls deep across cross-filing comparison and macro correlation
- Scored 96% on a 100-question eval set by enforcing structured JSON schemas and stateful tool chaining to block hallucinated figures, with rate-limited cached retrieval holding p95 latency at 1.1s across heterogeneous APIs

LEADERSHIP & AWARDS

Winner, Princeton AI Hackathon: voice chatbot handling insurance, appointment, and referral inquiries by phone

1st Prize, National Hack-AI-thon, Indian Institute of Technology (IIT) Roorkee

Co-President, Pan Asian Graduate Students Assembly (PAGSA) at UPenn, running an annual budget of \$25K+

President, National Service Scheme, the largest student-run club on campus during undergraduate studies

TECHNICAL SKILLS

Languages: Python, TypeScript, JavaScript, SQL, C++, Java

Agents & LLM: LangGraph, LangChain, MCP, RAG, tool calling, structured outputs (Pydantic), RAGAS, LangSmith, Chroma

Backend & Systems: FastAPI, asyncio, Node.js, Express, REST, GraphQL, WebSockets, PostgreSQL, distributed systems

Frontend & Infra: React, Docker, AWS, GitHub Actions, CI/CD, pytest, MLflow, Weights & Biases, Git