

# Ayush Rathod

Pune, India | +91 9359682937 | ayushjrathod7@gmail.com | linkedin.com/in/ayushjrathod | ayushjrathod.live

## SUMMARY

Full Stack Engineer with a backend-heavy foundation, specializing in production-grade AI systems. I ship complete products across React frontends, FastAPI and Node.js backends, and real-time data infrastructure, with hands-on depth in LLM integration, RAG pipelines, context engineering, and evaluation. Experienced with PostgreSQL, Kafka, Redis, Docker, and AWS.

## WORK EXPERIENCE

### DevDynamics • Software Engineering Intern

Jul 2025 - Present

Products: DevDynamics (Engineering Metrics Platform) · Refacto.ai (AI Code Review Tool)

Pune, India (On-site)

- Built a **code review evaluation pipeline** for Refacto.ai, benchmarking AI review quality against a **100-PR human-curated dataset** using automated **LLM-as-judge scoring**. Enabled version-over-version quality regression tracking and direct comparison against competitor tools.
- Solved the core problem of LLMs lacking codebase-aware context in code review by developing two complementary systems: an **AST-based codebase context engine** (Tree-sitter, 19 languages) extracting diff-scoped, codebase-aware context mapping symbol level changes to their dependency graph to give the LLM impact-aware code understanding, and a **production LLM-as-judge filtering layer** scoring each comment for relevance before surfacing to developers, jointly **raising precision from 12% → 54% and recall from 39% → 71%**.
- Optimized the codebase context engine via a **two-tier caching strategy** (in-memory LRU cache + DB-backed persistence) which **reduced peak RAM from average 4.5 GB to under 700 MB and processing time from average 2 min to 30 sec** across cold AWS Lambda starts.
- Deployed an **MCP server** and contributed to **Dyna AI**, exposing DevDynamics' engineering metrics as **LLM-callable tools** and a **natural language interface**, enabling AI agents and engineering leaders to query live DORA metrics, Tickets and PR analytics, and team health data.
- Integrated **Claude Code, Cursor, Windsurf, and GitHub Copilot** APIs into DevDynamics' AI Impact & Adoption module for per-user acceptance rates, session activity, and PR contribution metrics; and **Jira** into Refacto.ai's review pipeline to validate code changes against ticket requirements and acceptance criteria.

## PROJECTS

### Advista – AI Competitive Intelligence Platform

Demo GitHub

FastAPI, LangGraph, Groq, Celery, Redis, React, PostgreSQL, Prisma, AWS Lambda, Firebase

- Built an **AI competitive-intelligence platform** that turns multi-turn user chat into structured research briefs using a **LangGraph workflow** with schema-enforced LLM extraction, then orchestrates multi-source research and synthesis into **7-section market reports** with persisted session history and report artifacts in **Postgres via Prisma**.
- Engineered the platform's async research infrastructure by decoupling long-running **LLM jobs from the HTTP request cycle** using **Celery + Redis** for task dispatch and polling, with **Firebase Auth** for identity and session ownership, deployed via **Docker** on **AWS Lambda**.

### ValueX – Wealth Management AI Microservice

Demo GitHub

FastAPI, OpenAI, Python, Server-Sent Events, pytest

- Engineered a **raw multi-agent AI pipeline** (no LangChain/LangGraph) with a **single-call structured intent classifier** routing across 8 specialist agents, a **pre-LLM safety guard** completing in under 10ms, and **real-time streaming via SSE** — all under a **\$0.05/query cost ceiling** at gpt-4.1 pricing with p95 first-token latency under 2s.
- Built a **labeled eval suite** against 100+ labeled queries achieving **≥85% routing accuracy** and **≥95% safety recall**, with full **CI coverage via mocked LLM** and a Portfolio Health agent producing structured risk outputs (concentration, benchmark alpha, actionable observations) across edge-case user profiles.

## EDUCATION

### Vishwakarma Institute of Information Technology

Jun 2022 - May 2026

Bachelor of Technology in Artificial Intelligence and Data Science

CGPA: 8.56

## AWARDS

**Grand Finalist • Smart India Hackathon 2024:** Top 5 nationally (AICTE, Ministry of Law & Justice)

**5x Hackathon Winner:** Recognized for developing scalable cloud-native and AI-driven solutions in various hackathons

## TECHNICAL SKILLS

**Languages:** JavaScript, TypeScript, Python, C++

**Tech:** React.js, Next.js, Node.js, FastAPI, GraphQL, REST APIs

**AI/ML:** LLM Integration, Prompt Engineering, Context Engineering, RAG, LangChain, LangGraph, Tree-sitter, MCP, Transformers, LLM Evaluation, Vector Databases

**Data:** PostgreSQL, MongoDB, Redis, Apache Kafka

**Cloud/DevOps:** AWS (EC2, S3, IAM, VPC, CloudWatch), GCP (Compute Engine), Docker, GitHub Actions CI/CD, Linux, Bash