

Chaitanya Mhetre

Software Engineer, AI Backend · Mumbai, India

mhetrechaitanya06@gmail.com · github.com/chaitanya-mhetre · linkedin.com/in/mhetrechaitanya · chaitanyamhetre.me · +91-7030486255

Summary

Full Stack Developer with 1+ year of full-time experience (plus internships) building production backend and AI systems (Python, FastAPI, PostgreSQL, Vertex AI, Gemini) on a platform with 2.2 crore registered users. I build LLM infrastructure end to end: hybrid-retrieval RAG with evaluation, permissioned tool-calling agents, and a multi-provider LLM gateway. Top 15 of 9,700+ teams, Google Agentic AI Day 2025.

Skills

Languages: Python, Java, Go, TypeScript, Dart, SQL

Backend: FastAPI, Spring Boot, NestJS, REST, SQLAlchemy/JPA, Kafka, Celery/arq, WebSockets

AI/LLM: Vertex AI, Gemini, LlamaIndex, OpenAI & Anthropic APIs, RAG (BM25 + pgvector hybrid, reranking), tool calling/agents, LLM evaluation, prompt-injection defence

Data: PostgreSQL (RLS, indexing, migrations), Redis (Streams, caching, rate limiting), pgvector, Firestore

Infra: GCP, AWS, Docker, GitHub Actions, Terraform, Kubernetes/Helm, Prometheus, Grafana, OpenTelemetry

Mobile/Web: Flutter, React, Next.js

Experience

Full Stack Developer, Agribid Private Limited, Mumbai

Aug 2025 – Present

Software Engineering Intern, May 2025 – Jul 2025. Platform with 2.2 crore registered farmers, traders and cooperatives (~10,000 active users); work spans Agribid and NestIQ, the group's AI interior-design product.

- Co-built, in a team of 4, a **Vertex AI + LlamaIndex RAG** decision-support system inside the live product; Python backend services and REST APIs on GCP and AWS serve 3,000–5,000 requests a day.
- Rebuilt NestIQ's Python/FastAPI Gemini image-generation service: backoff retries that skip safety refusals, failure classification into clear errors, and an image-quality gate (Pillow/OpenCV) with automatic credit refunds.
- Replaced random product selection with a **scored SQL retrieval step** (style, colour, budget) that sends the 8 most relevant catalog items to generation instead of 30 random ones, linking every generated room to products users can buy.
- Designed a **server-driven UI template system** (Flutter + Firestore): versioned templates, dev → prod promotion with one-tap rollback, Firebase custom-claim admin access, and a shared resolver package with 47 unit tests.
- Delivered the Agribid customer app in 5 phases (registration state machine for 4 customer types, MPIN security, auctions/bidding, marketplaces, mandi prices) with 82 logic tests; shipped Play Store and App Store releases.

Software Developer (Part-time), DK Techno's, Remote

Jan 2024 – Feb 2025

- Built Flutter and JavaScript app modules, REST API integrations and authentication workflows for client projects during my final year of B.E.

Projects (personal, github.com/chaitanya-mhetre)

rag-engine · Python, FastAPI, PostgreSQL/pgvector, Redis

- Hybrid retrieval with hand-written BM25 and pgvector (RRF fusion, reranking), cited answers, multi-tenant access control; rag-eval CLI measures recall@k, MRR and faithfulness and gates CI on regressions.

ai-agent-platform · Python, FastAPI, Postgres, Redis

- Tool-calling agent runtime with per-tool permissions, human approval, taint-based prompt-injection defence, SSRF-safe tools and crash-safe durable runs; 48-case eval suite (in offline tests the taint policy blocked every injection attempt).

ai-gateway · Python, FastAPI, Redis, Prometheus

- OpenAI-compatible gateway over 4 LLM providers with routing, fallback, circuit breaker, caching, quotas and cost tracking; an outage benchmark exposed an unenforced-timeout bug that failed 100% of requests with a hung provider, and the fix brought failures to 0.

Recova (payment collections) · Java 21, Spring Boot 3, Kafka, PostgreSQL, Redis, Next.js

- B2B receivables platform: JWT refresh rotation, transactional outbox to Kafka with idempotent consumers, Resilience4j around the gateway, signed webhooks; 60 Testcontainers tests; k6 measured ~687 payments/s at p95 21 ms with 0 reconciliation mismatches.

Education

B.E., Artificial Intelligence & Data Science, Dr. D. Y. Patil Institute of Technology, Pune

2022 – 2025

Diploma, Information Technology, Government Polytechnic Solapur

2019 – 2022

Achievements

Top 15 of 9,700+ teams (final pitching round), Google Agentic AI Day (Hack2Skill), Bengaluru

2025