# HEMANTH JANAPALA

AI Software Engineer · GenAI · RAG · LangGraph · Java · Spring Boot · Event-Driven Systems

San Francisco, CA — (940) 977-5188 — hemanthjanapala@gmail.com — LinkedIn — GitHub

## SUMMARY

AI Software Engineer with 3+ years building GenAI and data systems for production: RAG pipelines, LLM integrations, and LangGraph agents on event-driven microservices and high-throughput Azure pipelines serving millions of retail members, with a focus on idempotent processing, grounded retrieval, and proactive monitoring.

## PROFESSIONAL EXPERIENCE

**Walmart Global Tech** Jun 2025 – Present
*Software Development Engineer* *Bentonville, AR*

- Cut processing time from 5–6 minutes to under 3 minutes and eliminated Databricks costs by leading end-to-end migration of a core promotion-targeting data pipeline to an event-driven SQL architecture (Blob Storage → Data Factory → Event Hub → Spring Boot → Azure SQL), processing millions of records daily.
- Achieved 99.9%+ delivery to 3 downstream systems by building an event audit and recovery service from scratch, using SQL state tracking and automated retries to reprocess unacknowledged events within a 5-minute SLA.
- Enabled natural-language querying and validation of promotion and audit data by building an LLM-based conversational agent (LangGraph, Azure OpenAI), grounded in internal data sources.
- Delivered sub-second UI reads with exactly-once semantics by building a dual-store persistence layer pairing Cosmos DB for high-throughput writes with OpenSearch index mappings, using idempotent processing and dead-letter queue handling.
- Delivered a zero-downtime migration by engineering a fault-tolerant Spring Boot batching engine with idempotency, retries, and JDBC batching, and coordinating a dual-write cutover with the cart-validation team.
- Strengthened platform security by migrating two services to auto-rotated Azure workload identity and redesigning a vulnerable header-based service-to-service auth scheme with token verification.

**Neotechsoft** Jan 2025 – May 2025
*Software Engineer Intern* *Remote*

- Developed an AI-driven product insights feature for an e-commerce platform, analyzing customer view and purchase activity to identify trending and top-performing products.
- Built Python and FastAPI services to collect and process product interaction events (views, add-to-cart, purchases) and serve insights through REST APIs.

**KPMG US** Jul 2024 – Oct 2024
*Software Developer* *Montvale, NJ*

- Built a personalized GenAI chatbot using LangChain and Retrieval-Augmented Generation (RAG), letting users upload their own documents and query them in natural language with context-grounded answers.
- Engineered the document ingestion pipeline (parsing, chunking, and embedding files into a vector database), enabling accurate semantic retrieval across large, unstructured corporate documents.

**Ernst & Young (EY)** Mar 2021 – Nov 2022
*Software Engineer* *Hyderabad, India*

- Improved reliability and uptime of Java backend services for fintech and enterprise applications through ongoing design and maintenance.
- Accelerated feature delivery by building Spring Boot microservices integrating relational databases and MongoDB.
- Implemented Jenkins CI/CD pipelines automating build, test, and deployment, and resolved performance bottlenecks, improving scalability.

## PROJECTS

**AI Customer Support Copilot**
*Python, FastAPI, OpenAI API, PostgreSQL, Docker*

- Built an LLM-powered customer support application with REST APIs for contextual responses; achieved 10/12 correct classifications on a 12-query intent-classification evaluation set.
- Implemented request validation, prompt handling, structured error responses, and conversation management, with logging and automated tests to support reliability.

**Enterprise Document Intelligence Platform**
*Python, FastAPI, OpenAI API, PostgreSQL, Docker*

- Built a retrieval-augmented generation (RAG) pipeline for document question answering, combining semantic retrieval with source-grounded citations for answer traceability.
- Implemented document processing, embedding generation, API-key authentication, and per-IP rate limiting; added 86

passing tests, a 36-question evaluation harness, and a load-test run of 1,603 requests with zero errors.

**Agentic Research & Analysis Platform**
*Python, LangGraph, LangChain, OpenAI, RAG, PostgreSQL*

- Developed an agentic workflow using LangGraph to decompose user requests into multiple steps, coordinate retrieval and tool execution, and generate structured, cited outputs.

## SKILLS

- **AI / GenAI:** RAG, LangGraph, LangChain, OpenAI API, Azure OpenAI, Vector Databases, Prompt Engineering
- **Languages:** Java, Python, SQL, JavaScript
- **Frameworks:** Spring Boot, Spring MVC, FastAPI
- **Data & Messaging:** Kafka, Azure Event Hub, Data Factory, PostgreSQL, MySQL, SQL Server, Cosmos DB, MongoDB, OpenSearch
- **Cloud & DevOps:** Microsoft Azure, Docker, Kubernetes, GitLab CI/CD, Jenkins, Azure Monitor, Application Insights
- **Security & Architecture:** JWT, OAuth2, Azure AD Workload Identity, distributed systems, event-driven microservices

## EDUCATION

**University of North Texas** Jan 2023 – Dec 2024
MS, Computer Science Denton, TX

**VNR Vignana Jyothi Institute of Engineering & Technology** 2017 – 2021
BTech, Computer Science Hyderabad