ROSHAN SHETTY

Boston, MA — shetty.ros@northeastern.edu — 858-341-9356
roshanshetty.dev — linkedin.com/in/roshanshetty271 — github.com/roshanshetty271

SUMMARY

Software Developer with 3+ years building AI-powered platforms and enterprise applications. Delivered RAG-based document systems using OpenAI embeddings and reduced support tickets by 40%. Skilled in Python, Java, React, and distributed systems.

EDUCATION

Northeastern University – Boston, MA, USA Dec 2025
Master of Science in Information Systems — GPA: 3.8

Mumbai University – Mumbai, India Nov 2020
Bachelor of Engineering in Electronics and Telecommunication — GPA: 3.5

SKILLS

Languages: Python, Java, JavaScript, TypeScript, SQL
AI: RAG (Retrieval-Augmented Generation), LLMs, OpenAI API, Word2Vec, Embeddings, Semantic Search, Vector Databases
Backend & APIs: Flask, FastAPI, Spring Boot, Node.js, Express.js, REST APIs, Microservices
Databases: Qdrant, pgvector, Redis, Elasticsearch, PostgreSQL, MongoDB, Neo4j
Distributed Systems: Akka Cluster, Actor Model, RabbitMQ, Event-Driven Architecture
Frontend: React.js, Next.js, TypeScript, Redux, HTML5, CSS3, D3.js
DevOps & Tools: Docker, Docker Compose, CI/CD, Git, GitHub, Maven, Agile

PROFESSIONAL EXPERIENCE

Aosenuma – Texas, USA **Jan 2025 – May 2025**
AI Software Developer

- Built intelligent document querying system using RAG pipeline with OpenAI embeddings, Supabase/pgvector, and Neo4j knowledge graphs, automating processing of PDF, Excel, and Word files for enterprise clients.
- Developed semantic search enabling natural language queries with 300-dimensional embeddings and cosine similarity ranking, improving search relevance across large datasets.
- Created Python data pipelines with Flask REST APIs for document ingestion, integrating PostgreSQL and Redis to enable real-time analytics for stakeholders.
- Delivered full-stack platform (React/Next.js/TypeScript + Python) with 15 D3.js visualizations in 4-person Agile team, using Docker and CI/CD to accelerate feature delivery.

Capgemini – Mumbai, India **Nov 2020 – Jun 2023**
Software Developer

- Reduced support tickets by 40% building enterprise web applications (JavaScript, HTML5, CSS3, Java, SAP ABAP) serving 100+ global users with improved UX and reliability.
- Improved issue resolution time by 20% through performance optimization and database query tuning, enhancing system responsiveness.
- Managed complete SDLC in Agile environment; trained 5 team members, reducing onboarding time by 30% and improving team velocity.

PROJECTS

DocuMind – Distributed AI Document Intelligence System **Sep 2025 – Dec 2025**

- Architected distributed RAG system using Akka Cluster with fault-tolerant nodes and Actor Model, enabling high availability and horizontal scalability for concurrent query processing.
- Built semantic search pipeline with Word2Vec (300-dim embeddings), Qdrant vector database, and sliding-window chunking (60% overlap) with keyword re-ranking, achieving high recall for document retrieval.
- Integrated OpenAI GPT via Spring AI with custom prompt engineering, reducing hallucination by grounding responses in source documents.
- Developed Java Spring Boot REST API with async endpoints and React frontend, containerized with Docker for consistent deployments.

Medical Insurance Plan API – RESTful Microservice **Sep 2025 – Dec 2025**

- Designed Python Flask microservice with Redis decomposed storage enabling O(1) partial updates, reducing write latency for complex nested documents.
- Built smart PATCH operations with deep merge logic; integrated RabbitMQ for async processing; secured with Google OAuth 2.0/RS256 JWT for enterprise-grade authentication.
- Orchestrated 4-service Docker Compose infrastructure with JSON Schema validation and ETag-based caching, ensuring data integrity and optimized cache efficiency.

RadioX Healthcare Platform **Sep 2023 – Dec 2023**

- Reduced medical imaging processing time by 25% with full-stack platform (React, TypeScript, Redux, Express/Node.js), improving clinical workflow efficiency.
- Implemented 22 REST API endpoints across 6 MongoDB collections with 9 interactive visualization tools, enabling data-driven clinical insights.