Sushma Avala

[sushmaa297@gmail.com](mailto:sushmaa297@gmail.com) | (216) 640-4221 | USA | [LinkedIn](https://www.linkedin.com/in/sushmareddyavala/)

# Technical Skills

**Languages:** Java, JavaScript, Python, HTML, CSS

**Databases:** SQL, Postgres, MongoDB, Redis

**Frameworks, Libraries & Technologies:** Spring Boot, Spring, React, NodeJS, AWS, Docker, Kubernetes, Kafka, Redis

**AI/ML & NLP Tools:** LLMs (OpenAI, Hugging Face), Retrieval-Augmented Generation (RAG), FAISS, Pinecone, Semantic Search

**Operating Systems:** Linux, AIX, Windows

# Professional Experience

**Full Stack Developer**, *Cleveland State University* 09/2023 – 05/2025|Ohio, USA

**Project: Faculty Research Collaboration Portal**

* Engineered an AI-driven internal web platform that streamlined faculty collaboration, reducing manual coordination efforts by **40%** and enabling real-time sharing of research outputs across departments.
* Built end-to-end features with **Java (Spring Boot)** and **Python (FastAPI + LangChain)** for backend microservices, and **React.js + Redux Toolkit** for the frontend, boosting user engagement and cutting response latency by **35%**.
* Designed and deployed a **RAG (Retrieval-Augmented Generation) pipeline** leveraging **LLMs (OpenAI/GPT, Hugging Face models)** with **MySQL + Vector DB (FAISS/Pinecone)**, enabling semantic search and Q&A over 10,000+ research documents.
* Developed secure **REST and GraphQL APIs** for project creation, member roles, document uploads, and AI-powered summarization, improving research document retrieval time by **50%**.
* Implemented **role-based access control** for faculty, graduate assistants, and department admins using **Spring Security, JWT, and OAuth2.0**, ensuring compliance with university data security policies.
* Automated deployment pipelines with **Docker and GitHub Actions**, reducing release cycles from weeks to **under 2 days** and improving reliability of AI model updates.

**Full Stack Java Developer**

**Client: CodeFish Technologies, Hyd , India**09/2021 – 01/2023

* Designed and implemented **Spring Boot microservices (Java, Node.js)** for product catalog, cart, orders, and payments, exposing REST APIs consumed by **React.js/Angular frontends**.
* Integrated **Stripe and Razorpay payment gateways** with transaction validation, retry logic, and webhook-based reconciliation, ensuring PCI-compliant payment flows.
* Optimized **MySQL queries and indexing strategies** and applied **Redis caching + AWS CloudFront CDN**, reducing response latency by **40%** under peak traffic loads.
* Containerized services using **Docker**, orchestrated deployments via **Jenkins CI/CD pipelines**, and provisioned infrastructure on **AWS EC2, RDS, S3, CloudFront**.
* Implemented **monitoring, logging, and alerting** with **AWS CloudWatch, Nginx logs, and Linux shell scripts**, enabling proactive incident detection and resolution.

# Projects

**AI Research Summarizer**

* Built a full-stack **AI-powered research assistant** using **Python (FastAPI + LangChain)** for backend services, **React.js** for frontend, and **MySQL + FAISS** for structured + vector storage.
* Implemented a **Retrieval-Augmented Generation (RAG) pipeline** integrating **LLMs (OpenAI GPT + Hugging Face models)** to summarize research papers and enable semantic search over 15,000+ academic documents.
* Developed secure APIs to handle research uploads, intelligent metadata extraction, and AI-driven Q&A, improving faculty research productivity by **45%**.
* Containerized with **Docker** and deployed on **AWS (EC2, S3, Lambda)** with CI/CD pipelines using **GitHub Actions**, cutting deployment cycles to under 2 days.

**Conversational AI Student Helpdesk**

* Designed and developed a **conversational AI system** to assist students with course FAQs, scheduling, and document guidance using **Java (Spring Boot) + Python (FastAPI)** for backend orchestration.
* Integrated **LLMs with a RAG pipeline** connected to **university course material and policies** stored in **PostgreSQL + Pinecone**, enabling contextual responses with **90%+ accuracy**.
* Built a **React.js + TailwindCSS** frontend chat interface supporting real-time AI response streaming via **WebSockets**.
* Improved query resolution speed by **60%**, reducing dependency on manual staff support.
* Deployed on **AWS (ECS, RDS, S3, CloudWatch)** with containerization via Docker and CI/CD pipelines for continuous updates.

# Education

**Master of Science in Computer and Information Science**

Cleveland State University

# Certifications

* [AWS Certified Solutions Architect – Associate | Amazon Web Services](https://www.credly.com/badges/79ef9114-0eb2-48d3-9bb6-4ed58b5be375/linked_in_profile)