# Shruti Varade

Plantation, FL, USA, 33324 | (617)606-8865 | varadeshruti27@gmail.com | https://shrutivarade.github.io/portfolio/

### **SUMMARY**

Full Stack Software Engineer with over 5 years of experience delivering end-to-end web and cloud solutions. Proficient in Java, Fast API and JavaScript/TypeScript, designing scalable, serverless AWS architectures and expertise in multithreading, systems design, and debugging.

Experienced in **Al/ML integration**, building **RAG** pipelines, leveraging **LLMs**, **Vector DB** for my next **Agentic AI application**, and tools like **Cursor** for improving productivity. I am <u>currently prototyping a custom AI-powered application to automate complex engineering tasks.</u>

#### **EXPERIENCE**

## Full Stack AI Platform Engineer [Early-stage Startup]

March 2025 - Present

Boston, MA, USA

Curvepoint.ai

- o Designed and deployed a scalable, highly available, serverless AWS backend (RESTful API, Lambda, Kinesis, S3, DynamoDB, Web Sockets) to ingest over ~21,600 ML model outputs per hour from Runpod and stream actionable insights to the frontend for analytical purposes.
- o Helped secure \$1 M in pre-seed funding, enabling rapid expansion of our MVP and accelerated go-to-market for camera-free safety solutions.
- o Worked at a fast-paced startup environment on Wi-Fi based user interface using React, NodeJS, Vercel enabling real-time, privacy-preserving threat detection in schools via passively interpreting Wi-Fi sensing technology—no cameras or wearables required.
- o Developed and maintained **infrastructure-as-code** using **SST** (Serverless Stack Toolkit) to build new things from **0-1 quickly**, accelerating cloud migrations by **~60%** (**using "write-once**, **deploy-everywhere"**) and improving resource maintainability for the existing AWS infrastructure.

#### Software Engineer

July 2024 - November 2024

Massachusetts General Hospital – Harvard Medical School [Health Tech]

Boston, MA, USA

- o Developed a 3D visualization tool using Typescript to display the Brain MRI data for Neural Connections on the web browser.
- o Reading and understanding the legacy code to add additional feature to the existing platform.
- o Reduced the time to fetch 1GB of data from AWS Cloud from 15 seconds to about 3 seconds, achieving an 80.91% improvement.

#### Software Engineer (AI/ML)

May 2023 - March 2025

Machine Psychology Lab, University of Massachusetts, Boston [Health Tech]

Boston, MA, USA

- o Developed Boostlet.js, **a JavaScript library** enables real-time **Al/ML** processing of medical images in web browsers, created a comprehensive solution supports various data formats (e.g., EEGs, ECGs, X-rays, MRIs, and CTs).
- o Designed a feature for image segmentation, data visualization, and integration with **Hugging Face's machine-learning models** to enhance existing medical image processing libraries with client-side processing capabilities.
- o Improved teamwork and iterative development processes using **Git** for Version Control, **NodeJS** for backend development, **GitHub Actions** for test automation and **Submodules** for external library integration.

**Software Engineer** 

Sept 2020 - June 2022

TATA Group

Mumbai, India

- o Implemented a clean and simple full stack recommendation platform following microservices based architecture with React as frontend framework, Spring Boot for developing a Java microservices for user management, analytics and recommendation engine.
- o Implemented JSON-based user data storage with in-memory session handling and RESTful APIs for frontend-backend data exchange.
- o Used transport layer like http, containerized using docker and orchestrated with docker compose for local development and integration.
- o Established a testing protocol, **JUnit and Jest for unit testing**, and **python for integration testing** to validate workflows across the application.

## **EDUCATION**

## Master's Degree of Science in Computer Science

Sept 2022 - May 2024

University of Massachusetts, Boston | GPA 3.9/4.0 (Teaching Assistant)

Boston, MA

## **PROJECTS**

(Code Assistant) Continuous chat bot running in terminal using Agentic AI - OpenAI API, chain of thought for reasoning

RAG app that indexes the given pdf doc and performs vector search - Tokenizer, Vector Embedding, LangChain, Vector DB Qdrant

Matching Researchers with Students via Machine Learning deployed Web App - Python, Natural Language Processing, Streamlit.py, LLM, Al

An app that prevents race conditions and deadlocks - Java, Design patterns, Multithreading, Memory Management, Stream API

Health Monitoring web dashboard to display graphs and charts - Python, Django - MVT architecture, Chart.js, Postgres SQL

#### **SKILLS**

Al Skills: LLM, RAG, Qdrant, Tokenization, Vector Embeddings, Vector DB, LangChain, LangGraph

Al Tools: OpenAl API, ChatGPT, Claude, Perplexity, Cursor, GitHub Copilot, Midjourney, NotebookLM, Zapier, Make.com, n8n,

Frontend: HTML, CSS, JavaScript, Jest, TypeScript, React.js, Next.js, MUI, ECMAScript, Restful APIs, WebGL, SEO, Sanity, Prisma ORM, Redis Backend: Java, Android, JUnit 5, Apache Ant, Gradle, Tomcat, Python, Django, Express, Node.js, C++, REST API, JSON, HTTP, Docker, Linux Databases: SQL, MySQL Relational, Postgres NoSQL, GraphQL, MongoDB, Cloud Computing – Amazon Web Services, Kafka, RabbitMQ

Tools: AI/ML, Git, GitHub, Postman, CI/CD pipeline, Agile Methodology, SDLC, Data Structures and Algorithms, MATLAB