

Shreenidhi Acharya

New York, USA | shreenidhiacharya7@gmail.com | (585) 490-5750 | [LinkedIn](#) | [Github](#)

EDUCATION

Master of Science, Computer Science, Rochester Institute of Technology, Rochester, NY (August 2022 - May 2025) | GPA 3.8 / 4.0
Bachelors of Engineering, Computer Engineering, University of Mumbai, India (June 2015 - May 2019) | GPA 4.0 / 4.0

TECHNICAL SKILLS

Backend & Systems:	Python, Java (Spring Boot), FastAPI, Asynchronous Programming, Shell Scripting, Nginx
Data & Databases:	Redis Stack VSS, PostgreSQL, MongoDB, MySQL, Pydantic, Apache Airflow, Alembic
AI & LLM Orchestration:	RAG (Retrieval-Augmented Generation), Agentic AI Workflows, LangChain, Semantic Caching
Web Technologies:	TypeScript, ReactJS, HTML5/CSS3, RESTful APIs, Figma, Lovable
Cloud & DevOps:	AWS (Boto3/SDK), Docker, CI/CD (Jenkins) Prometheus, Grafana, Git/GitHub, Linux/Unix
Networking & Security:	SSH, TLS/SSL, TCP/IP, HTTP, RPC, gRPC
AI-Assisted Development Tools:	Claude, Gemini LLM, Cline CLI, GitHub Copilot, Cursor

PROFESSIONAL EXPERIENCE

Saayam For All

San Jose (Remote), USA

Software Engineer

June 2025 – Present

- Engineered a stateful notification feature for the volunteer microservice using Java Spring Boot, designing 2 core REST APIs to retrieve notification counts and content for timely user alerts from PostgreSQL database.
- Optimized notification delivery by persisting a watermark via Spring Data JPA to implement timestamp-based retrieval logic and reduce redundant data fetches.

London Stock Exchange Group

Buffalo, NY

Software Engineer Intern

September 2024 – December 2024

- Engineered a hybrid AWS/on-prem data pipeline with Python backend services and Lambda, automating the ingestion of 100K+ index records to PostgreSQL database and cutting latency by 50% (from 20 to 10 minutes).
- Architected and implemented 13 REST APIs on AWS Lambda, orchestrating critical database workflows for 5 internal data teams while maintaining 99.9% uptime on 1K daily requests.
- Deployed 7+ serverless REST API endpoints using AWS Lambda and Python to process over 100,000 data transfer transactions weekly on Linux servers.

Software and Data Engineer Intern

September 2023 – December 2023

- Automated ETL workflows using Python and Apache Airflow, increasing data pipelines reliability by 60% and scaling daily data ingestion of 100K+ records to PostgreSQL database to support business analytics.
- Reduced ETL workflows incident response time from hours to under 15 minutes by implementing four automated pre- and post-checks (e.g., holiday checks, file validation with Pandas) with unit tests in PyTest.
- Implemented a multibranch CI/CD pipeline using Shell scripting in Jenkins that reduced the deployment of Airflow-based data pipelines by 50% in the development and production environments.

Tata Consultancy Services

Mumbai, IN

Software Engineer

June 2019 – July 2022

- Accelerated integration testing workflows by partnering with QA teams, leveraging Python and shell scripting, boosting unit test execution speed by 65%, and reducing post-release defects by 25%.
- Introduced a robust CI/CD pipeline using Python scripting in Jenkins, with integrated automated test suites for system installations, enabling reliable server setups for a 25-member engineering team.
- Designed and implemented a backup solution for Linux servers in High Available setups using advanced shell scripting, reducing backup and restore time by 80% (from 20 to 4 hours) and significantly minimizing system downtime.

PROJECTS

Panther AI (Hackathon Project), [Github](#)

December 2025 – Present

- Led a 7-day sprint to architect and deliver a functional MVP for a scalable AI-powered Freelance Operating System that automates contract management, reporting, and stakeholder communication with a **TypeScript and React** frontend.
- Reduced LLM API costs by **25%** and response latency by **40%** by implementing a **Redis-based semantic caching layer**, bypassing redundant model inference while maintaining **sub-100ms responses**.
- Achieved a 100% query resolution rate** by engineering an **Agentic fallback workflow** using **Gemini 1.5 Flash tool-calling**. Developed an automated **SMTP escalation agent** that autonomously routes complex queries to human experts.

RAG-Powered Q&A Assistant: Adult Immunization, [Personal Project](#)

August 2025 – September 2025

- Built a Q&A RAG application using **Python**, FastAPI, Pydantic, **LangChain**, **ReactJS**, and **Chroma vector DB**, integrating **Gemini Flash 2.5 LLM with Hugging Face embeddings** on the WHO immunization dataset.
- Implemented **Redis Vector Search** for semantic user queries caching, reducing redundant LLM calls and improving response efficiency. Utilized dockerized Prometheus and Grafana monitoring to track cache hits, misses, and LLM calls, optimizing application performance.

AI-Powered Employment Recommendation System, [Capstone Project](#)

January 2025 – April 2025

- Developed and deployed a full-stack job recommendation system using **ReactJS**, FastAPI (**Python**) REST API services, and **AWS (S3 for frontend hosting, EC2 for backend APIs)** to enable resume analysis, role matching, and upskilling suggestions.
- Optimized K-Means clustering using TF-IDF and Word2Vec vectorization, combined with Cosine, Jaccard, and Euclidean distance metrics, achieving **90%** recommendation accuracy across **22** job domains.