

LEELA SRIJA ALLA

+1 (716) 705-1957 · Seattle, WA, 98119 · srijaalla1001@gmail.com · [LinkedIn](#) · [Portfolio](#) · [GitHub](#)

EDUCATION

University at Buffalo, SUNY

Aug 2023 - Dec 2024

Masters in Computer Science (AI/ML track)

- **Relevant Coursework:** Operating Systems, Computer Security, Analysis of Algorithms, Machine Learning, Deep Learning, Natural Language Processing, Data Models and Query Languages, Data Intensive Computing, Computer Vision

University of Hyderabad, India

Mar 2017 - Jul 2022

Integrated Masters in Technology, Computer Science

- **Publication :** Opinion Maximization in Signed Social Networks Using Centrality Measures and Clustering Techniques
- **Research Assistant:** Collaborated with Dr. Anjeneya Swami in developing heuristic algorithms with existing diffusion models in Social Network Analysis.

SKILLS & OTHER

Languages: C, Java, Python, JavaScript, PHP, HTML, CSS, Shell Scripting; **Frameworks:** TensorFlow, Django, Flask, Scikit-learn, NetworkX, Keras, PyTorch, Hadoop, Langchain, ReactJS, TypeScript, Node.js, Ruby on Rails; **Databases:** MySQL, MongoDB, PostgreSQL, Scala, Pinecone; **Technologies/Environments:** Unix/Linux, Windows, CI/CD, AWS EC2, SageMaker, Docker, Azure, Agile Scrum, Kubernetes, Git, Latex, Tableau, Microsoft Suite

PROFESSIONAL EXPERIENCE

Senior AI Engineer Intern, Flow, New York

Aug 2024 - Present

- Developed backend for Flow's corporate website **end-to-end** using **Django**, including designing the architecture.
- Connected backend services to **Salesforce CRM** for efficient business management. Contributed to creating responsive landing and support pages using **ReactJS**, ensuring a seamless user experience.

Tech Stack: ReactJS, Django, Node.js, Scrum, Python, HTML, TailWind CSS, JavaScript, Docker, Kubernetes, Selenium

Software Engineer, DKube, Hyderabad

Aug 2022 - Apr 2023

- Enhanced DKube UI using **ReactJS and TypeScript**, reducing model deployment time of ML models by 9%.
- Optimized backend with Node.js and **Fast API**, improving API response times by 20%, and increasing server performance.
- Implemented **CI/CD with Jenkins**, automating **Docker** image creation and **zero-downtime deployments on Kubernetes**.
- Integrated **Prometheus and Grafana** for monitoring, enhancing early issue detection and system performance reliability.
- Created a microservice architecture with Docker and Kubernetes, ensuring scalable deployment across clusters.

Tech Stack: ReactJS, TypeScript, Fast API, Python, Node.js, Express.js, Docker, Kubernetes, Prometheus, Grafana, Linux, AWS EC2, Jenkins, Jira

Full Stack Web Developer (Intern), Jindal Trading Co., Delhi

Oct 2020 - Dec 2020

- Headed e-commerce platform overhaul using **ReactJS**, and **PostgreSQL**, boosting daily traffic by 21% and sales by 15%.
- Incorporating unit testing with Jest and end-to-end testing with **Cypress**, enhancing code quality and reducing bugs.
- Established **CI/CD pipelines** with **Jenkins** and **Docker**, boosting development efficiency and ensuring seamless integration.
- **Led** a team of **4 developers**, managing milestones and delegating tasks to ensure project success and business growth.

Tech Stack: ReactJS, PostgreSQL, Cypress, Docker, Linux, Jest, Jenkins

PROJECTS

Patient-Centric AI Health Assistant with Knowledge Graph (RAG)(Langchain, PostgreSQL, Django)

Sep 2023 - Oct 2023

- Built a health-focused chatbot using **OpenAI's LLM** and **Django**, where patient data in **PostgreSQL** is used for interactions about their conditions and treatment plans.
- Integrated **Neo4j** for a knowledge graph, enabling efficient retrieval of patient-specific data. Leveraged **Langchain** and **Graph RAG** to dynamically query relevant information, enhancing response accuracy and continuity.

Personalized LinkedIn Bot using Gen AI (Selenium, Gemini AI)

Mar 2024 - May 2024

- Developed and deployed a personalized AI chatbot using Google's **Gemini 1.0 Pro LLM** and **AI studio**, fine-tuning the model with my WhatsApp conversations.
- Automated the bot to reply to comments on a LinkedIn post, mimicking my conversational style for public interaction.

Textual Entailment in Clinical Trial Data (PyTorch, Hugging Face)

Mar 2024 - May 2024

- Achieved an F1 Score of 72.34 by building multi-evidence NLI model with Clinical Trial Reports for breast cancer treatments.
- Improved numerical reasoning by fine-tuning **DeBERTa-v3**, applying Sebis Pipeline Model and experimented with different **BERT** models and **LLMs**.

RAG Based Conversational Framework (Langchain, Hugging Face, Pinecone, Streamlit, AWS)

Feb 2024 - Mar 2024

- Engineered **Generative AI** for academic planning, using Langchain and Pinecone for vector database creation and retrieval.
- Created and deployed a chatbot with **Meta Llama-7b-chat-hf LLM** on **AWS Sagemaker**, enhancing query responses using **all-MiniLM-L6-v2** sentence transformers.
- Implemented an end-to-end chatbot using **OpenAI** and **Flask, TypeScript** for cost efficiency and performance.

AWARDS, VOLUNTEER AND LEADERSHIP ACTIVITIES

- Received the **Best Student Paper Award** in the Intelligent Technology track for my paper titled - "Opinion Maximization in Signed Social Networks using Centrality Measures and Clustering Techniques" at the **19th International Conference on Distributed Computing and Intelligent Technology**
- **Winner of UB AI hackathon by Blackstone Launchpad** for creating an end-to-end OCR and RAG based AI solution for the NPO Lighthouse Medical free clinic.