Sri Harsha Vallabhaneni

vsriharsha814@gmail.com | +13035226589 | linkedin.com/in/vallabhanenisriharsha | github.com/vsriharsha814 | vsriharsha814.vercel.app

Education

University of Colorado Boulder   **Aug 2024 – May 2026**
Master of Science in Computer Science (GPA: 4.0)   *Boulder, CO*
Relevant Coursework: Datacenter Scale Computing, NLP, Big Data Architecture, Neural Networks and Deep Learning, Data Mining

Birla Institute of Technology and Science (BITS Pilani)   **Aug 2018 – May 2022**
Bachelor of Engineering in Electronics and Communication Engineering   *Hyderabad, India*
Relevant Coursework: Object Oriented Programming, Digital Image Processing, Operating Systems, Software Engineering, IoT

Skills

AI / Machine Learning: GenAI, LLMs, RAG, NLP, LangChain, PyTorch, Sentence Transformers, Vector Databases
Backend / Systems: FastAPI, Node.js, Flask, Microservices, REST APIs, SQL
Cloud / Infrastructure: Docker, Kubernetes, Kafka, Redis, GCP, Cloud Deployment
Languages / Tools: Python, JavaScript, TypeScript, Bash/Shell, MongoDB, HTML/CSS

Professional Experience

AI Project Engineer / Developer at Cisco Systems (Capstone Partnership)   **Aug 2025 – May 2026**
- Implemented an **enterprise-grade incident triage system** for Cisco's internal infrastructure, utilizing **LLMs** and **semantic search** to automate anomaly detection in large-scale distributed logs.
- Architected containerized **FastAPI microservices** using **Docker**, **MongoDB**, and internal API gateways, embedding anomaly detection into diagnostic workflows and **reducing incident triage latency**.

AI Engineering Intern at Empowered Margins Inc. (Colorado Springs, CO, USA)   **May 2025 – Jul 2025**
- Built and deployed an **agentic AI decision pipeline** using **Flask**, **LangChain**, and **Sentence Transformers** to extract and normalize data from unstructured Excel files, reducing manual processing by **85%**.
- Engineered multi-agent workflows for **semantic header mapping** and field detection, improving consistency across diverse vendor formats and outperforming rule-based baselines by **40%**.

Software Development Engineer 2 (SDE → SDE2) at Darwinbox (Hyderabad, India)   **Jun 2022 – Jun 2024**
- Led development of **RAG-based Generative AI systems** using **LangChain**, **OpenAI APIs**, and **Pinecone**, achieving **95% hallucination reduction** through retrieval grounding and prompt orchestration.
- Developed Microsoft Teams-integrated chatbots and responsive UIs using **Node.js**, **MongoDB**, **Bot Framework**, **JavaScript**, **HTML**, and **CSS**, boosting user engagement by **50%** and browser interactions by **30%**.

Software Developer Intern at Malhar Industries (Nagpur, India)   **May 2020 – Jul 2020**
- Led a team of 5 to build a full-stack **Django–MySQL CRM** with **RESTful APIs**, **Bootstrap UI**, and advanced reporting dashboards, streamlining sales ops, enhancing client engagement by **30%**, and cutting reporting time by **60%**.

Projects

- **Exactly-Once Event Processing System**
  Implemented **exactly-once event processing** using **idempotency keys** and a persistent **deduplication state store**, ensuring correctness under retries and redelivery. Analyzed **failure windows** (crash before/after acknowledgment) and validated behavior under **at-least-once delivery semantics** using failure injection and invariants.

- **FaithCircle – Multi-User Mobile Platform & Admin Dashboard**
  Built a **production-grade** mobile application with **collaborative** user flows, **authentication**, and persistent **state management**. Designed and deployed a web-based **admin dashboard** on Vercel to manage live application content, quizzes, and metadata without app redeploys. Implemented **role-based access**, **real-time data updates**, and **scalable** backend APIs to support multi-user collaboration and feature expansion.

- **Urban Crime Forecasting and Hotspot Detection**
  Developed a full ML pipeline using **Python**, **scikit-learn**, **XGBoost**, and **Pandas** to predict high-risk crime zones using 1M+ Chicago crime records; performed feature engineering, cross-validation, and hyperparameter tuning, achieving an **F1-score of 0.83**. Built geospatial heatmaps and temporal trends using **Folium**, **Seaborn**, and **Matplotlib**.