Srikar Veluvali

Hyderabad, Telangana, India
+91 ********** srikarv100@gmail.com LinkedIn GitHub LeetCode HuggingFace Portfolio

Education

Keshav Memorial Institute of Technology, Hyderabad **2022 - Present**
B.Tech in Information Technology *9.68 CGPA*

Sri Chaitanya Junior College, Hyderabad **2020 - 2022**
Maths Physics Chemistry (MPC) *98.4 %*

Experience

Microsoft **Jan 2025 – Present**
Research Intern *Bengaluru, Karnataka*

- Researching on **optimization techniques** to enhance the efficiency and scalability of **GPU Kernels**, with a focus on minimizing latency and maximizing throughput of **LLMs**.
- Worked on **parsing, re-structuring, and improving the infrastructure of compiler intermediate codes**, favoring performance improvement using **Graph Transformations**.

Defence Research Development Laboratory - DRDO **Jun 2024 – Dec 2024**
HPC Software Engineering Intern *Hyderabad, Telangana*

- Optimized **GPU processing time by 25.93% for CFD simulations**, significantly reducing missile flow simulation runtime from **30 to 22 days**, accelerating project timelines.
- Designed and refined **parallel computing algorithms** in collaboration with engineers, achieving higher **GPU core utilization** and enhanced execution efficiency for fluid dynamics simulations.
- Performed **comprehensive performance profiling, debugging, and kernel optimization** using advanced tools like **gprof**, ensuring efficient execution and stability for **Reynold's Equations**.

Projects

Astor AI: A Chatbot for Medical Queries | *LLMs, Generative AI, React.js, Flask, Fine-tuning*

- Designed and developed a medical chatbot tailored to answer complex medical inquiries using the **Llama 3** language model, which was **fine-tuned specifically on medical datasets** to enhance the model's relevance and accuracy in the medical domain.
- Implemented **Retrieval-Augmented Generation (RAG)**, integrating external medical databases to ground responses in reliable, up-to-date information and select research papers. This approach ensured that the responses were medically sound. This model has also achieved **400+ downloads on Huggingface**.

Heart Health Web Application | *MERN, Machine Learning, Data Science, Flask*

- Developed an advanced **MERN stack web application** for **heart disease prediction** and promoting heart health through personalized recommendations, utilizing **machine learning models** trained on large medical datasets to analyze key risk factors and predict heart disease likelihood.
- Integrated **Google's Gemini AI** to deliver tailored **dietary and exercise plans**, providing users with a customized and engaging experience based on their unique health profiles.
- Implemented a **location-based feature** to help users locate nearby healthcare services, enhancing **accessibility to cardiologists and medical facilities**, and earned a **nomination for the People's Choice Award** in the **Google Gemini API Developer Competition**.

Technical Skills

Programming Languages: C/C++, Python, Java, JavaScript, Go (Golang)
Web Development: HTML, CSS, Node.js, React.js, Express.js, Flask, Next.js, Bootstrap, Tailwind
Databases & Data Management: MySQL, MongoDB, Chroma, Pinecone
Tools & Platforms: Azure, AWS, Git, GitHub, Docker, Postman, Tableau, Unix/Linux, Google Colab, StarUML
Relevant Coursework: Software Engineering, Computer Networks, Operating Systems, Parallel Programming, DBMS

Achievements

- Solved **over 800 problems on Leetcode** achieving a **competitive programming rating of 1713**.
- Heart Health was nominated for **People's Choice Award in the Google Gemini API Developer Competition**.
- Awarded **Silver Medal (Top 5%) in Python for Data Science** from **IIT Madras**, scoring 84%
- Awarded **Silver Medal (Top 5%) in Java Programming** from **IIT Kharagpur**, scoring 78%