Saarang Srinivasan

510-706-1247 | srini158@purdue.edu | saarang123.github.io | LinkedIn/saarang-srinivasan | GitHub/saarang123 | US Citizen

Education

Purdue University — Expected Graduation: May 2026
Bachelor of Science in Computer Science (Honors) *GPA: 3.94* — *West Lafayette, Indiana*

Concentrations: Machine Intelligence, Algorithms, Systems Software
Relevant Coursework: Data Structures & Algorithms, Data Mining & Machine Learning, Operating Systems, ML Systems (G), Motion Planning (G), Introduction to AI, Randomized Algorithms (G), Advanced Theory of Algorithms, Systems Programming, Reasoning About Programs (G), Theory of Computation, Parallel Computing (G), Compilers, Deep Learning (G). (G - graduate level)

Experience

AI/ML Engineering Intern — Dec 2024 – Present
Avoca AI (YC '23) — *New York, NY / Remote*
- Working with AI Consumer Support team on sales call outcome analysis using NLP (RoBERTa, DeBERTa among others).
- Optimizing AI sales agent through classification of call history to analyze mistakes and potential improvements.

Software Engineering Intern — May 2024 – Aug 2024
Collins Aerospace / Raytheon Technologies — *Cedar Rapids, IA*
- Built and certified flight control software for the military Boeing KC-46 Pegasus.
- Designed and implemented a new hardware configuration format for KC-46 Pegasus start-up, reducing file size by 72%.
- Created C/C++ tests to verify correctness of flight software based on composition analysis and MC/DC coverage.
- Automated data parsing and tracking using Python and ML techniques, improving speed and accuracy by 30%.

Software Engineering Intern — May 2023 – June 2023
Hacklab Solutions Pvt Ltd — *Bangalore, India*
- Built containerized AI product architecture with server/client modules using Docker and Kubernetes for rapid deployment.
- Designed a main server integrated with a AI workers, front-end dashboard, supporting Postgres/Redis for persistent storage.
- Implemented self-healing nodes, CI/CD pipelines, and load balancing, ensuring scalability (MLOps/DevOps).
- Optimized memory accesses, boosting deployment speed by 80% and reducing latency by 30%.

Projects

GNN Cost Model for Tensor Program Optimization | *PyTorch Geometric, TVM, NLP, Multithreading* — Nov 2024
- Designed a Graph Neural Network (GNN)-based cost model to predict runtime of tensor programs.
- Integrated into TVM's search framework, model auto-optimizes tensor programs for deep learning on NVIDIA V100 GPUs.
- Engineered features from TensorIR Abstract Syntax Trees (ASTs) using FastText embeddings trained on random walks.
- Surpassed existing runtime prediction model XGBoost's accuracy and validated model on the TenSet dataset.
- Slides: https://tinyurl.com/GNN-Project, trying to publish

Pulse - Convert Lectures to TikToks | *Next.js, Flask, PyTorch, OpenCV, Firebase, MongoDB* — Jan 2024
- Developed a React.js web app to convert lecture videos/slides into TikTok-style clips, summarizing key topics.
- Built a robust two-layer MongoDB/Firebase DB architecture to increase video generation speed by 80%.
- Integrated with React Native iOS app for viewing generated videos, including deepfake videos and tweet summaries.
- Utilized OpenAI's Whisper API and Modal/PyTorch for video processing and Firebase/MongoDB for data storage.

TinyVerify: Translation Validation for Tinygrad | *Z3 SMT Library, Formal Verification, ML Compilation* — Nov 2024
- Built translation validation framework for the Tinygrad ML Compiler using SMT solvers (Z3).
- Verified correctness of GPU kernel optimizations on the Tinygrad Intermediate Representation.
- Extended support for ALU/memory ops, FP/int/vector types, enabling rigorous formal verification.
- Added support for GPU-specific thread variables and CPU control flow and range operations.
- Slides: https://tinyurl.com/tinyverify, trying to publish

Honors & Awards

Gold Medalist: Indian National Olympiad in Informatics 2022 (Top 10 in India).
ICPC Regionals: Rank 18 representing Purdue in the ACM International Collegiate Programming Contest Regionals 2022/23.
International Olympiad in Informatics Training Camp: Top 30 in India to be selected in 2021 and 2022.

Technical Skills

Languages: Python, C/C++, Java, JavaScript, Unix, R, MATLAB, ARM/x86, HTML/CSS, Rust
Frameworks/Tools: Pandas, Open-CV, Torch, DGL, Docker, Kubernetes, SQL, Flask, Firebase, Git, React.js, Node.js, Linux
Certifications: MITx: 6.431x Probability and Uncertainty of Data, Data Structures & Algorithms Certification by CodeChef
Fields: System for ML, NLP, Machine Learning, Computer Vision, Full Stack Development, Algorithms

Activities

Course Development: CS381 Analysis of Algorithms (Fall '24, 400+ students), CS182 Discrete Math (Spring '24, 800+ students)
Teaching Assistant: CS381 (Fall '24), CS182 (Spring '24), CS311 Competitive Programming 2 (Spring '23)
Research: With Dr. Simina Branzei on developing reinforcement learning algorithms for the repeated fair division problem.
USACO Tutor: Instruct Bay Area students in advanced algorithms for the USA Computing Olympiad with X-Camp Academy
Club Treasurer/Officer: Competitive Programming Union Club 2023-2024