

# SARANG SRIDHAR

+1-267-279-8568    saresri@seas.upenn.edu    linkedin.com/in/sarang-sridhar    sarang-sridhar.github.io

## Education

University of Pennsylvania

Aug 2024 - May 2026

M.S.E in Computer and Information Sciences

Philadelphia, PA

Birla Institute of Technology and Science Pilani

Sep 2020 - May 2024

B.E in Computer Science and Engineering

Pilani, India

## Experience

University of Pennsylvania | Graduate Research Assistant

Advisor: Prof. Insup Lee

Jan 2025 – Present

- Building an **AI agent for ophthalmology**, leveraging **reasoning-oriented reinforcement learning**, to execute necessary tool calls, retrieve critical info, & dynamically modify inputs to a **multimodal LLM** to ensure accurate diagnosis

Advisor: Prof. Kevin Johnson

Aug 2024 – Dec 2024

- Designed an **evaluation framework** to benchmark **SOTA LLMs** in real-time summarization of clinical conversations
- Developed a system that enhances real-time summarization by improving information extraction and context aggregation throughout the conversation. **Published at the AAAI 2025 workshop on GenAI4Health.**

Penn Medicine | Undergraduate Research Intern, Advisor: Prof Walter Witschey

Aug 2023 – Dec 2023

- Developed a **class-agnostic few-shot segmentation algorithm** using the **Segment Anything Model** and semantic correspondence-based point prompts to segment the liver in the MRI scans of **2000+ patients** with **71% mIoU**
- Finetuned the Total Segmentator model to segment the heart in CT scans of patients with Lipomatous Metaplasia (fat deposition), **automating the removal of over 80% of noise** and quantification of fat content

Philips Healthcare | Research Intern

Jun 2023 – Aug 2023

- Developed inference pipelines for segmentation of lung nodules on Chest CT scans and disease detection on Lung XRays
- Ensured the efficient scheduling of a large number of inference requests by **parallelizing various steps to maximally use GPUs** with Docker and PyTorch in the deployment of the models leading to up to **3x faster performance**.
- Code transferred to servers in Philips Middle East and South Asia and run on over **1.8 million CT scan images**.

Carscan.ai | Web Development Intern

May 2022 – Jul 2022

- Built a **dashboard** in React and integrated REST APIs through which the product team could quickly update the web app without having to approach any developers; **Reduced 2+ weeks turnaround time** in app updates.

## Projects

Multimodal Deepseek R1 | Advanced NLP, U. of Penn

March 2025 – April 2025

- Engineered an **optimized multi-GPU distributed training pipeline** for a multimodal reasoning LLM based on LLaVA techniques, **reducing memory footprint by over 60%** and doubling throughput during pretraining; Enabled efficient instruction tuning of R1-based vision models supporting scalable VLM training on limited hardware

Lung Cancer Detection and Segmentation | [Link](#)

Aug 2022 – May 2023

- Worked with [Prof Vinay Chamola](#), BITS Pilani, to apply **transfer learning** in PyTorch to finetune a large pre-trained CNN on a dataset of CT scans to detect & segment lung cancer nodules; Built a **feedback loop mechanism** to identify annotated CT scans and improve annotation prediction using downstream cancer detection, achieving **70% accuracy**

ERPlag Custom Compiler | [Link](#)

Feb 2023 – Apr 2023

- Created a compiler for a custom language (ERPlag) in C, with specially designed data structures & algorithms for **optimized memory consumption & reduced compile time**; Built the lexer, parser, semantic analyzer, and code generator modules and **reduced generated assembly code size by 25%** through various optimization techniques.

The Studyzone Platform | [Link](#)

Oct 2021 – Dec 2021

- Leveraged Redux and React to develop a student productivity app for **crowdsourcing academic resources**. The app has **3000+ verified documents** and **700+ GB of lecture recordings** to serve students in the university.

The Taxicab Web Dashboard | [Link](#)

Jan 2022 – Mar 2022

- React-based dashboard, integrated with REST APIs for the universities' taxicab service admins to analyze real-time ride statistics and manage rates/packages. This service has **750+ completed trips** and helps **2500+ students**.

## Technical Skills

Languages: Python (PyTorch, Tensorflow, Pandas, Numpy, Django), Javascript (Node.js, React.js, Next.js), C, C++, Java

Tools: AWS, Linux, Git, Firebase, Postman, Docker, OpenCV, MONAI

## Relevant Coursework

Graduate: Natural Language Processing, Machine Learning, Theory of Computation, Advanced NLP, Big Data Analytics

Undergraduate: Deep Learning, Data Structures and Algorithms, Image Processing, Database Systems, Operating Systems

Teaching Assistant: Data Structures and Algorithms, BITS Pilani

Aug 2022 – May 2023