

PARTH BHALERAO

Phone: (+1) 408-343-9562 ◊ Email: pbhalerao@scu.edu

Homepage: pvggeek.github.io

Google Scholar ◊ Github ◊ LinkedIn

EDUCATION

Santa Clara University (SCU)

Jun-2025 - current

PhD in Computer Science and Engineering - AI Specialization

PhD guided by - Dr. Oana Ignat

Current Research Work: Building AI Agents, Multi-modal AI Image+Video+Audio, RAG

Santa Clara University (SCU)

Sep-2023 - Jun-2025

M.S. in Computer Science and Engineering

GPA: 3.71/4.0 & Thesis: Multi Agent Image Generation System

Selected courses: Directed Research - AI & NLP, Distributed Systems, Algorithms

Ramdeobaba Univeristy India

Aug-2019 - May-2023

B.E. in Electronics and Computer Science

GPA: 9.4/10.0

Selected courses: Artificial Intelligence, Machine Learning, Data Analysis, Software Engineering

RESEARCH INTERESTS

I am interested in Agentic AI and Multimodal AI (text, image, video, audio), with a focus on building intelligent agents and scalable systems. I also contribute to creating novel datasets for low-resource languages, enabling inclusivity and broader accessibility in AI research. My work includes optimizing large AI models on multi-GPU setups and distributed infrastructures, leveraging sequential and parallel computation strategies for efficiency and real-world deployment.

RESEARCH EXPERIENCE

Mentorship4All: Multi-Agent QA Extraction for Long-Form Mentorship Videos

June 2025 – current - Under Experimentation and Review (soon to be released)

- Proposed a multi-agent framework for QA from long-form mentorship and educational videos.
- Developed a novel chunking algorithm, benchmarking single vs. multi-agent performance.
- Completed RAG comparisons, confirming multi-agent superiority.
- Found significant QA metric gains in faithfulness, relevance, and coherence across three languages.
- Finished by showcasing scalable multilingual capabilities for accessible learning.

MoSAIG - Multi-Agent Multimodal Models for Multicultural Text to Image Generation

October 2024 - May 2025 - Under Review ArXiv Page

- Proposed MosAIG, a Multi-AI-Agent framework for multicultural image generation.
- Found multi-agent models outperform simple baselines.
- Completed and open-sourced a 9,000-image multicultural dataset and the multi-agent pipeline.
- Modified workflows with agent-based captioning.
- Finished with steps for fairness and multilingual improvements.

Performance Analysis of YOLOv5 for ASL Detection

November 2022 – Apr 2023 – SSRN Page

- Proposed YOLOv5 evaluation for ASL detection using PyTorch, TensorFlow, and multi-GPU training.
- Found consistent misclassifications in specific letters across devices and frameworks.
- Completed 4,500 automated experiments across Intel CPUs, Raspberry Pi, and Jetson Nano GPUs.

- Modified workflows with automation scripts and parameter sweeps (image size, weights, thresholds).
- Finished by open-sourcing all model weights, code for community use, and suggested the most optimal and best framework setup across various platforms.

PUBLICATIONS

ECG Classification Using Machine Learning on Wave Samples for the Indian Population

Bhalerao P, Essaji H, Korde M. — IEEE InCACCT, 2023 — PDF

Design of a Dynamic Traffic Signal System with IoT and Digital Circuit Integration

Bhalerao P, Thakre P, Dongre A — IEEE ICCCNT - Top Conference, 2023 — PDF

Point of Care Device for Measurement of Vital Parameters

Bhalerao P, Korde M — Springer SmartCom International Conference & Patent granted, 2023 — PDF

WORK EXPERIENCE

Research Assistant — AIM Lab

Santa Clara University, Santa Clara, CA — June 2025 – Present

- Tech: Python, PyTorch, CUDA, multi-GPU optimization, multimodal datasets, AI agent frameworks.
- Developing AI agents for advanced multimodal research, integrating image, text, and video understanding tasks.
- Designed and implemented chunking algorithms to efficiently preprocess and manage large multimodal datasets.
- Researching and prototyping optimized deployment strategies for sequential and parallel AI agents across multi-GPU systems, improving scalability and throughput.

Software Automation Developer

Santa Clara University, Santa Clara, CA — Jan 2024 – Aug 2025 | Part-Time

- Tech: Workday, Python, ELK Stack (Elasticsearch, Logstash, Kibana), JavaScript/HTML/CSS
- Built automation scripts for Workday client processes, streamlining student enrollment and salary calculations and reducing administrative effort.
- Implemented structured logging with the ELK Stack, introducing correlation IDs and custom middleware for request tracking, improving end-to-end traceability and cutting issue resolution time by 60%.
- Developed and maintained frontend components for SCU websites, enhancing UI/UX design and creating new pages for academic and administrative use.

Research Assistant — HASO Labs

Santa Clara University, Santa Clara, CA — Sep 2023 – Dec 2023

- Tech: Python, CLIP, Mediapipe, CUDA, AWS (Lambda, SQS, API Gateway), GPU optimization.
- Designed and implemented a GPU-optimized video processing pipeline using Python, CLIP, and Mediapipe for a 300GB+ dataset.
- Parallelized frame-level operations, reducing vector embedding generation time from 50+ hours to ~21 hours.
- Deployed trained models on AWS cloud, integrating SQS queues, Lambda, and API Gateway for scalable query handling.
- Delivered an end-to-end system optimized for Meta VR headset integration, improving latency and reliability of video-based research tasks.

Machine Learning Intern

Innovative Technologies, New Delhi — June 2022 – Nov 2022 | Internship

- Tech: Python, scikit-learn, NumPy/Pandas, signal processing, XML automation, deployment.
- Researched and trained ML models for 3-lead ECG rhythm prediction, achieving ~93% accuracy with medically graded sensor integration.
- Contributed to a novel 3-lead ECG dataset, including collection standards and labeling guidelines.
- Built XML automation pipelines for data extraction and preprocessing to accelerate experimentation.

Systems Programmer Intern

ECDS, Nagpur, MH, India — Dec 2021 – Apr 2022 | Internship

- Tech: C++, custom libraries, IoT, system performance optimization.
- Developed C++ libraries for system software, optimizing hardware–software interaction.
- Reduced IoT transfer latency from 5–7s to milliseconds, drastically improving end-to-end system performance.

SKILLS

Programming	Python, C/C++, CUDA Programming
Frameworks & Libraries	PyTorch, TensorFlow, scikit-learn, HuggingFace, LangChain, LlamaIndex, CrewAI
Databases	Vector Databases & RAG pipelines, SQL
Data Science Tools	NumPy, Pandas