

ARATI GANESH

☎ +1(404)2039552 ✉ aganesh83@gatech.edu [in arati-ganesh](#) [🔗 Portfolio](#) 🏠 Atlanta, GA

EDUCATION

Georgia Institute of Technology, Atlanta

Aug 2023 - May 2025

Master of Science, Electrical and Computer Engineering

GPA - 4.0/4.0

Advanced Programming Techniques, Generative and Geometric DL, Random Processes, Machine Learning, Hardware and Software Codesign for ML. Researcher under [Prof. Alexey Tumanov](#) and [Prof. May Wang](#)

BMS College Of Engineering, Bengaluru

Aug 2017 - July 2021

Bachelor of Engineering in Electrical and Electronics

GPA - 9.57/10

WORK EXPERIENCE

AI Accelerated HPC CPU Intern

May 2024 - Present

AMD, Austin

- Optimizing TPCxAI Benchmark, used to evaluate AI/ML workloads, for multicore processing on smaller scale factors (SF1 - SF30), focusing on reducing the 3-hour benchmark execution time.
- Implemented extensive parallelization for data loading and preprocessing across 10 ML models, using custom OpenMP framework and Apache Arrow, achieving 2x-3x speedup on AMD EPYC Server CPUs.

Graduate Teaching Assistant - OMSCS 7643, PHYS 2212

Aug 2023 - May 2024

Georgia Institute of Technology, Atlanta

- CS 7643: Assisted students with Deep Learning concepts, graded assignments, and provided detailed feedback.
- PHYS 2212: Led weekly group problem-solving sessions and addressed student doubts in modern physics.

Machine Learning Engineer

Jan 2022 - July 2023

Sony India Software Centre, Bengaluru

- Optimized object detection and image matting models, improving accuracy by 10-15% via hyperparameter tuning. Implemented CUDA algorithms for 2x speed and accuracy improvements.
- Developed C++ ALPR application for iMX8 edge device, combining iMX500 AI camera inference with on-device processing. Optimized with TensorFlow Lite and ONNX for efficient inference.
- Accelerated model development cycles by 20% through a custom benchmarking application that extracted KPIs and enabled model caching. Integrated MLflow to streamline the ML workflow, reducing deployment time in production
- Engineered a scalable cloud application utilizing AWS Lambda, S3, Batch, and EC2, achieving a 25% improvement in batch image processing throughput.

Engineer

Aug 2021 - July 2023

Ignitarium Technology Solutions, Bengaluru

- Benchmarked early vision algorithms and SOTA Deep Learning models for object pose estimation in retail robots.
- Engineered custom CUDA kernels, achieving a 40% efficiency boost in image processing pipelines, validated and fine-tuned using Nsight Systems profiler.

Embedded Engineering Intern

Feb 2021 - July 2021

Honeywell Technology Solutions, Bengaluru

- Engineered firmware for ADuCM355-based Single Gas Detectors, enhancing device performance. Optimized design through simulations using Mentor Graphics tool.

Robotics Intern

Jul 2020 - Dec 2020

Robert Bosch Center for Cyber Physical Systems, Indian Institute of Science, Bengaluru

- Developed an IMU-based hand pose estimator for tele-robotic control after analyzing various methods. Simulated a ROS tele-operated pick-and-place robot, enhancing control and manipulation. [\[Report\]](#)

TECHNICAL SKILLS

Languages: Python, C, C++, CUDA, Java, Shell Scripting, MATLAB, OpenMP, OpenMPI, OpenGL

Frameworks/Tools: PyTorch, TensorFlow, Scikit-Learn, Numpy, Matplotlib, ROS, AWS Cloud, Docker, MLFlow, Git

Publications: Mobile Covid Sanitization Robot [\[Link\]](#)

PROJECTS

Staged Aggregation for Efficient Inference in Graph Neural Networks - Systems for AI Lab | *Deep Graph Library*

- Implemented layer-wise inference in GNNs, addressing neighborhood explosion and enhancing scalability. Developed degree-balanced data loading for uniform batches, significantly improving memory efficiency and processing speed. [\[Slides\]](#)

Variational Autoencoders for Collaborative Filtering | *Pytorch, Data Preprocessing & Analysis*

- Engineered a VAE-based recommendation system in PyTorch, integrating composite prior and beta annealing techniques to enhance collaborative filtering performance. [\[Paper\]](#)

Efficient LLM Finetuning | *PyTorch, Transformers*

- Implemented Context Distillation (CD) using GPT-4 as teacher and OPT 125M as student. Compared CD + PEFT (LoRA, QLoRA) with traditional fine-tuning methods, evaluating out-of-domain accuracy and GPU Utilization. [\[Code\]](#) [\[Paper\]](#)

Beyond FLuID: Exploring Sparsity and Algorithm Variability in Federated Learning | *Flower, PyTorch*

- Extended FLuID to further mitigate straggler effects in Federated Learning. Explored FedAdam and FedAvg, evaluating sparsity thresholds and model quantization to enhance convergence and efficiency. [\[Code\]](#)[\[Paper\]](#)

Enhancing Pulmonary Embolism Diagnosis and Treatment | *Scikit-Learn, PyTorch*

- Implemented and evaluated ML models (SVM, CatBoost, K-Means, DBSCAN) for pulmonary embolism prediction and diagnosis using EHR data. Assessed performance with accuracy, F1 score, AUROC, and silhouette scores. [\[Code\]](#) [\[Website\]](#)