Arjun Parthasarathy

905 Mockingbird Ln, Palo Alto, CA 94306
650-796-0394 | arjun.p@columbia.edu | https://github.com/ArjunParthasarathy

EDUCATION

Columbia University **New York, NY**
B.A Computer Science; Applied Math Expected May 2027

- **GPA:** 4.03 / 4.0
- **Honors:** I.I Rabi Scholar
- **Relevant Coursework:** Computer Science Theory, Fundamentals of Computer Systems, Advanced Programming, Linear Algebra, Differential Equations, Discrete Math, Data Structures and Algorithms

WORK & LEADERSHIP EXPERIENCE

HazyResearch, Stanford AI Lab, Stanford University **Stanford, CA**
Research Assistant Jul 2024 – Present

- Co-author of paper in submission to International Conference on Learning Representations
  - ThunderKittens: Implementing CUDA kernels for latest NVIDIA GPUs
- Designed gated-convolution kernels used in upcoming state-of-the-art DNA foundation model

Center for Theoretical Neuroscience, Zuckerman Institute, Columbia University **New York, NY**
Research Assistant Nov 2023 – Present

- Meta-learning synaptic plasticity ML model to model odor pathways in piriform cortex
- Previous work: connectivity between mossy fiber (MF) and granule cell (GrC) layers in the cerebellum
- ML and graph theory skills in PyTorch, NumPy, SciPy, and NetworkX

Autonomous Networks Research Group, USC **Los Angeles, CA**
Research Assistant Apr 2021 – Jun 2023

- First author of three research papers on the optimal partitioning, placement, and robust deployment of deep neural networks across clusters of low-resource edge devices
- Best Paper Award at COMSNETS MINDS Workshop
- Reached communication latency within 9% of theoretically optimal using custom graph algorithms
- Designed custom Kubernetes cluster with multi-node fault tolerance, tested on up to 50 nodes and devices w/ 128 to 512 MB RAM

SELECTED PUBLICATIONS

SEIFER: Scalable Edge Inference for Deep Neural Networks
2022 Workshop on Challenges in Deploy and Monitoring ML Systems, Conference on Neural Information Processing Systems (NeurIPS)

Partitioning and Placement of Deep Neural Networks on Distributed Edge Devices to Maximize Inference Throughput
2022 32nd International Telecommunication Networks and Applications Conference (ITNAC)

DEFER: Distributed Edge Inference for Deep Neural Networks
2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS)

SKILLS, ACTIVITIES & INTERESTS

Languages: Fluent in English; Conversational Proficiency in Spanish, Tamil
Technical Skills: Python, C/C++, CUDA, Java, Golang, SQL, HTML/CSS/JS
Certifications & Training: DeepLearning.ai NLP Specialization, Java Spring Boot Certification
Activities: New York Classical Guitar Orchestra
Interests: Classical/Flamenco Guitar, Formula 1, Cycling