

# Vikranth Srivatsa

vsrivatsa@berkeley.edu ◊ github.com/vikranth22446

## EDUCATION

---

**University of California San Diego - Ph.D.** Oct 2023 - May 2027  
Computer Science: Systems and Machine Learning

**University of California Berkeley - Bachelor & Masters of Science** Aug 2018 - May 2023  
Electrical Engineering and Computer Science

### Skills

Languages: Python, Golang, C++, Java, C, Scala, Rust, Verilog, Risc-V/x86, CUDA  
Other: Pytorch, Tensorflow, Natural Language Processing, Keras, OpenCV

## WORK EXPERIENCE

---

- **Augmented Language Models - WukLab UC San Diego** Sept 2023 - Present
  - Built a inference system that improves latency/throughput of language models that call APIs
  - Built load testing tool that profiles different augmented language models vLLM, OpenAi, ToolFormer
- **NASA AMES and SkyLab** June 2021 - July 2023
  - Created a serverless scheduling framework across client, 5G networks, & cloud (AWS, GCP, Azure, K8s)
  - Wrote heuristic and linear programming algos to save 70% cost reduction and 30% performance increase
  - Profiled and optimized throughput, implemented write ahead logging and fault tolerance
- **New Relic Pixie Labs** *Systems Software Engineering Intern* June 2022 - May 2022
  - Designed a new system that automatically parses system level protocols used in Kubernetes in C++
  - Architected RabbitMQ and Kafka monitoring and performance analysis tooling in Kubernetes
  - Worked on kernel level memory profiling and flame graph creation via kernel monitoring protocol eBPF
- **UC Berkeley RISELab** *ML Researcher - Published at NeurIPS Workshop* Feb 2021 - Present
  - Investigating machine learning explainability with large models and distribution shift
  - Designed and ran research experiments to test model size, subgroup robustness and counts, and pretraining
  - Automatically constructed spurious correlation large scale image datasets with Imagenet
- **Amazon Alexa** *Software Development Engineer Intern* May 2020 - August 2020
  - Designed and built dev tool for experiment tracking, managing A/B model testing, and dashboards
  - Built parameter tracking to process models used in ranking content in the Alexa Mobile App
- **Amazon AWS** *Software Development Engineer Intern* May 2019 - August 2019
  - Implemented core open source features on AWS SAM-CLI used in 13% of production workflows
  - Designed cloudformation and cloudformation intrinsic templating resolution
  - Built a Serverless Testing Framework to benchmark and test AWS Serverless infrastructure
- **Circles - Startup** *Chief Technology Officer* September 2018 - May 2019
  - Lead a team of 10 engineers to build out a curated news platform
  - Created an efficient news recommendation system using feed ranking algorithms and collaborative filtering

## PROJECTS

---

- **Passive Listening Smart Speakers - Berkeley Blues Security Lab** Feb 2020 - Present
  - Created a system to study privacy implications of always listening smart assistants via machine learning
  - Built machine learning NLP models to classify different intentions for passive listening and speech
  - Wrote noise filtering & detection algorithms in order to improve and detect intents
  - Published at UC Berkeley Center for Longterm Cybersecurity
- **Mixclaves - UC Berkeley Sky Lab** *Paper in Review* Fall 2022 - December 2022
  - Built a system to anonymize private metadata in a network vis mixnets and enclaves
  - Built a live demo with a realtime network metrics, auto scaling, and anonymous administration

## RESEARCH

---

- **Augmented Language Models - WukLab UC San Diego**  
*ICML 2024* Present
- **Effect of Model Size on Worst-group Generalization**  
*Published at Neurips Workshop* 2021
- **Cloudless Serverless across multi cloud**  
*Research in Progress SOSP 2024* 2023
- **Cloudless Serverless across multi cloud and Mixclaves: secure private messaging**  
*Published at UC Berkeley Masters Thesis* 2023
- **Passive Listening Smart Speakers - Berkeley Security Lab**  
*Published at Center of Long Term Cybersecurity* 2022