# LEI GAO

213-255-8040 ◇ leig@usc.edu

## EDUCATION

**University of Southern California** — August 2022 - Present
Ph.D. in Electrical Engineering — *Los Angeles, California, USA*
Advisor: Dr. Murali Annavaram

**University of Southern California** — January 2020 - May 2021
M.S. in Electrical Engineering — *Los Angeles, California, USA*

**University of California, Santa Barbara** — September 2015 - March 2019
B.S. in Electrical Engineering — *Santa Barbara, California, USA*

## TECHNICAL SKILLS

| | |
|---|---|
| **Programming Languages** | C/C++, Python, CUDA, Bash |
| **Python Packages** | PyTorch, TensorFlow, TFLite |
| **Software & Tools** | CMake, Docker, TensorRT |
| **Hardware Description Languages** | Verilog, SystemVerilog |

## RESEARCH INTERESTS

**Resource-efficient Machine Learning Systems**
**Federated Learning**
**Large Language Models**

## WORK EXPERIENCE

**FedML Inc** — May 2022 - July 2022
*Applied Researcher Internship* — *Los Angeles, California, USA*

- Developed an Android on-device training engine using C++ for the FedML library, integrating MNN and PyTorch, to facilitate advanced machine learning directly on mobile devices.
- Conducted comprehensive federated learning experiments across a diverse range of Android smartphones, utilizing various models and datasets to test effectiveness and scalability.

**OmniVision Technologies Inc** — July 2021 - April 2022
*SoC Design Engineer* — *Santa Clara, California, USA*

- Engineered DVP and MIPI receivers for CMOS image sensors, ensuring compliance with MIPI CSI-2 and DPHY specifications, and enhancing signal processing capabilities.
- Participated in the complete ASIC design flow for image sensors, encompassing Verilog RTL coding, simulation, synthesis, timing closure, static timing analysis, and formal verification.
- Contributed to the development of cutting-edge image sensor technology, meeting the demands of vehicular applications.

**USC ISI Application Specific Intelligent Computing Lab** — January 2021 - April 2021
*Research Assistant Internship* — *Los Angeles, California, USA*

- Analyzed and optimized energy consumption and memory requirements for CNN hardware accelerators.
- Executed Monte Carlo simulations in HSpice to model magnetic tunnel junction circuits in MRAM and SRAM arrays.
- Evaluated and compared the power, performance, and area of wafer-scale on-chip memory architectures based on magnetic and optical devices against traditional SRAM designs.

**Chipltech Technologies Ltd** — September 2019 - December 2019
*ML ASIC Software Engineer Internship* — *Shenzhen, Guangzhou, China*

· Implemented and benchmarked the inference and training processes for ResNet50 and XGBoost algorithms using a cycle-accurate instruction set simulator, providing crucial insights into ML accelerator architecture performance.
· Enhanced the simulation performance of ML accelerator architectures by maximizing utilization of the systolic array unit and innovated a patented method that accelerates batch normalization layer computations.

## PUBLICATIONS

- **Lei Gao***, Yue Niu*, Tingting Tang, Salman Avestimehr, Murali Annavaram "Ethos: Rectifying Language Models in Orthogonal Parameter Space" NAACL Findings 2024.
- **Lei Gao***, Yue Niu*, Tingting Tang, Salman Avestimehr, Murali Annavaram "Ethos: Rectifying Language Models in Orthogonal Parameter Space" AAAI 2024 Responsible Language Models Workshop (spotlight).
- Tuo Zhang, **Lei Gao**, Sunwoo Lee, Mi Zhang, Chaoyang He, Salman Avestimehr "TimelyFL: Heterogeneity aware Asynchronous Federated Learning with Adaptive Partial Training" CVPR 2023 Federated Learning for CV workshop.
- Tuo Zhang, **Lei Gao**, Chaoyang He, Mi Zhang, Bhaskar Krishnamachari, Salman Avestimehr "Federated Learning for Internet of Things: Applications, Challenges, and Opportunities" IEEE Internet of Things Magazine.
- Tuo Zhang, Chaoyang He, Tianhao Ma, **Lei Gao**, Mark Ma, Salman Avestimehr "Federated Learning for Internet of Things: A Federated Learning Framework for On-device Anomaly Data Detection" ACM SenSys 2021 AIChallengeIoT Workshop.
- Feng Yun, Yunkun Lin, Lou Yunfei, **Lei Gao**, Vaibhav Gera, Boxuan Li, Vennela Chowdary Nekkanti, Aditya Rajendra Pharande, Kunal Sheth, Meghana Thommondru, Guizhong Ye, Sandeep Gupta "Fault-coverage Maximizing March Tests for Memory Testing" 2022 IEEE International Test Conference.

## HONORS AND AWARDS

**Ph.D. Annenberg Fellowship, USC** August 2022

**College of Engineering Honors Program Graduation Scholar, UCSB** June 2019

## TEACHING EXPERIENCE

**EE 599: Machine Learning Systems** Fall 2023, Spring 2024
*Teaching Assistant* *University of Southern California*

## PROFESSIONAL SERVICE

**Reviewer: IEEE Computer Architecture Letters** September 2023

**Reviewer: ACL Rolling Review** January 2024

## PATENT

- Xiaoming Chuang, Yifan YangGong, Hanxun Zheng, **Lei Gao**, Juzhe Zhong "Data processing method, device, chip and computer readable storage medium" China Patent CN111814983A, October 23, 2020.