

EDUCATION

University of California Los Angeles

Masters of Science— Electrical and Computer Engineering— GPA: 4.0/4.0

Los Angeles, CA

Sep 2024 – June 2026

University of California Los Angeles

Bachelors of Science, Electrical Engineering, GPA: 3.8/4.0

Los Angeles, CA

Sep. 2021 – June 2024

RESEARCH EXPERIENCE

Graduate/Undergraduate Researcher

April 2023 – Present

Advised By Professor Lin F. Yang

UCLA

- Conducted research on efficient large-language-model inference through compression and length-aware reasoning
- Proposed NoWag, a unified normalization and layer-wise proxy loss that enables data- and compute-efficient pruning and vector quantization of large language models; accepted as a first-author paper at the Conference on Language Modeling (COLM 2025)
- Showed that the NoWag-VQ 2-bit vector-quantization method outperforms state-of-the-art one-shot 2-bit vector quantization baselines across a range of open-weight LLMs while using $48\times$ less calibration data
- Developed ARMOR, a semi-structured pruning method that factorizes LLM weight matrices into a hardware-efficient 2:4 sparse core with adaptive block-diagonal error-correcting wrappers, optimized via a block-coordinate algorithm with theoretical convergence guarantees
- Demonstrated that ARMOR-pruned LLaMA and Qwen models retain up to 50% more performance than models pruned with existing 2:4 sparsity methods, while maintaining inference speedups and memory savings

Undergraduate Researcher

February 2022 – October 2024

Advised By Professor Vwani Roychowdhury

UCLA

- Formulated self-supervised VAE framework that automatically clusters epileptic brain signal morphologies in latent space
- Identified a new class of pathological brain signals in the latent space that are more correlated with surgical outcomes and easier (thus cheaper) to detect than established biomarkers; validated on 686K events from 185 patients.
- Developed PyHFO, an end-to-end clinical software platform integrating optimized signal detection algorithms with lightweight CNN classifiers, achieving $50\times$ speedup over traditional methods
- Led a multidisciplinary team to build PyHFO, a Python GUI tool for real-time brain signal recording and biomarker detection, now actively used by pediatric neurologists at UCLA Mattel Children's Hospital
- Results published in *Journal of Neural Engineering* and *Epilepsia*

TEACHING EXPERIENCE

Teaching Assistant— Deep Learning 2 (ECE 239AS)

UCLA Electrical and Computer Engineering Department

April 2024 – June 2024

Los Angeles, CA

- Volunteered as TA during the first offering of a graduate-level Deep Learning course as an undergraduate, leading weekly office hours and reinforcement-learning-focused discussion sections for 200+ graduate students
- Designed two reinforcement learning programming assignments (DDPG for continuous control, DQN for discrete actions) adopted for course curriculum, enabling hands-on implementation experience with modern RL algorithms

PUBLICATIONS

Journal/Conference Papers

- **Lawrence Liu**, Inesh Chakrabarti, Yixiao Li, Mengdi Wang, Tuo Zhao, Lin F. Yang. “[NoWag: A Unified Framework for Shape Preserving Compression of Large Language Models](#)”, In *Second Conference on Language Modeling (COLM)*, 2025
- Yipeng Zhang, Atsuro Daida, **Lawrence Liu**, et al. “[Self-Supervised Data-Driven Approach Defines Pathological High-Frequency Oscillations in Epilepsy](#)”, *Epilepsia*, 2025 [Accepted]
- Yipeng Zhang, **Lawrence Liu**, Yuanyi Ding, et al. “[PyHFO: Lightweight Deep Learning-powered End-to-End High-Frequency Oscillations Analysis Application](#)”, In *Journal of Neural Engineering* , 21, 2024.

Preprints/Under Review

- **Lawrence Liu**, Alexander Liu, Mengdi Wang, Tuo Zhao, Lin F. Yang, “[ARMOR: High-Performance Semi-Structured Pruning via Adaptive Matrix Factorization](#)”, 2025.
- Chang Liu, Yiran Zhao, Yaoqi Ye, **Lawrence Liu**, Csaba Szepesvari, Lin F. Yang, “[LACONIC: Length-Aware Constrained Reinforcement Learning for LLM](#)”, 2025
- Tonmoy Monsoor, Atsuro Daida, Prateik Sinha, Yipeng Zhang, **Lawrence Liu**, et al. , “[Mini-Seizures: Novel Interictal iEEG Biomarker Capturing Synchronization Network Dynamics at the Epileptogenic Zone](#) ”, 2025.

Workshop Papers

- Preliminary Results of the NoWag framework presented at **ICLR 2025 Sparsity in LLMs Workshop**

AWARDS, HONORS, & SERVICES

Dean's List — *UCLA Henry Samueli School of Engineering* — Fall 2021, Fall 2022, Winter 2023, Spring 2023, Spring 2024

Eta Kappa Nu (HKN) Honor Society — *UCLA* — Fall 2023

Volunteering - Reviewer: *ICLR 2026*, Mentor: *UCLA ACM AI AISF program*

TECHNICAL SKILLS

Languages: Python, C/C++, R, Matlab, Octave, Julia, Mathematica

Frameworks & Tools: Git, Bash Scripting, Pytorch, Sklearn, Tensorflow, Pandas, NumPy, Matplotlib, Simulink, Gymnasium, Optuna, Cvxp, Mosek, Cuda, Transformers

Specialized Techniques:

- **Machine Learning:** Deep Learning, Reinforcement Learning, Graph Neural Networks, Generative AI, Model Parallelism
- **Signal Processing:** Time Series Analysis, Statistical Analysis, Data Mining, Data Visualization
- **Optimization:** Numerical Analysis, Convex Optimization
- **Systems:** Control Theory, High-Performance Computing