Eric Liu

Email: eliu4913@usc.edu · Cell: (650) 865-8068 · eliu4913.github.io/

EDUCATION

University of Southern California, Viterbi School of Engineering – Los Angeles, CA   Aug. 2020 – Dec. 2025
B.S. in **Computer Science & Business Administration** | **GPA 3.92/4.0**
M.S. in **Computer Science (Progressive Degree Program)** | **GPA 4.0/4.0**

- Courses: Optimization for Machine Learning, Deep Learning and its Applications, Machine Learning, Applied NLP, Computer Systems, Professional C++, Analysis of Algorithms
- Honor: Dean's Honor List in all semesters.

RESEARCH EXPERIENCE

Research Assistant   May 2025 – Present
University of Southern California

- Worked with Professor Sai Praneeth Karimireddy and PhD students Yavuz Faruk Bakman and Duygu Nur Yaldiz on researching LLM unlearning and evaluation.
- Conducting in-depth literature review on existing methods in the field of LLM unlearning.
- Implementing adversarial attacks and prompting on existing models to test the likelihood of sensitive information leakage.

Research Assistant   March 2025 – Present
University of California, San Diego, PICASSO Lab

- Worked with Prof. Yufei Ding and PICASSO on ML-systems research, focusing on context parallelism for long-context LLM training.
- Implemented SOTA baselines (Per-Sequence and Per-Document context parallelism), and co-developed core components of *FlashCP*.
- Ran large-scale experiments on 8×H100 GPUs and AWS clusters; benchmarked models up to 70B parameters across three public datasets, showing avg. 1.45× end-to-end speed-ups, and up to 1.55× vs. the baseline in attention computation and training throughput.
- Co-authored and submitted the resulting manuscript to NeurIPS 2025 (under review).

Research Assistant   March 2025 – Present
University of California, San Diego, SimWorld Team

- Worked with Prof. Lianhui Qin and the SimWorld Team on *SimWorld*, an Unreal Engine 5 simulator that generates large-scale, open-ended urban environments for embodied-AI research.
- Contributed to the paper's second experiment, a LLM-driven delivery-agent benchmark, by designing task variants, implementing configuration/launch scripts, implementing baseline and ablation setups and debugging the experiment pipeline.
- Summarized the performance findings into plots and tables for the NeurIPS 2025 submission (under review).

Research Assistant   May 2024 – Present
University of California, Los Angeles

- Worked with Prof. Lin Yang and PhD student Yunfan Li on *Trajectory Graph Learning*, a method to directly learn policies from preference-ranked trajectories data.
- Performed an in-depth literature review to guide discussions on experiment methodology and paper positioning: research questions, project framework and paper storyline.
- Implemented the full experimental pipeline: set up environments, reproduced baseline algorithms, and implemented TGL variants.
- Co-authored and submitted the resulting manuscript to NeurIPS 2025 (under review).

TEACHING EXPERIENCE

Course Producer (CSCI270 Introduction to Algorithms and Theory of Computing)   Jan. 2024 – Dec. 2024
University of Southern California, Viterbi School of Engineering

- Conducted weekly office hours and exam review to assist 150+ students with concepts such as Greedy Algorithms, Divide and Conquer, Dynamic Programming, NP-Complete, and Network Flow.
- Proctored exams, graded assignments, and answered questions from students on discussion forums.

ACADEMIC PROJECTS

ZeRO-Offload Stages and Model Size Impact   Jan. 2025 – Apr. 2025

- Reproduced the implementation of ZeRO-Offload stages to transfer optimizer states, gradients, and model parameters from GPU VRAM to CPU memory.
- Conducted experiments on different ZeRO-Offload strategies and evaluated metrics such as peak and average memory use, final accuracy, workload tradeoffs (data transfer vs. forward/backward computations), and total throughput.
- Identified optimal strategies to balance training efficiency and scalability for large-scale transformer models.

Impact of Object Detection Methods for Game-Playing Agents   Aug. 2024 – Dec. 2024

- Developed a game-playing agent that leverages advanced object detection pipelines (e.g., YOLOv8, detection transformers) to inform decision-making, which is then used to determine the tradeoffs between detection accuracy with real-time inference performance.
- Designed and executed experiments that evaluated precision, recall, mAP, and latency, correlating these metrics with in-game performance to quantify trade-offs in detection methods.

Virtual Machine and Compiler (C++, Bison, x86 Assembly)   Oct. 2023 – Dec. 2023

- Created a Virtual Machine in C++ with fifteen 32-bit integer registers, 1 KB of stack space, and 3-bit color graphics that support turtle graphics by using template metaprograms, exceptions, and smart pointers.
- Designed a compiler that enabled the generation of assembly instructions from an Abstract Syntax Tree that streamlined code compilation and increased efficiency of 15% compared to open-source solutions.
- Optimized register usage and improved execution efficiency of operations of VM by implementing Linear Scan Register Allocation algorithm to map virtual registers to real registers.

SKILLS

Research Interests: ML Systems, efficient and trustworthy LLMs, large-scale training, RL, and applications of deep learning.

Languages and Frameworks: Python (PyTorch, TensorFlow), C++, C, and Unreal Engine.

Bilingual in Mandarin and English.