Lawrence Feng

925-819-8690 lawrence.feng2017@gmail.com lawrencefeng17.github.io

Education

Carnegie Mellon University **Fall 2022 – Spring 2026**

Bachelor of Science in Statistics and Machine Learning *Pittsburgh, Pennsylvania*

- Additional major in Artificial Intelligence and minor in Mathematics
- GPA: 3.90

Relevant Coursework

- Introduction to Deep Learning (PhD)
- Deep Reinforcement Learning (PhD)
- Probabilistic Graphical Models (PhD)
- Probability and Mathematical Statistics (PhD)
- Algorithm Design and Analysis
- Parallel and Sequential Data Structures
- Advanced Methods for Data Analysis
- Monte Carlo Methods and Applications

Experience

Inflection AI **Summer 2025**

Member of Technical Staff Intern

- Co-developed post-training tools for the **open-source** torchtune library, including support for distributed Qwen2.5-VL fine-tuning (PR #2868); required deep understanding of multimodal RoPE and fusion architectures
- Proposed and built benchmark refinement tool leveraging Item Response Theory to transform statistically insignificant LLM-as-a-judge evaluations into reliable discriminative assessments, now part of the company's core evaluation suite
- Core contributor to a company-wide effort on EQ evaluation; surveyed prior work and collaborated on designing a suite of benchmarks, prototyped persona-based LLM simulators, and built synthetic data generation pipelines

Carnegie Mellon Computer Science Department **Fall 2024 – Summer 2025**

Research assistant advised by Professor Wenting Zheng

- Developed novel watermark embedding algorithm which makes cryptographic watermarking practical
- Sole programmer managing all aspects: experimental design, implementation, analysis, and progress reporting
- Synthesizing complex experimental results into clear visualizations and accessible summaries for research collaboration

Robotics Institute at Carnegie Mellon **Spring 2024 – Fall 2024**

SURA Summer Research Intern

- Explored adapting CoTracker's retrospective dense point tracking into predictive models that output future robot states
- Built automated MuJoCo pipeline that dynamically generates environments and processes incoming model checkpoints
- Uncovered significant robustness failures in state-of-the-art models through systematic environmental variations

Carnegie AI Safety Initiative **Fall 2023 – Present**

President (since Fall 2025), Member

- Leading a 200+ member community promoting AI safety; quadrupled fall recruitment through bold campus initiatives, expanding leadership team, and continuing to lead the introductory AI alignment fundamentals course

Projects

Honors Undergraduate Research Thesis *advised by Professor Aditi Raghunathan* **Fall 2025 – Present**

- Investigating training dynamics and catastrophic forgetting in large language models, focusing on how knowledge is acquired, retained, and lost over the course of pretraining and fine-tuning.

Interpreting Vision Language Models | *Python, Hugging Face, TransformerLens* **Fall 2024**

- Led a project investigating vision language models using tools like Hugging Face and TransformerLens.
- Found that a language-only sparse autoencoder can provide insights into a multimodal model's black-box behavior
- Demonstrated the ability to alter model behavior predictably by intervening on intermediate activations

Solving Jigsaw Puzzles using Reinforcement Learning | *Python, PyTorch, Pandas, Git* **Fall 2023**

- Led a team of four in designing, implementing, and testing a deep reinforcement learning system inspired by AlphaGo
- Achieved $> 90\%$ on image reassembly task through innovative model architecture and tree search
- Managed project timeline, code integration, and technical direction while coordinating team efforts

Technical Skills

Programming Languages: Python, R, C, C++, Java, SQL, SML, LaTeX, Bash
Developer Tools: GCP, AWS EC2, Git, WandB, Slurm, Conda, uv
Frameworks and Libraries: PyTorch, TensorFlow, Hugging Face, vLLM, OpenCV, NumPy, Pandas, Matplotlib, ROS