Michael Li

bearseascape@gmail.com ▪ ml5885.github.io ▪ github.com/ml5885

EDUCATION

Carnegie Mellon University, *B.S. Statistics & Machine Learning* August 2022 - May 2026
Dean's List, High Honors (All Semesters)
Relevant Coursework Machine Learning, Natural Language Processing, Deep Learning Systems, Parallel and Sequential Data Structures, Introduction to Computer Systems, Linear Algebra, Advanced Methods for Data Analysis, Causal Inference, Statistical Computing

SKILLS

Languages: Python, C/C++, Java, SML, C#, JavaScript, TypeScript, Go, SQL, R, HTML/CSS
Frameworks & Libraries: PyTorch, JAX, TensorFlow, Hugging Face, React, Flask, Django, Node.js, PostgreSQL, OpenGL
DevOps/MLOps: Git, Docker, Kubernetes, Apache Kafka, Prometheus, Google Cloud Platform, AWS

WORK

Software Engineering Intern, Google Sunnyvale, CA • May 2025 – Aug 2025

- Built preprocessing pipelines to extract network data via **GraphQL** and store features in **Cloud Spanner** for downstream ML workflows.
- Designed and trained a **GNN autoencoder** (**Python/JAX**) with masked feature prediction on Google's B4 network, producing high quality embeddings for network state representation.
- Deployed embeddings for **anomaly detection** via cosine distance & z-score, achieving 92% precision in surfacing network anomalies.
- Presented findings within Google Global Networking, informing roadmap for the Autonomous Network Operations project.

Machine Learning Intern, Epirus Los Angeles, CA • June 2024 – August 2024

- Researched and developed autonomous targeting policies for countering heterogeneous, delayed-feedback drone swarms using **deep reinforcement learning** (**PPO**, **PointNet**, **PyTorch**), improving simulated mission success rates over baseline heuristics.
- Designed a fully vectorized RL environment and multithreaded training pipeline, reducing training time from days to minutes.
- Built a 2D/3D drone-swarm simulator with **PyGame** and **OpenGL** to generate diverse training scenarios and evaluate learned policies.
- Demonstrated research to company leadership, showcasing deep RL's long-term potential alongside traditional control systems.

Software Engineer Intern, Beaver Health Palo Alto, CA • May 2023 – Aug 2023

- Led a team of interns to build **generative AI** dialogue models digitizing evidence-based dementia interventions for older adults.
- Architected a **multimodal data pipeline**, experimented with lightweight **NLP** models, helping shape the product roadmap on an **NIH**-funded R&D project.
- Built the conversation backend in **Python/Flask** and **TypeScript/Express/Node/TypeORM**.

RESEARCH

Researcher, Carnegie Mellon University, Language Technologies Institute Pittsburgh, PA • Apr 2025 – Present

- Mentored by Nishant Subramani, researching **model interpretability** with a focus on deepening our understanding of neural networks and using those insights to build better systems.
- Model Internal Sleuthing: Finding Lexical Identity and Inflectional Morphology in Modern Language Models (COLM2025 Interplay Workshop)

Researcher, University of Victoria, Department of Mathematics and Statistics Remote • Jul 2022 – May 2023

- Advised by Prof. Xuekui Zhang on COVID-19 forecasting using **TCN**, **CNN**, and **LSTM** models, with statistically significant improvements over CDC ensemble models.
- Processed demographic and mobility data for 3000+ US counties using reproducible **R** and **Python** pipelines.
- Predicting the daily counts of COVID-19 infection using temporal convolutional networks (Journal of Global Health, 2023).

PROJECTS

CLaiM – Cal Hacks 11.0 (Won $1000 for best use of Hyperbolic cloud inference) https://devpost.com/software/autoclaim-q8who1
Created a machine learning system with **Meta's SAM 2** and **YOLOv8** to automate natural disaster home insurance claim processing.

Ad Lunam – To the Moon and Hack (3rd place) https://devpost.com/software/ad-lunam
Developed a physics simulation in **Unity** with procedurally generated planets in a VR environment.

Shipworthy – HackItShipIt (1st place) https://devpost.com/software/shipworthy
Used **OpenCV** to track a physical steering wheel in real-time as a controller for a custom Unity sailing game.

Stance – Data Day Grind (Best Data Visualization) https://devpost.com/software/stance-taking-a-stand-against-hate-speech
Applied **scikit-learn** and **LIME** to build and explain models for detecting toxic online comments.

AWARDS

Science Fairs: 1st Place – Alameda County (2021 & 2022), California Science Fair Presenter (2021 & 2022)
Music: Gold Medalist – U.S. Open Music Competition, Showcase Senior Division (2021 & 2022)