

George Liu

📞 770-365-8764 | ✉️ georgeli@andrew.cmu.edu | 🌐 /georgeliu59 | 🌐 georgeliu59.github.io | 🌐 /georgeliu59

EDUCATION

Carnegie Mellon University

Pittsburgh, PA

B.S. Artificial Intelligence, Additional Major Mathematics, GPA: 4.0/4.0

May 2028

Coursework: Generative AI (PhD), Machine Learning, AI Representations, Probability and Computing, Computer Systems, Data Structures and Algorithms, Theoretical Computer Science, Discrete Mathematics, Matrix Theory

EXPERIENCE

Incoming Software Engineering Intern

May 2026 – Aug. 2026

Databricks

Mountain View, CA

Machine Learning Research Assistant

Dec. 2025 – Present

CMU Robotics Institute (Advised by Prof. Deva Ramanan)

Pittsburgh, PA

- Leading research on cinematographic alignment for **Moodio**, implementing GRPO and DPO to achieve a **6x increase** in instruction adherence and reasoning over base models.
- Architected a high-throughput pipeline to curate **5M+ cinematic videos**, utilizing SOTA VL models for automated dense captioning and VQA-based semantic filtering.
- Scaled Moodio to **100+ paid users**, managing a team of annotators and specialists to curate expert-level datasets and incorporate real user feedback into model alignment and reward signal design.

Teaching Assistant

Aug. 2025 – Present

CMU Computer Science Department (15-259/559: Probability and Computing)

Pittsburgh, PA

- Teaching **140+ students** topics spanning probability theory, stochastic processes, statistical estimation (MLE/MAP), and simulation, leading weekly solo recitations for 20+ students and authoring exam problems.

Software Development Engineer Intern

May 2025 – Aug. 2025

Amazon Web Services (AWS)

Herndon, VA

- Built autonomous network triage system using MCP, LangChain, SageMaker, and Bedrock to automatically diagnose critical infrastructure issues from telemetry logs and test reports.
- Engineered multi-agent AI workflow processing **1M+ daily** telemetry events with **95%** diagnostic precision.
- Improved IP hostname resolution process by **98.5%** (400s → 6s), significantly enhancing oncall engineer productivity and expediting network troubleshooting workflows.

Machine Learning Research Assistant

Sep. 2024 – Present

CMU Computer Science Department (Advised by Prof. Tai Sing Lee)

Pittsburgh, PA

- Researching how structural and sparsity constraints improve **robustness** and **generalization** in visual learning, bridging insights from human cognition and computer vision.
- Developed structure-first pretraining strategies that induce more **human-aligned inductive biases** in deep neural networks, improving resilience to distribution shift.
- Co-first authored a paper presenting a novel **sparse pretraining** method that improves **data efficiency by 30%** and creates 40% more compact neural representations (<https://arxiv.org/pdf/2508.06696>).

Data Science Intern

Sep. 2023 – May 2024

VDart

Alpharetta, GA

- Built intelligent chatbot for data querying using OpenAI API and RAG (Retrieval-Augmented Generation), reducing data retrieval time by **35%** across company-wide databases.
- Developed ML algorithms for candidate skill categorization and contract data analysis, processing **15,000+** records and improving recruitment efficiency by **15%** with **30%** higher data accuracy.

Machine Learning Research Assistant

Dec. 2022 – Oct. 2023

Georgia Institute of Technology (Advised by Prof. Guillermo Goldsztein)

Remote

- Developed a 3D CNN for automated dementia diagnosis on **1,324 MRI scans**, achieving **76.25% accuracy** with a 6% improvement over comparable methods through atrophy analysis (DOI: 10.47611/jsrhs.v12i3.4569).

AWARDS AND SKILLS

Awards: 2025 Putnam Top 650, 5x AIME Qualifier, 2x USNCO National Finalist, National Merit Scholar

Languages: Python, C/C++, Java, SQL, SML/OCaml, Swift, JavaScript, LaTeX

ML & Research: RLHF (GRPO/DPO), Multimodal Learning, Video Understanding, Inference-Time Scaling, Distributed Training, Computer Vision, Sparse Pretraining, RAG, Multi-Agent Systems, Human-AI Alignment

Technical Skills: System Design, Distributed Systems, High-throughput Pipelines, Algorithms & Complexity