LIAM CAWLEY

+1 (917) 319-9518 ◇ New York, NY
cawleyl@umich.edu

EDUCATION

UNIVERSITY OF MICHIGAN, **College of Engineering** Expected May 2025
Bachelor of Engineering, Computer Science Engineering; Minor in Mathematics

COURSEWORK

• Random Numerical Linear Algebra • Principles of Machine Learning • Science of Language Models • Large Language Models
• Graph Algorithms • Mathematics of Finance • Probability Theory • Randomness and Computation • Matrix Algebra • Computer Vision

PAPERS AND WORKSHOPS

- **International Conference on Machine Learning (ICML) 2025 Workshop** "Learning in low-resource settings": Paper Submission, "LoRAMBo: Fighting LoRA Memory Bottlenecks with Adaptive Rank Selection" Liam Cawley

PROJECTS

Local LLM Research and Coding Agent Python, PyTorch, Docker, Ollama, DeepSeek

- Developed specialized coding assistant using DeepSeek R1, implementing custom LoRA adapters with 8-bit quantization to achieve 4GB VRAM footprint. Integrated with Ollama for local deployment and Vast.ai for scalable GPU access.
- Engineered context-injection system leveraging DeepThink's architecture to maintain coherent multi-turn dialogue while preserving code context across interactions. Achieved 87% reduction in context window usage.
- Built custom evaluation pipeline comparing agent performance against human solutions on competitive programming tasks. Demonstrated 65% success rate on medium-difficulty LeetCode problems.
- Created distributed training setup on Vast.ai using A4000 GPUs, achieving 3.2x cost reduction compared to traditional cloud providers while maintaining 94% GPU utilization.

EXPERIENCE

High Performance Computing Intern May 2024 - August 2024
KLA Corporation *San Jose, CA*

KLA is a global leader in semiconductor process control and yield process analysis; customers include Nvidia, Intel, & TSMC.

- Developed a custom 34-layer ResNet to classify nanometer-scale defects on engineered features to approximate thousand-attribute model with a singular attribute.
- Used Grad-CAM and Explainable AI to analyze filters and learned features in data-limited and noisy environments.
- Achieved a false negative rate below 0.001% with a false positive rate under 5%.

Machine Learning Intern May 2023 - August 2023
EMAG Technologies, Inc. *Ann Arbor, MI*

EMAG provides software and hardware solutions for radio frequency & wireless systems analysis, characterization and diagnosis.

- Aerospace Unit, focus on medium range hypersonic missile antennae.
- Developed an original calibration heuristic for digital coherent beamforming in phased arrays.
- Designed a CNN to scale the calibration heuristic for large-scale systems with vast state spaces.
- Applied the calibrator with adaptive nulling to optimize SDR arrays with continuous phase and attenuation controls.
- Achieved a 650% performance increase over traditional sample matrix inversion methods.

Software Engineering Intern May 2022 - August 2022
RTX Fintech & Research, LLC *New York, NY*

- Developed full-stack features for an interest rate derivatives exchange platform, and Price/Time Priority algorithm for trade execution strategies.
- Built API endpoints using Django, OAuthLib, and Serpy to construct Par, Swap, and Forward Rate curves.
- Developed and deployed features in HTML, TypeScript, and Java to an OpenFin environment.

SKILLS

Programming Languages: Mojo, Python, C++, C, Java, TypeScript, HTML, SQL, LaTeX, MATLAB

Tools and Frameworks: TensorFlow, PyTorch, Docker, CUDA, Kubernetes, Django, OpenFin, Git, JIRA and Confluence

Activities: UM HuggingFace Admin, UM Hackers Quant Team, FPV Building/Racing, Archery