Eric Xu

236-881-1361 ericx057@gmail.com linkedin.com github.com

Education

University of Waterloo **2025 – 2030**
Bachelor of Software Engineering *Waterloo, Ontario*

Experience

Co-Founder & Lead Systems Engineer **2025 – Current**
Stealth AI *Waterloo, ON*

- Architected a high-efficiency training engine for **Llama-3.1-70B** on 48GB VRAM, utilizing FA2 and Liger Kernels to reduce MLP memory overhead by **60%**, achieving **high-rank QLoRA fine-tuning at <10% of cloud costs**.
- Engineered fault-tolerance system, achieving **99.9%** training continuity by serializing state during OOM events and implemented protocol to forcibly reclaim fragmented VRAM before hot-swapping worker processes.
- Developed custom memory management routines in Python/CUDA, including surgical precision casting (BF16/FP32 hybrid) and manual page eviction, to **resolve fragmentation issues inherent in native PyTorch garbage collection** on consumer hardware.
- Designed a topology-aware sharding strategy to balance embedding tax across uneven GPU loads, stabilizing unreliable on-premise infrastructure for enterprise-grade workloads.
- Led a 5-engineer team; defined technical roadblocks and coordinated exact execution across the stability, dynamics, and safety modules.

Full-Stack Software Engineering Intern **2023 – 2023**
University of British Columbia *Vancouver, BC*

- Optimized search query performance, reducing average response time by approximately **75.4%** through database indexing and query refactoring.
- Improved data retrieval efficiency by transitioning from a nested search approach to a linear-time $O(n)$ **hash-based lookup**.
- Utilized SQL, React, and Python to develop and refine production software solutions.

Projects

Roblox Game | *Lua* **2024**

- Designed, scripted, and optimized game systems in Lua, driving adoption to **3M+** players and generating **$1,000** in revenue.
- Implemented scalable gameplay logic and performance-conscious scripting, supporting high peak concurrent player loads without degradation.

Quantitative Trading System | *Python, PineScript, Pandas* **2025**

- Designed and implemented quantitative trading strategies leveraging statistical modeling and Gaussian-based interval analysis.
- Conducted rigorous backtesting with out-of-sample validation, achieving **+241%** simulated portfolio growth and an **Alpha of 1.1%** over benchmark.
- Validated performance via paper trading on TradingView, recording **+8%** in Q1 2025 and **+15%** in Q2 2025 simulated growth.
- Collaborated with a team of three engineers to develop and deploy algorithm implementations.

Technical Skills

Languages: Python, C++, SQL, JavaScript (React), Lua, C, HTML/CSS
Frameworks & Tools: PyTorch, Pandas, NumPy, Git, Docker/Kubernetes, CUDA, Linux/Unix