

Evan Kazakin

✉ eckazaki@uwaterloo.ca | 🌐 github.com/classicluna/ | 🔗 linkedin.com/in/evankazakin/

Education

University of Waterloo

Expected April 2027

Bachelor of Mathematics, Combinatorics & Optimization and Pure Mathematics; Minor in Computer Science

Waterloo, ON

- **Relevant Coursework:** Data Structures & Algorithms, Object-Oriented Software Development, Neural Networks, Numerical Computation, Graph Theory, Nonlinear Optimization

Experience

Achievers

May 2026 – Present

Software Engineering Intern

Toronto, ON

- Redesigned a Go cleanup pipeline for millions of MySQL records, replacing concurrent sweeps with a self-republishing workflow; cut peak DB threads from **60 to 20** and peak row-lock waits from **71 to 5 seconds**
- Eliminated a Pub/Sub redelivery loop caused by database contention, reducing dead-lettered messages from **2K to 0** and weekly internal-error alerts from **153 to 21**
- Built cross-service event scheduling in each user's local timezone for **4.7M** users, with message deduplication, idempotent processing, and atomic MySQL transactions
- Fixed a distributed job-state race that killed slow-running jobs; added an intermediate termination state and promotion daemon so completion and failure results reach downstream analytics

Intuit

Jan 2024 – Aug 2025 (3 internships)

Software Engineering Intern

Toronto, ON

- Built core React/TypeScript UI components for BizTax, a new product that generated **\$100K+** in first-month revenue
- Instrumented BizTax flows with 15 Splunk metrics, found a major drop-off and drove a redesign that lifted completion by **30%**
- Launched Tax Prep View in Intuit's Expert Platform, moving **12,000+ tax professionals** from a legacy desktop app to the cloud
- Migrated QuickBooks Online ProTax to TypeScript micro-frontends, enabling independent feature development and deployment
- Built React-based MultiSign support for **1,500+ users**, increasing eSign batch capacity from **1 to 5 forms per request**

Projects & Open Source

Local LLM Inference on AMD | llama.cpp, C++/HIP, Vulkan, ROCm, Python

GitHub

- Ran a **27B** model entirely in **16 GB** of VRAM on an RX 7800 XT, choosing the quant by VRAM budget and a perplexity A/B
- Raised decode from **33 to 59 tok/s** with MTP speculative decoding and KV-cache quantization once the **624 GB/s** bandwidth ceiling ruled out further kernel gains; cut an agent benchmark from **12.7 to 8.3 min** at 8/8 pass
- Ported a 1.75-bpw ternary dequant kernel to HIP, fused Hadamard rotation with Q8 activation quantization, and batched GQA flash-attention decode on RDNA3: **+53%** prefill, **+10%** decode, verified token-identical against CPU

Homelab | Debian, Docker Compose, Traefik, Pi-hole, Prometheus, Grafana, Loki, Tailscale, Bash

- Run **26** containers in 8 Git-versioned Compose stacks behind Traefik TLS on a private CA, with Tailscale remote access
- Built a Prometheus/Grafana/Loki stack with **19** alert rules provisioned as code and routed to Discord
- Wrote a VPN watchdog that checks the tunnel and forwarded port every 5 min, resyncs the client port and re-pins slow exits; traced a tunnel restarting every **2 min** to an ICMP health check broken by NAT id rewriting

Cine-Vector | Python, FastAPI, React, Transformers

GitHub

- Built semantic search for **4,800+ movies** using BERT-based transformers and **384-dimensional embeddings**
- Reduced query latency by **82%** to **<5 ms p95** using an in-memory NumPy vector store with vectorized cosine similarity

Open Source Contributions | Swift

GitHub

- **cmux** (**27K+** stars): fixed the active pane border desyncing from zoomed split panes by driving its geometry from the visible zoom container and clamping it to the overlay canvas; added unit tests covering every panel type

Skills

Languages: Go, TypeScript/JavaScript, Python, SQL, C/C++, Java, Bash

Frameworks & Libraries: React, Node.js, Express, FastAPI, Spring Boot, PyTorch, NumPy

Cloud, Databases & Tools: Docker, Kubernetes, MySQL, AWS, GCP, Linux, Git, gRPC, Grafana, Splunk

Interests: Hiking, Snowboarding, Photography, Birdwatching, Free and Open Source Software