Ri Hong

riri.hong@gmail.com | github.com/Ri-Hong | linkedin.com/in/ririhong

Education

University of Waterloo 2022 – 2027

Bachelor of Computer Science (Co-op) · GPA: 3.9/4.0

Skills

Languages: Python, Java, JavaScript, C++, Go, TypeScript, HTML, CSS
Technologies: React, Next.js, Docker, Kubernetes, Terraform, FastAPI, AWS, Linux, PostgreSQL, gRPC

Experience

Groq Sept 2025 – Dec 2025

Software Engineering Intern *Toronto, ON*

- Quickly ramped on existing infrastructure codebase and pushed production-ready code for multi-region infrastructure with Terraform and Kubernetes, enabling horizontally scalable AI/ML workloads and cutting deployment time by 24% using **Python**.

Base Power Jan 2025 – Apr 2025

Software Engineering Intern *Austin, TX*

- Identified reliability problem in trading simulations, outlined optimal cloud-native solution accounting for tradeoffs, and pushed production-ready code using Temporal Cloud, enabling 1000s of auto-retriable workflows with few iterations.
- Discovered performance bottleneck in real-time market data transformation, analyzed edge cases and tradeoffs, and implemented optimal O(n log n) solution with async publishing, achieving a 32% speedup; communicated effectively with algorithm developers and infrastructure teams.

Trend Micro May 2024 – Aug 2024

Software Engineering Intern *Ottawa, ON*

- Quickly ramped on legacy codebase and pushed production-ready code, upgrading Deep Security Manager from JDK 8 to JDK 11 while accounting for compatibility edge cases, modernizing the system for over 250 million global customers using **Java**.
- Identified CI/CD pipeline limitations, outlined optimal solution considering deployment tradeoffs, and completed the Jenkins pipeline revamp efficiently with few iterations, achieving a 35% increase in automation efficiency; communicated effectively across DevOps and QA teams.

Walnote.ai Aug 2025 – Present

Founder & CTO *Toronto, ON*

- Identified scalability problem in video generation pipeline, outlined optimal solution accounting for latency and cost tradeoffs, and pushed production-ready code with Celery + **FastAPI** using **Python**, processing 1,000+ animations efficiently with few iterations.

Projects

Trasee | *Pyodide, React, TypeScript, JavaScript* Oct 2025

- Identified problem in algorithm visualization, outlined optimal two-phase solution accounting for tradeoffs, and delivered production-ready code visualizer using **Python**, **JavaScript**, and **TypeScript** with few iterations.

Neural Style Transfer Engine | *PyTorch, CUDA, FastAPI, Docker* Sept 2025

- Analyzed performance bottlenecks and edge cases, outlined optimal solution with custom CUDA kernels accounting for speed-quality tradeoffs, and pushed production-ready code achieving 8x faster style transfer with **FastAPI** and **Python**.

Publications

Pruned Graph Generating State Space Models (in preparation) Oct 2025

Research Paper

- Proposes a pruning strategy to sparsify graph structure in GG-SSMs, reducing FLOPs while preserving accuracy and improving computational efficiency for high-dimensional data.