Ri Hong

ri.hong4@gmail.com | github.com/Ri-Hong | linkedin.com/in/ririhong

10× Hackathon Winner · Applied ML Engineer · Cloud and Distributed Systems Engineer

Education

University of Waterloo 2022 – 2027

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

- Relevant coursework: Algorithms, Operating Systems, OOP, Artificial Intelligence, Distributed Systems, Linear Algebra, Numerical Methods, Machine Learning, Computer Vision

Skills

Languages: Go, Python, C, C++, Java, JavaScript, TypeScript
Technologies: CUDA, Kubernetes, Docker, Jenkins, Ansible, Terraform, AWS, Linux, Bash, SQL, Git, Node, React
AI/ML: PyTorch, TensorFlow, CUDA, cuDNN, scikit-learn, MLflow

Experience

Groq Sept 2025 – Dec 2025

Software Engineering Intern *Toronto, ON*

- Engineered multi-region infrastructure with **Terraform** and **Kubernetes** on GCP, enabling scalable **distributed systems** AI/ML workloads and cutting deployment time by 24%.
- Designed observability platform measuring engineering productivity and AI tool adoption, providing actionable insights for 400+ engineers.

Base Power Jan 2025 – Apr 2025

Software Engineering Intern *Austin, TX*

- Improved trading simulation reliability using cloud-native **distributed systems**, enabling 1000s of auto-retriable workflows and eliminating single-node failure risks.
- Resolved performance bug reducing complexity from $O(n^2)$ to O(n log n); implemented async publishing using **Go** routines and **Protobuf/gRPC** for **Python-Go microservices** communication, achieving 32% speedup and 20% serialization overhead reduction.

Trend Micro May 2024 – Aug 2024

Software Engineering Intern *Ottawa, ON*

- Upgraded legacy system from JDK 8 to JDK 11 using **Java**, modernizing codebase for 250M+ customers; revamped **Jenkins** CI/CD pipeline, achieving 35% increase in **automation** efficiency.
- Resolved 40+ installation and deployment issues through **debugging** on **Linux AWS** EC2 instances; refactored monolithic codebases into **microservices**, cutting deployment errors by 30%.

Walnote.ai Aug 2025 – Present

Founder & CTO *Toronto, ON*

- Launched AI platform combining GPT-5 with Manim for auto-generated explainer videos; accelerated generation using **distributed systems GPU computing** rendering, reducing latency from 12s to 2s; scaled with **Docker**, FFmpeg, and PostgreSQL.

Projects

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

- Engineered high-performance **deep learning** Neural Style Transfer system with custom **CUDA** kernels, achieving 8x faster style transfer through **GPU computing**; built scalable REST API with FastAPI, containerized with **Docker**, delivering production-quality results in under 30 seconds using **accelerated computing**.

DistilBERT Sentiment Analysis | *PyTorch, GCP, **Kubernetes**, Terraform* Sept 2025

- Engineered production-grade sentiment analysis service achieving 92.5% accuracy using **deep learning** DistilBERT, deployed on GKE with Terraform; optimized with mixed precision and **distributed systems GPU computing** training on **Kubernetes**, reducing training time by 40%.

HomeLab | *Linux, Proxmox, **Ansible**, **Kubernetes*** Apr 2025

- Deployed **Kubernetes** cluster on Proxmox with Terraform + **Ansible**, enabling scalable, self-healing **microservices** and secure HTTPS routing.