

Ri Hong

ri.hong16@gmail.com | github.com/Ri-Hong | [linkedin.com/in/ririhong](https://www.linkedin.com/in/ririhong)
10× Hackathon Winner · Applied Machine Learning Engineer

EDUCATION

University of Waterloo

2022 – 2027

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

- Relevant coursework: Artificial Intelligence, Machine Learning, Algorithms, Object-Oriented Programming

SKILLS

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

AI/ML: PyTorch, TensorFlow, scikit-learn, LangChain, OpenAI GPT-4/5, CUDA, MLflow, DVC

Technologies: Docker, Kubernetes, Terraform, Temporal, gRPC, GCP, AWS, PostgreSQL, BigQuery

AI Tools: GitHub Copilot, Claude Code, Cursor, ChatGPT

EXPERIENCE

Groq

Sept 2025 – Dec 2025

Applied AI Engineering Intern

Toronto, ON

- Designed and launched an observability platform to measure engineering productivity and **AI coding tool** adoption (Claude Code, Cursor, AmpCode), providing actionable insights for 400+ engineers
- Developed dashboards and analytics workflows that surfaced correlations between **AI tool usage** and code quality, driving data-informed adoption strategies

Walnote.ai

Aug 2025 – Present

Founder & CTO

Toronto, ON

- Launched an **AI platform** combining **GPT-5** with Manim to auto-generate explainer videos; processed 1,000+ animations with a Celery + FastAPI pipeline
- Accelerated generation by segmenting code generation and running distributed **GPU rendering**, reducing render latency from 12s to 2s; leveraging a pipelined streaming workflow to deliver near real-time playback
- Scaled secure rendering and delivery with Docker, FFmpeg, Cloudflare R2, and PostgreSQL; raised \$20k pre-seed and led product strategy from prototype to production

Base Power

Jan 2025 – Apr 2025

Software Engineering Intern

Austin, TX

- Reduced simulation time by 5x by identifying bottlenecks with OpenTelemetry and redesigning the simulation engine to support parallel processing via Kubernetes-based horizontal scaling
- Improved simulation speed and reliability by migrating local simulations onto Temporal Cloud, storing intermediate state in AWS S3 to allow for retry on failure

Microgreen Solar

Sept 2023 – Dec 2023

Software Engineering Intern

Toronto, ON

- Built a real-time telemetry ingestion pipeline using AWS IoT Core to stream battery data into AWS RDS, enabling persistent storage and downstream processing for off-grid energy systems
- Developed a responsive Progressive Web App with React, Next.js, and TailwindCSS to visualize live and historical telemetry; introduced Redis caching to reduce query latency from 5000ms to 100ms

PROJECTS

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

Sept 2025

- Engineered a production-grade sentiment analysis service achieving 92.5% accuracy using DistilBERT, deployed on GKE with Terraform
- Implemented end-to-end **MLOps** pipeline with MLflow for experiment tracking, DVC for data versioning, and BentoML for model serving
- Optimized training by 40%, using mixed precision and distributed **GPU** training with Kubernetes

FlaimBrain | *LangChain, GPT-4, MongoDB, Flask*

Oct 2023

- Engineered an **AI** study assistant that converts uploaded notes into summaries and flashcards using GPT-4 and **LangChain**
- Led the **RAG** architecture with MongoDB Vector Search + Flask backend, earning 4th place at Hack the Valley 8 (80 teams)