

Ji Sung Han

+1 647-321-9811 | jishan.han@mail.utoronto.ca | GitHub | LinkedIn | Personal Website

EDUCATION

University of Toronto

Jan. 2022 – Dec. 2026

Honours Bachelor of Science in **Computer Science and Statistics**

- **Technical Director** at UofT Blockchain, **Teaching Assistant** (Calculus), Focus in AI & Machine Learning.
- **Relevant Coursework:** Databases, Operating Systems, Software Engineering, Algorithm Design, Deep Learning.

EXPERIENCE

NVIDIA

Winter 2026

Systems Software Engineer Intern

- On the Performance Lab team, improving AI/HPC performance by profiling and debugging C++ services on Linux and analyzing CPU–GPU bottlenecks to reduce workload latency.
- Build Python profiling and telemetry pipelines to standardize performance KPIs detect performance regressions and reduce triage time for inference performance issues.

Tesla

Sep. 2025 – Jan. 2026

Data Engineer Intern

- Built a real-time streaming pipeline (Kafka, Spark Streaming, Kubernetes) with auto-scaling to process **50M+ events/day** and cut data lag from minutes to seconds (p95 latency **under 1 second**).
- Productionized a batch ETL + ML pipeline (Airflow, Spark) integrating time-series risk scores from a scikit-learn model for inactive installer prediction driving **about 5 reactivations/week** and an estimated **\$15M/year** in value.
- Optimized Airflow DAGs using vectorized pandas and SQL partitioning/indexing on **1M+ rows/day** cutting runtime from **5 hours to 5 minutes** and enabling daily decision cycles.
- Deployed an XGBoost Top-K model with FastAPI human-in-the-loop review and rule-based signals to match **500K+ entities/day** achieving **97% accuracy** and automating **90%** of manual operations.
- Reduced release cycles by **40%** by automating CI/CD for Airflow workflows with reliable deployments plus retries and alerting.

Colliers

May 2025 – Aug. 2025

Software Engineer Intern

- Built a semantic search and summarization platform (Python FastAPI, Angular, LangChain, Vertex AI) on GCP enabling **40+ brokers** to analyze leases **3x faster**.
- Reduced frontend query latency from **5.0 seconds** to **3.5 seconds** via Redis caching and query optimization improving UX with async summarization (Python asyncio).
- Built a distributed OCR + LLM ELT pipeline (Airflow, PySpark) extracting **60+ fields** from **10K+ leases** cutting hourly batch runtime by **52%** (from **58 minutes** to **28 minutes**).
- Replaced a 1–3 day manual extraction workflow with automated hourly ELT into BigQuery powering Power BI dashboards enabling **1,000+ analysts** to refresh in **under 1 hour** and saving **\$25K/year**.

University of Toronto

May 2024 – May 2025

Research Assistant

- Built a transformer ensemble (BERT/RoBERTa/XLNet) in PyTorch and TensorFlow using GNN-augmented encoders and Integrated Gradients attribution improving crypto fraud detection F1 by **12%**.

PROJECTS

Lifomation (Co-founder) | Angular, Azure OpenAI, Redis, BullMQ, Jenkins, Docker, Dynatrace

- Co-developed a cloud-native AI document platform in an 8-person team building an Angular frontend and scalable RAG workflows powered by Azure OpenAI Redis and BullMQ.
- Improved Azure OpenAI privacy features for **1,000+ users** and supported a product demo presented to Sequoia Capital.

TECHNICAL SKILLS

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

Frameworks: FastAPI, Flask, Node.js, Express.js, Angular, React, PyTorch, TensorFlow

Cloud / DevOps / Infra: AWS, GCP, Azure, Docker, Kubernetes, CI/CD (Jenkins, GitHub Actions), gRPC