Roger Chan

(647)-823-6889 | roger.chan@uwaterloo.ca | linkedin.com/in/rogerzchan/ | github.io/rogerchan

Technical Skills

Languages: Python, SQL (Redshift, PostgreSQL), R, Bash/Shell, TypeScript, JavaScript
AI / ML: PyTorch, Transformers, Hugging Face, XGBoost, LightGBM, scikit-learn, statsmodels, RAG, Uplift Modeling
Data & Visualization: dbt, Preset (Apache Superset), Tableau, Pandas, Polars
Tools & Platforms: AWS (Bedrock, Claude API, EFS), Apache Airflow, Kubernetes, FastAPI, Firebase

Experience

Wealthsimple – *Data Science Intern* Jan 2026 – Present | Toronto, ON

- Tuned class-weight dampening and regularization on a **PyTorch foundation model** (46M params, 6-layer transformer) trained on **8.1M financial event sequences**, improving next-event prediction accuracy from 70.3% to **75.0%** and macro F1 by **4.7%** across 2.55M user embeddings used for fraud detection.
- Built an **LLM-powered AML alert triage system** using Claude (AWS Bedrock) that achieved an **87% match rate** against analyst dispositions across 1,805 cases, reducing dangerous false dismissals to **0.27%**.
- Built a **dbt pipeline monitoring bot** (custom Airflow operator + Slack) that streams live progress for multi-hour data-mart builds; engineered its cross-pod logging on **Kubernetes** via shared **EFS** storage to carry it through the platform's **Nomad-to-K8s migration**, and shipped it as the first **cross-team** adopter.

Vale Base Metals – *Data Science Intern* May 2024 – Dec 2025 | Toronto, ON

- Developed and maintained scalable **data pipelines** in Python and Power Automate to centralize Continuous Improvement data from **four global regions**, ensuring data quality and governance.
- Integrated **classification ML models** into PowerApps to recommend optimal CI methodologies, automating project routing and boosting submission throughput by **25%**.

Ontario Securities Commission (OSC) – *Data Science Intern* May 2023 – Dec 2023 | Toronto, ON

- Designed and deployed a **RAG (Retrieval-Augmented Generation) system** over **thousands of regulatory PDFs and unstructured financial disclosures**, enabling analysts to query fund-related filings in natural language and reducing manual document review time by **60%**.
- Developed **Python audit-automation scripts** and a **Streamlit dashboard** that gave auditors self-serve access to variance checks and filing lookups, cutting audit review time by **55%** and eliminating **300+ manual Excel workflows per quarter**.

Research Paper

Base-Learner Sensitivity in Uplift Meta-Learner Benchmarks Dec 2025

- Evaluated **60 uplift-modeling configurations** (6 meta-learners × 5 base learners) on Criteo v2.1 (**11.2M rows**) and Hillstrom; diagnosed a novel **S-learner degeneracy** (**0% of splits** on the treatment column, Qini **46**), contradicting UpliftBench (2026)'s #1 ranking.
- Proved meta-learner leaderboards are dominated by base-learner choice: **Class-Transformation Qini swung 4,527 → 8,261 (+83%)** by swapping XGBoost for Logistic Regression, and cross-dataset rank correlation was only **Spearman $\rho = 0.18$**.

Community Impact

City of Iqaluit Smart Water Delivery App Jan 2024 – Apr 2024

- Developed a mobile application with the **City of Iqaluit** to modernize trucked water delivery for **470+ households**, integrating **IoT sensors** with Firebase for real-time tank monitoring and automated scheduling and reducing operational costs by **$250K/year**.
- Implemented an **ARIMA-based ML forecasting model** to predict household water demand and minimize off-schedule deliveries.

Education

University of Waterloo, Faculty of Engineering Waterloo, ON
Master of Management Engineering (MMSc), Specialization in Data Science *Expected Sept 2027*

University of Waterloo, Faculty of Health Waterloo, ON
Bachelor of Science, Minor in Computing *Graduated 2024*