

# Joseph Ngai

647-638-5400 | [j5ngai@uwaterloo.ca](mailto:j5ngai@uwaterloo.ca) | [linkedin.com/jojo.ngai](https://linkedin.com/jojo.ngai) | [github.com/jojo.ngai](https://github.com/jojo.ngai)

## EDUCATION

### University of Waterloo

Bachelor of Engineering in Management Engineering

Waterloo, ON

Sep 2021 - May 2026

## EXPERIENCE

### Software Engineer

Sept 2025 – Dec 2025

Opener

Toronto, ON

- Engineered data-cleaning workflows in SQL, designing various rules while preserving foreign-key relationships
- Developed a client-facing bulk-upload feature: built the UI, scheduled background jobs with BullMQ, and created REST/GraphQL APIs to validate incoming data, insert new stores, and maintain data integrity.
- Implemented a recommendation engine using Elasticsearch vector search

### Machine Learning Engineer

May 2025 – Aug 2025

MissingWealth

Toronto, ON

- Deployed an ETL pipeline on GCP to extract, transform, and cleanse email data using Python and Jupyter
- Developed a classification system using Gemini and zero and few-shot learning, achieving an F1-score above 0.90

### Data Scientist

Jan 2025 - Apr 2025

RBC

Toronto, ON

- Built an ETL pipeline from a conversational API platform to Postgres DB hosted on an in house DBaaS
- Configured and deployed an application on Helios, a Docker orchestration platform
- Clustered conversational data using a BERTopic model with K-Means and HDBSCAN clustering algorithms
- Summarized each cluster and generated curated training phrases using openAI's gpt 3.5 api LLM

### Data Analyst

May 2024 – Aug 2024

RBC

Toronto, ON

- Automated financial and employee reporting by developing Python scripts, eliminating *all* manual intervention
- Optimized employee movement tracking in Superset by creating 50+ SQL views, validation queries, and charts
- Built data cleansing and preparation scripts that processed 10000+ data points, enabling real-time BI dashboards

### Data Governance Analyst

Sep 2023 – Dec 2023

Toyota Canada

Toronto, ON

- Improved Python-based resume parser with Pandas, PyPDF2 and text classification. Used to hire following interns
- Developed a weather API dashboard with Power BI, Python scripts for data extraction, and SQL DB integration

## PROJECTS

### TradeWatt | Python, Neural Networks, Regression, Decision Trees, MongoDB

May 2025 - Aug 2025

- Designed an automated ETL pipeline from energy market data into a structured MongoDB database
- Developed a machine learning pipeline for price prediction, integrating ETL, preprocessing, and model training
- Implemented and evaluated multiple ML models to optimize predictive accuracy

### MindfulMeals | Python, Javascript, React, Node.js, Firebase, API, SQL

Jan 2024 - Apr 2024

- Built a recipe management app with Firebase authentication, profiles, dietary filtering, and a recipe finder
- Implemented key features including real-time search, grocery cart, saved recipes, user reviews, and recipe filtering.

## TECHNICAL SKILLS

**Languages:** Python, Java, Typescript SQL (Postgres), JavaScript, HTML/CSS, R, GraphQL

**Frameworks:** React, Node.js, Django, Jenkins, RestAPI, FastAPI

**Developer Tools:** Git, Docker, VS Code, Visual Studio, Jupyter Notebook, Postman, Eclipse, Cursor

**Libraries:** Pandas, NumPy, Matplotlib, Scikit-learn, Seaborn, TensorFlow, PyTorch

**Databases and Cloud Services:** MySQL, Postgres, MongoDB, Firebase, AWS, Azure, Hasura