Dylan Dai



🏶 dylanydai.github.io 💟 dylan.dai@uwaterloo.ca in linkedin.com/in/dylanydai 🖸 github.com/dylanydai

EDUCATION

University of Waterloo

Expected May 2028

Bachelor of Computer Science (Honours)

Waterloo, Ontario

Coursework: Data Structures & Algorithms, Compilers, Object Oriented Programming, Computer Architecture

SKILLS

Languages: Python, C++, C, SQL, Bash

Frameworks & Tools: Concurrency, Pandas, Scikit-learn, PyTorch, NumPy, Git, PostgreSQL, GCP, Cursor Interests: Fashion, Digital Art, Tetris, Puzzle Games, Thrifting, Guitar, Rubik's Cube Puzzles, Travelling

WORK EXPERIENCE

Stealth September 2025 – December 2025

Software Engineer Intern

San Francisco, CA

- Seed stage startup backed by **Sequoia Capital** to enhance pharmaceutical market research
- Building data pipelines to parse pharmaceutical surveying data for synthetic data creation and backtesting
- Building synthetic data creation pipelines by training machine learning models and creating evaluation tooling

Cohere

May 2025 - August 2025

Software Engineer Intern

Toronto. ON

- Saved \$100,000+ monthly and 35% in runtime by adding batching to all company-wide AI model calls
- Reduced evaluation effort by 30% by building a tool to access all Cohere's AI model benchmark statistics
- Saved 20% in GPU runtime by adding cost tracking and aggregation for all company-wide AI model calls
- Reduced server load for storing AI model queries by 90% from implementing decision trees for item indexing

Cohere

September 2024 – December 2024

Data Engineer Toronto, ON

- Managed coding datasets used to train state of the art machine learning model Command-A
- Implemented web-scraper to extract 1,000+ questions from programming websites for LLM training datasets
- Designed and solved 200+ data structure and algorithm problems to train and evaluate Cohere's LLM models

PROJECTS

AI Dataset Undersampler [7] | NumPy, TypeScript, Three.js, Scikit-learn

- Diversifies AI model training datasets by 30% by building a tool to analyze and filter data using k-means
- Visualized data by embedding data then using Principal Component Analysis for vector compression

Music Tracking Game [7] | MATLAB, Flask

- Evaluated audio similarity in real-time with cross-correlation for lag correction and amplitude scaling
- Built Flask backend for music performance game via audio stream to MATLAB

AWARDS AND ACHIEVEMENTS

Canadian Computing Olympiad Bronze Medalist | Placed 14th out of 10,000+ participants

National Speedcubing Competitor | Ranked top 50 nationally with best Rubik's cube solve of 6.22 seconds

National Band Festival Winner | Mentored group of 30+ clarinet musicians to gold award performance

National ranked Tetris player | Top 50 global for tetris.com

Hack the 6ix Winner | Won \$1,000 for best Vellum project out of 500+ participants

GenAI Genesis Hackathon Winner | Won \$800 for best DEI project out of 700+ participants

UTRAHacks Winner | Won \$300 for best Databricks project out of 400+ participants