Dylan Dai

dylanydai.github.io | dylan.dai@uwaterloo.ca | linkedin.com/in/dylanydai | github.com/dylanydai

EDUCATION

University of Waterloo — Expected May 2028
Bachelor of Computer Science (Honours) — *Waterloo, Ontario*
Relevant coursework: Functional Programming, Compilers, Calculus, Linear Algebra, Computer Organization & Design, Algorithms

Canadian Computing Olympiad | Bronze Medalist
- Placed **14th out of 10,000+ participants** in national computing competition with **less than a year** of programming experience

SKILLS

Languages: Python, C++, C, SQL, Bash
Frameworks & Tools: AsyncIO, Pandas, Weights & Biases, PyTorch, NumPy, Git, PostgreSQL, MongoDB, GCP, Cursor
Interests: Fashion, Digital art, Tetris, Puzzle games, Thrifting, Guitar, Rubik's Cube puzzles

WORK EXPERIENCE

Stealth — September 2025 – December 2025
Software Engineer Intern — *San Francisco, CA*
- Seed stage startup with sufficient seed funding backed by general catalyst and sequoia to speed up pharmaceutical market research
- Collaborating with researchers to build data pipelines to parse medical surveying data for synthetic data creation and backtesting
- Building synthetic data creation pipelines by training machine learning models and backtesting via existing data

Cohere — May 2025 – August 2025
Software Engineer Intern — *San Francisco, CA*
- Saved **$100,000+ monthly** and **35%** in runtime by adding **batched query processing** on company-wide AI model calls
- Reduced management effort by **30%** by building a tool to access all of Cohere's AI model benchmarks and runtime 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 indexing

Cohere — September 2024 – December 2024
Data Engineer — *San Francisco, CA*
- 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+** advanced data structure and algorithm problems to train and evaluate Cohere's LLM models

PROJECTS

AI Dataset Undersampler
- Diversifies AI model training datasets by **30%** by building a tool to analyze and filter data using **k-means**
- Visualized data by implementing vector compression via transformers

Music Tracking Game
- Implemented **cross-correlation** and lag adjustment to compare live audio to in-game music through vectorizing amplitudes

AWARDS AND ACHIEVEMENTS

National Speedcubing Competitor | Fastest Rubik's cube solve of **6.22 seconds**, ranked **top 50** nationally

National Band Festival Winner | **First chair Clarinetist**, led section of **30+** musicians to **gold award** performance

National ranked Tetris player | **Top 50** global for tetris.com, **top 150** nationally for tetr.io

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

GenAI Genesis Winner | Won **$800** for best Diversity Equity Inclusion project out of **700+ participants**

UTRAHacks Winner | Won **$300** for best use of Databricks out of **400+ participants**

JAMHacks Winner | Won **$100** for best use of MATLAB out of **300+ participants**