Dylan Dai

dylandai.vercel.app DylanYDai@gmail.com linkedin.com/in/dyland06 github.com/suoeh

SKILLS

Languages: Python, C++, C, Java, JavaScript, TypeScript, Bash, Racket, Scheme, Clojure, HTML, CSS, SQL, Haskell
Technologies: Git, React, React Native, Flask, Google Cloud Platform, Terraform, MongoDB Atlas, Arduino, AWS, Node.js, Next.js, React.js, PyTorch, Linux, NumPy, Pandas, Puppeteer, BeautifulSoup, MATLAB, LangChain, Whisper

EDUCATION

University of Waterloo Waterloo, Ontario
Bachelor of Computer Science
Faculty GPA: 3.7/4.0
Relevant coursework: Compilers, Advanced Functional Programming, Calculus, Linear & Abstract Algebra

WORK EXPERIENCE

Cohere | Python Toronto, Ontario
Machine Learning Engineer Intern *May 2025 – Present*
- Developing tooling to evaluate large language models and improve training data

Cohere | Python, C++, Java, JS, HTML, CSS, React, Bash Toronto, Ontario
Machine Learning Data Consultant *August 2024 – May 2025*
- Oversaw and managed large-scale code datasets used to train Cohere's Command A learning model, which outperforms Deepseek-V3 and GPT-4o on the RepoQA benchmark by a **margin of 10%**
- Designed and solved advanced data structure and algorithm problems to train and evaluate Cohere's LLM models
- Optimized and reviewed over **700** coding test entries for evaluating the quality and accuracy of LLM-generated code
- Improved coding abilities of large language models including Command R+, helping achieve **71.4%** on the MBPPPlus and **22.2%** on the LBPP benchmarks by providing reinforcement learning from human feedback

PROJECTS

Training Data Undersampling tool (Github) | Python, Three.js, Next.js, Flask *March 2025*
- Won best **Diverse AI Hack** from **620+ participants** for building a tool to diversify and analyze AI training data
- Made a **web-based data-processing platform** to identify similar datapoints for dataset filtering and diversification
- Vectorized uploaded raw text using a Transformer, then used **k-means clustering methods** to segment data for bias analysis
- Added **bias identification** by querying LLMs to pinpoint attributes in clustered results.
- Implemented data visualization using **PCA vector compression** to render clustered data points using **Three.js**.

Exercise Assistant (Github) | MongoDB Atlas, Terraform, AWS, Arduino, Python, LangChain, DataBricks *January 2025*
- Won **Best use of DataBricks** from **340+ participants** by developing a web-based physiotherapy game
- Stores and recreates exercises by translating movement vectors into absolute position from controller data
- Incorporated live feedback via a GenAI-powered voice assistant to enhance user engagement and retention using **LangChain**
- Deployed the platform with **Terraform** on **GCP** for highly scalable infrastructure and reliable performance
- Designed a **Flask** backend to integrate with **MongoDB** on **AWS** for persistently storing user heartrate and exercise data

Music Tracking Game (Github) | MATLAB, Flask, HTML, CSS, JavaScript *June 2024*
- Won **Best use of MATLAB** from **200+ participants** by developing a musical accuracy tracking game
- Implemented **cross-correlation** to compare two audio waves by extracting two vectors of amplitudes
- Tracked musical accuracy by adjusting lag from both audio waves, giving live feedback

Waste Sorter App (Devpost) | Google Cloud, Hugging Face, Flask, JavaScript, Python, React Native *May 2023*
- Made an application built with **React Native** to identify and sort waste using a camera for object detection with **90% accuracy**
- Used **Google Cloud's Vision AI** for object detection and classification to send to manually parsed waste database with **JS**
- Combined embeddings with a **384 dimensional dense vector space** using AI model **all-MiniLM-L6-v2**
- Sent response data and associated weights to the **REST API** of the **Flask backend** to identify disposal strategies

AWARDS

Canadian Computing Olympiad | Bronze Medalist *May 2024*
- Placed **14th out of 10000+ participants** in national-level computing competition, invited to Canada's team selection contest
- Solved algorithmic problems using data structures, graph theory, and combinatorics in C++