

VIVIAN DAI

Ithaca, NY: 607-280-7928 | Toronto, ON: 416-620-0366
vzd2@cornell.edu | [linkedin.com/in/viviandai](https://www.linkedin.com/in/viviandai)

EDUCATION

Cornell University BS, Computer Science

Aug 2019 – May 2023
Ithaca, NY

- GPA: 4.03, Dean's List (all semesters)
- Relevant Courses: Machine Learning, Algorithms, Data Structures, Natural Language Processing, Computer Vision, Operating Systems, Functional Programming, Object-Oriented Programming, Probability and Stats, Linear Algebra
- Activities and Involvements: Cornell Electric Vehicles, Association of CS Undergrads, CS course staff, Hortus Forum

EXPERIENCE

Software Engineering Intern Google

May 2022 - Aug 2022
Sunnyvale, CA

- Investigated Woodshed, a new training data storage system to improve modeling workflow for Cloud AI Retail Search ML infrastructure, becoming the first Google Cloud users of the storage system.
- Implemented Flume/Beam writer and TFX reader pipelines with Python and C++, greatly reducing the efforts and rollout time of implementing new experimental features and decreasing example generation redundancy.
- Became one of the first adopters of Woodshed-TFXIO integration across the company, gaining recognition for the project's use case.

Software Engineering Intern Meta

Sept 2021 - Dec 2021
Remote

- Created multiple new ML features for Instagram creator classification model to improve the model's performance and runtime via internal ML and data frameworks (Dataswarm, Fluent2), Python, and SQL.
- Developed Dataswarm pipeline to resolve rater labelling errors present in user data utilized by all IG creator models.
- Implemented NLP techniques and algorithms (TF-IDF, MinHash) to improve text feature embeddings coverage.
- Benchmarked the performance of 20+ GBT and neural network classification models, considering feature importance and hyperparameter sweeping.

STEP Intern Google

June 2021 - Sept 2021
Remote

- Developed a full-stack application to display room and rate-plan metadata on Google's Hotel Center site, improving the onboarding and debugging capabilities of hotel partners who previously could not see the data they sent.
- Extended the Hotels Price Feeds ARI reporting service to read and respond to queries with room and rate-plan data using C++, ProtoBufs, Bigtable, and Google's internal tooling (Cider, Blaze, Critique, etc.).
- Created an intuitive and functional page with TypeScript, AngularJS, and Java that presents data in a user-friendly way to be viewed and used by all Google hotel partners.

Software Team Lead Cornell Electric Vehicles

Oct 2019 – present
Ithaca, NY

- Leading a team of 15+ team members with the design and development of an autonomous vehicle software system.
- Worked with PointPillars object detection algorithm using PyTorch and Open3D for an autonomous vehicle in the Shell Eco-Marathon.
- Benchmarked new data metrics in the validation pipeline for residual neural network object-detection algorithms.

Teaching Assistant Cornell University

Sept 2020 – present
Ithaca, NY

- Held in-person discussion sections, office hours, and graded assignments for Cornell's Data Structures and Functional Programming course (CS 3110, Fall 2022, Spring 2021), Intro to Algorithms course (CS 4820, Spring 2022), and Discrete Mathematics course (CS 2800, Fall 2020).

SKILLS

Languages: Python, Java, C++, OCaml, SQL, C

Tools: Git, Mercurial, VS Code, Protocol Buffers, Flume/Beam