

Stanley Z. Hua

Toronto, Canada

(+1) 4379863444 | stanley.hua@mail.utoronto.ca | [linkedin](#) | [github](#) | [website](#)

EDUCATION

University of Toronto

Honours BS Computer Science Specialist, Statistics Minor (GPA: 3.84/4)

Toronto, CA

Sept. 2022 – June 2024

University of Toronto

Honours BS Bioinformatics & Computational Biology Specialist

Toronto, CA

Sept. 2019 – Sept. 2022 (Switched Majors)

TECHNICAL SKILLS

Languages: Python, SQL, Shell Script, Git, Javascript, HTML/CSS, C/C++, Java, R, Assembly

Software Tools: Airflow, AWS, Docker, GitHub, Redis, Figma

Data Skills: Data Engineering, Data Visualization, Machine Learning, Computer Vision, NLP

Python Libraries: Pandas, NumPy, Matplotlib, Dask, Ray, psycpg2

ML Libraries: PyTorch, Tensorflow, Scikit-Learn, MLFlow

CV/NLP Libraries: Transformers, Langchain, Guardrails, nltk, OpenCV

RESEARCH EXPERIENCE

ML Research Assistant

University of California, Berkeley (Supervisor: Prof. Irene Chen)

May 2024 – Present

Remote

- Investigating the impact of quantization on fairness in LLMs for healthcare

ML Research Assistant

Vector Institute (Supervisor: Prof. Rahul Krishnan)

Jan. 2024 – Present

Toronto, CA

- Exploring multi-agent prompting methods to improve zero-shot performance of open-source LLMs for rare diseases

Junior ML Specialist

The Hospital for Sick Children (Supervisors: Prof. Lauren Erdman, Alex Lu, Prof. Irene Chen)

May 2022 – Aug. 2023

Toronto, CA

- Demonstrated that (MoCo) **supervised contrastive image pre-training** can improve **in-domain** and **out-of-distribution generalization** for automatic view labeling on renal **ultrasound videos**.
- Trained and deployed CNN to forecast ER patient volumes, and explored Bayesian methods (GP, Bayesian NNs) for confidence interval estimation.

ML Research Assistant

The Hospital for Sick Children (Supervisors: Prof. Lauren Erdman, Prof. Anna Goldenberg)

Sept. 2021 – May 2022

Toronto, CA

- Adapted video-based deep learning methods for **feature aggregation** (Conv.Pooling, CNN-LSTM, TSM) to predict if a child needs kidney surgery from ultrasound images taken over multiple hospital visits.
- Demonstrated that ultrasounds from the first hospital visit alone is enough to predict the need for surgery.

ML Research Assistant

University of Toronto (Supervisors: Prof. Alan Moses, Alex Lu)

May 2021 – Sept. 2021

Toronto, CA

- Created the **CytoImageNet dataset** (890K images, 894 classes) from 20 TB of open-source microscopy images.
- Showed that CytoImageNet-pretrained models are competitive with ImageNet-pretrained models on downstream microscopy datasets, despite weak labels and minimal hyperparameter tuning (compared to ImageNet).
- The CytoImageNet dataset attracted attention on Kaggle (10696 views, 526 downloads).

ML Research Assistant

University of Toronto (Supervisor: Prof. Pascal Tyrrell)

Jul. 2020 – Jul. 2021

Toronto, CA

- Showed that choice of dimensionality following dim. reduction (PCA, autoencoder) is important for clustering (K-Means, DBSCAN, Agglomerative) of medical images under small sample sizes using Tensorflow.

INDUSTRY EXPERIENCE

Data Science Intern

June 2023 – Sept. 2023

Wealthsimple

Toronto, CA

- Created a **framework** to perform **zero-shot classification** with LLMs to **annotate any text data** for **downstream model training**, resulting in significantly lower labeling errors (5%) versus human-annotated labels (20%) for customer service ticket routing. Communicated with stakeholders to define requirements, implemented asynchronous batched calls and automated validation with guardrails.
- Designed, implemented and deployed model to an API endpoint that detects if a user entered invalid account details when initiating an account transfer, potentially saving \$12M in flagged transfers annually.

Software Engineer Intern

May 2022 – May 2023

Intel Corporation

Toronto, CA

- Refactored a **data extraction tool** that allows users to easily retrieve data for benchmarking experiments from the database and file system, resulting in improved efficiency, code architecture and test coverage (99%).
- Optimized **SQL** and Pandas code for existing dashboards, speeding up loading time by 400%.
- Developed dashboards and automated checks, to ensure CRON jobs are spaced apart and benchmarking jobs are not abusing high-priority job queues, increasing stability of jobs and saving cloud compute by +1K hours per month

PUBLICATIONS

Hua SBZ, Lu AX, Moses AM. CytoImageNet: A large-scale pretraining dataset for bioimage transfer learning. NeurIPS Workshop on Learning Meaningful Representations for Life. 2021 Dec.

Hua SBZ, Rickard M, Weaver J, Xiang A, Alvarez D, Velear KN, Sheth K, Tasian GE, Lorenzo AJ, Goldenberg A, Erdman L. From Single-Visit to Multi-Visit Image-Based Models: Single-Visit Models are Enough to Predict Obstructive Hydronephrosis. 18th Symposium on Medical Information Processing and Analysis (SIPAIM). 2022 Nov.

CONFERENCE POSTERS

Supervised Contrastive Learning for Improved View Labeling in Pediatric Renal Ultrasound Videos
ISBI 2023 (Cartagena, Colombia)

***Longitudinal Image-Based Prediction of Surgical Intervention in Hydronephrosis Patients: Perhaps Earlier Decision-Making Is Possible!**

ESPU 2023 (Lisbon, Portugal)

*Not presenter

INVITED TALKS

Towards Meaningful Pretraining Data (with Alex Lu and Alexander Lin)
Models, Inference & Algorithms Seminar, Broad Institute
Boston, USA, October 25th, 2023

HONORS & AWARDS

2024 **St. Michael's College In-Course Scholarship**, \$3000

2024 **Hosinec Family Scholarship**, \$3000

2023 **Samuel Beatty Fund Travel Grant**, \$600

2021 **University of Toronto CSB Undergraduate Research Award**, \$4000

2020-22 **Dean's List Award**

2021 **The F. M. Hill Scholarship in Biology**, \$1100