Stanley Z. Hua

Toronto, Canada
(+1) 4379863444 | stanley.z.hua@gmail.com | linkedin | github | website

Education

University of Toronto — Toronto, CA
Honours BS Computer Science Specialist (GPA: 3.84/4) — *Sept. 2022 – June 2024*

University of Toronto — Toronto, CA
Honours BS Bioinformatics & Computational Biology Specialist — *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, psycopg2
ML Libraries: PyTorch, Tensorflow, Scikit-Learn, MLFlow
CV/NLP Libraries: Transformers, Langchain, Guardrails, nltk, OpenCV

ML Research Experience

Data Scientist — June 2024 – Present
The Hospital for Sick Children (Supervisor: Prof. Lauren Erdman) — *Toronto, CA*

- Developing a ML system to classify kidney and bladder anatomical planes in pediatric renal ultrasound sequences to enable community point-of-care ultrasound. Validated on ultrasound image and sequences from SickKids and 3 external institutions.
- (With Michael Chua & Catherine Windrim) Training segmentation models (SAM2-UNet) for two use-cases: a) rare anomaly detection in prenatal ultrasounds (sagittal view), and b) penile region assessment in pediatric hypospadias from optical images.
- (With Irene Chen and Mandy Rickard) Assessing LLMs for intrinsic bias on assessing case urgency in real-world urology referral notes pre/post compression.
- (With Rahul Krishnan, David Pellow and Michael Brudno) Designing a retrieval-augmented LLM system to assist genomic analysts in identifying causal variants for rare diseases
- Presented workshops to the broader SickKids Research Institute on a) accelerating scientific computation with GPUs on the SickKids HPC SLURM cluster, and b) introduction to basic and intermediate SQL.

ML Research Assistant — May 2024 – Present
University of California, Berkeley (Supervisor: Prof. Irene Chen) — *Remote*

- (With Sanae Lotfi) Investigating the components of LLM compression (quantization) that impact fairness.

ML Research Assistant — Jan. 2024 – June 2024
Vector Institute (Supervisor: Prof. Rahul Krishnan) — *Toronto, CA*

- Benchmarked multi-agent prompting methods to improve zero-shot performance of open-source large multi-modal models (Mixtral 8x7B, LLaVA) for general and healthcare-specific MCQA datasets.
- (With Ian Shi and Philip Fradkin) Assisted in evaluation of mRNA foundation model IsoCLR against RNA foundation models (mRNA-FM and CodonBERT) to perform RNA-binding protein prediction

Junior ML Specialist — May 2022 – Aug. 2023
The Hospital for Sick Children (Supervisors: Prof. Lauren Erdman, Alex Lu, Prof. Irene Chen) — *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 — Sept. 2021 – May 2022
The Hospital for Sick Children (Supervisors: Prof. Lauren Erdman, Prof. Anna Goldenberg) — *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 May 2021 – Sept. 2021

University of Toronto (Supervisors: Prof. Alan Moses, Alex Lu) *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 Jul. 2020 – Jul. 2021

University of Toronto (Supervisor: Prof. Pascal Tyrrell) *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.

Engineering Experience

Co-Founder & Engineer July 2024 – Present

Joust *Toronto, CA*

- Led a team of 3 engineers and 2 marketing analysts to develop Joust, the last dating app for serious daters
- Under the Supabase Expo React Native app framework, co-designed the Postgres database schema, user authentication, match-making system and other guardrails to promote genuine interactions

Data Science Intern June 2023 – Sept. 2023

Wealthsimple *Toronto, CA*

- Collaborated with client-facing teams to create a **LLM framework** to perform zero-shot **topic/subtopic classification** of **customer service tickets** for ticket routing, resulting in 15% less errors versus humans.
- Designed, implemented and deployed model to an API endpoint that detects if a user entered invalid account details when initiating an account transfer, preventing a projected $12M in failed 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

Erdman L, Rickard M, Drysdale E, Skreta M, **Hua SBZ**, . . . , Lorenzo AJ, Goldenberg A. The Hydronephrosis Severity Index guides paediatric antenatal hydronephrosis management based on artificial intelligence applied to ultrasound images alone. Scientific Reports. October 2024

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

Machine Learning-Enabled Renal Ultrasound View Labeling to Expand Use of Point-Of-Care Imaging in Community Settings

Nature Conference on Precision Child Health 2024 (Toronto, Canada)

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