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 (PyTorch Lightning), Tensorflow, Scikit-Learn, Comet
CV/NLP Libraries: Transformers, vLLM, Langchain, Guardrails, nltk, OpenCV

ML Research Experience

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

- (With Lauren Erdman & Mandy Rickard) Develop a real-time EfficientNet-based ML system to guide clinicians in acquiring pediatric kidney and bladder ultrasound views to enable community point-of-care ultrasound, retrospectively validated on data from SickKids, Stanford Children's, CHOP and UIowa.
- (With Lauren Erdman) Develop HocusPOCUS, an OOD dataset for point-of-care ultrasound, highlighting the potential for even a simple post-hoc OOD detection method – maximum logit score to jointly identify irrelevant views and choose samples that the model can perform well on. We highlight further areas of improvement in preventing bias propagation and temporal instability of image-based OOD detection.
- (With Lauren Erdman, Alex Lu, Irene Chen, Alex Towbin & Nicholas Heller) Carry out a comprehensive public data review, being the first to show that less than 1% of public data is pediatric. On popular X-ray datasets, we demonstrate that adult trained X-ray models can perform increasingly worse on younger children.
- (With Lauren Erdman & Michael Chua) Trained a SAM2UNet model for penile anatomy segmentation to assist in surgical planning for children with hypospadias.
- (With Lauren Erdman, Tim Van Mieghem & Catherine Windrim) Developing an unsupervised medical anomaly detection algorithm for first trimester fetal ultrasounds using SAM2UNet-based image reconstruction.
- Organized workshops for the broader SickKids Research Institute on: a) deep learning for medical image analysis, b) accelerating scientific computation with GPUs on the SickKids HPC SLURM cluster and c) a two-part SQL tutorial on the Expedition cloud platform.

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

- (With Irene Chen, Sanae Lotfi & Mandy Rickard) Developed a framework to assess the causal impact of quantizing LLMs on social bias in discriminative and generative benchmarks, finding quantization does not worsen social bias on average but leads to behavior flipping at the instance-level. Collaborated with clinicians to further measure social bias in urology referral notes with SickKids REB/IRB approval.

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.
- Performed registration and skull-stripping on brain MRIs of children with vasculitis using FSL

Engineering Experience

Co-Founder Oct 2024 – Present

AltraML *Remote*

- Developing an edge platform for serving real-time AI/ML models for point-of-care ultrasound.
- Performing a literature review to identify real-time use cases of ultrasound AI and gathering potential leads.

Co-Founder & Engineer July 2024 – Oct 2024

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 mission-critical **data extraction tool** to extract FPGA benchmarking results from a SQL database or file system, resulting in improved efficiency, code maintainability and unit-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

In Progress/Review

Hua SBZ, Khondker A, ..., Erdman L. Longitudinal image-based prediction of surgical intervention in infants with hydronephrosis using deep learning: is a single ultrasound enough?. Submitted to PLOS Digital Health (**In Review**).

Hua SBZ, ..., Erdman L. HocusPOCUS: An OOD detection benchmark for point-of-care ultrasound. Submitted to MICCAI 2025 (**In Review**).

Hua SBZ, ..., Chua M, Erdman L. Machine Learning Application for Automatic Segmentation of Hypospadias Anatomy: A proof of concept. Submitted to ESPU 2025 (**In Review**).

Hua SBZ, He P, Towbin A, Heller N, Chen I, Lu A, Erdman L. Leaving Children Behind: The Bias Imposed by Missing Pediatric Data in Open Medical AI Datasets. Pending Submission to NEJM AI 2025 (**Pending Submission**).

Hua SBZ, Lotfi S, Chen I. Understanding Social Bias in Quantized Large Language Models. Pending Submission to NeurIPS 2025 (**Pending Submission**).

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

2020-24 **Dean's List Award**

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

Extracurricular Activities

Teaching Assistant June 2021 – Aug. 2021
University of Toronto – Independent Summer Statistics Community *Toronto, CA*
- Organized data science - related workshops for undergraduate statistics students.
- Presented hands-on workshop introducing the Pandas library.

Club President & Founder Sept. 2021 – Sept. 2022
University of Toronto – Offline *Toronto, CA*
- Led a team to build an online community of 200+ undergraduate students on Discord during the pandemic.
- Hosted multiple game nights and movie nights.

Events Coordinator May 2020 – May 2021
University of Toronto – Table Tennis Club *Toronto, CA*

- Led new initiatives to increase social media presence and spread information on table tennis players and competitions.

General Council Member Sept. 2020 – May 2021

University of Toronto – Bioinformatics & Computational Biology Student Union *Toronto, CA*

- Assisted in planning of online social activities for bioinformatics undergraduate students.

Volunteer Chess Teacher Aug. 2020 – Aug. 2020

Kensington Health – Second Mile Club *Toronto, CA*

- Organized, promoted and hosted beginner chess workshops for seniors.

Volunteer Photographer Feb. 2020 – Feb. 2020

West Neighborhood House *Toronto, CA*

- Took photographs and videos to help promote the not-for-profit's community services.