

## Education

---

- University of Toronto**, MSc in Computer Science Aug. 2019 - June 2022  
Machine Learning Group and DGP Lab  
GPA 4.0/4.0
- University of Toronto**, B.Sc. in Computer Science and Statistics Aug. 2015 - June 2019  
Dean's List Award (2016 - 2018)  
GPA 3.97/4.0 in advanced courses.

## Internship

---

- Research & Development Intern - Canada Pension Plan Investment Board** Jan. 2022 - Present
- Design & built NLP models for a national pension fund to analyze text transcripts between company management and media to predict future company metrics using underlying sentiments (PyTorch, Python)
    - Restructured native BERT model to accommodate for the very long financial text transcripts which cannot be handled by BERT directly.
- Research & Development Intern - Samsung Toronto AI Research** May 2018 - Dec. 2018
- Participated in the research & implementation of VASTA, a smartphone task automation system (Tensorflow, Python, Java)
    - Computer Vision: used object detection & optical character recognition to recognize UI elements on the cellphone screens in a way that is robust to positional and visual changes
    - NLP: used clustering model for user audio recognition and param prediction model to determine the automation parameters
    - Python TensorFlow: used GPU and CUDA to vectorize the image data (synthetic data we generated manually for our specific needs) to re-train YOLO and Retina Net models for a faster speed
  - Real-time video instance segmentation system on android devices (Java, Python)
    - Independently implemented the system using Mask RCNN model
    - Android Frontend (Java): collecting real-time video data, compressing & sending data using JSON stringification by HTTP POST (0.5s per video frame), and receiving data from Mask RCNN model from the server
    - Backend (Python): receiving video frames, using GPU clusters to compute the instance segmentation, offloading the computing burden from the android side with limited computing power to the GPU server
- Software Development Intern - Hospital for Sick Children** Jan. 2020 - Present
- Implemented an interactive time-efficient labeling framework for doctors to label biological waveform data used for AI models using semi-supervised learning with WaveNet (Python)
    - Used Matplotlib for data rendering to avoid potential out of memory issues in case of large data flows
    - Used by our clinical staff to label more than 20 hours of ECG data to train our AI models
    - Made a distributable package for users to install on their local machine
  - Led the design & development of a dynamic online medical data visualization & management system called *In-Sight* (fullstack with Highcharts.js, Node.js and MongoDB)
    - Frontend: customized highcharts.js source code to fit project's needs, e.g. synchronous zooming of multiple aligned charts

- Backend: saved user profiles/annotations, render annotations with Node.js, express server, and MongoDB
- Increased the doctors' speed of the morning rounds by about 30%
- Led the design & development of a website called *RESUSVis* used by the hospital's monthly cardiac review meeting to display the heterogeneous information in a more organized way (fullstack with Highcharts.js, Node.js and MongoDB)
  - Frontend: developed our own libraries to feed normalized/structured data to highcharts.js and the server
  - Backend: Node.js, express server, and MongoDB
  - The website is preferred by 90% of the users compared with their original way of the meeting

**Research Assistant - Computer Science - Vector Institute** Aug. 2018 - Present

- Used Javascript and C3.js to implement an intuitive visualization of medical AI models to improve model explainability, leading to better user trust in the AI models (Javascript, C3.js)
- Coded Flipout methods into our current AI model to improve the training efficiency of Bayesian neural networks (Python, Pytorch)

**Student Developer - University of Toronto** Apr. 2017 - Dec. 2017

- Developed the backend of our course management system called *Courseography* (Haskell)
  - Learned to use Haskell in two weeks to accommodate for the peculiarity of the system which does not support common languages
  - Used by about 1000 computer science students at the University of Toronto

## Projects

---

**Product recommender system based on Amazon reviews** (Python) Sept. 2020 - Dec. 2020

- Built a complete NLP system from scratch using BERT
  - Compared BERT to common NLP frameworks like tf-idf, LSTM, etc
  - With a final accuracy of 74.25% in predicting users' ratings for the products

## Languages and Technologies

---

**Programming Languages:** Python; Java; C++; SQL; VBA; Shell; Haskell;

Bootstrap; Node.js; React.js; jQuery; Vue; D3.js;

Tensorflow; Pytorch; Scikit-learn; R; MATLAB; OpenCV;

**Others:** Git; Linux; Jupyter Notebooks; GSuite; MongoDB; Adobe Creative Suite; AWS;

## Publications

---

**Minfan Zhang\***, Daniel Ehrmann\*, Mjaye Mazwi, Danny Eytan, Marzyeh Ghassemi, Fanny Chevalier. 2022. Get To The Point! Problem-Based Curated Data Views To Augment Care For Critically Ill Patients. In *Proceedings of CHI '22: ACM CHI Conference on Human Factors in Computing Systems*

Alborz Rezazadeh Sereshkeh, Gary Leung, Krish Perumal, Caleb Phillips, **Minfan Zhang**, Afsaneh Fazly, Iqbal Mohamed. 2019. VASTA: A Vision and Language-assisted Smartphone Task Automation System. In *Proceedings of IUI '20: ACM International Conference on Intelligent User Interfaces*.

Danny Eytan, Dmitrii Shubin, **Minfan Zhang**, Daniel Ehrmann, Sebastian Goodfellow. 2021. Time-efficient labeling framework for biological waveform data using semi-supervised learning and interactive visualization. In *MLHC '21: Machine Learning for Healthcare*.