Felix Zhang

felixfzhang@cs.toronto.edu | ff-zhang | felixfzhang

Education

University of Toronto Sept. 2025 – June 2027
Masters of Science in Computer Science

University of Toronto Sept. 2021 – June 2025
Honours Bachelor of Science in Computer Science; Major in Mathematics *3.96/4.0 cGPA*

Publications

PD3: Prefetching Data with DPUs for Disaggregated Memory May 2026
Sidharth Sankhe, Felix Zhang, Umayrah Chonee, Sherman Lim, Jason Hu, Jialin Li, Qizhen Zhang
23rd USENIX Symposium on Networked Systems Design and Implementation

Research Experience

Research Assistant; Far Data Lab, University of Toronto Sept. 2024 – Present
Supervisor: Prof. Qizhen Zhang

- Investigated offloading computations onto SmartNICs and data processing unit (DPUs)
- Parallelized the execution of *Monodepth2* in **Python** and **C++** with the latter achieving linear performance scaling with the number of threads
- Built a DPU-based prefetcher *PD3* with a team of 6 which intercepts network traffic to predict and prefetch data for tiered key-value stores
- Designed an external service for offloading shuffle operations from database management systems which supports either a memory or storage backend managed by DPUs

Research Assistant; University of Toronto March 2024 – Dec. 2024
Supervisor: Prof. Jack Sun

- Worked on a team of 11 to implement a pedagogical kernel *KidneyOS* in **Rust** to be used in an introductory operating systems course with **500+** students annually
- Enabled thread creation and destruction, multi-threading, pre-emptive scheduling within the thread system
- Led a team of 3 to implement POSIX-compatible syscalls and add support running user-space executables

Research Assistant; PRISM Lab, Bloorview Research Institute June 2024 – Aug. 2024
Supervisors: Erica Floreani and Prof. Tom Chau; Funded by: FUSRP

- Curated deep-learning models from the literature on denoising electroencephalogram (EEG) data in a team of 4 and benchmarked them on the *EEGDenoiseNet* dataset
- Investigated the applicability of end-to-end transformer models to denoise EEG signals and the impact of using signals' time-frequency representation as input on model performance

Research Assistant; Biological Physics Group, University of Toronto May 2023 – May 2024
Supervisor: Prof. Anton Zilman; Funded by: Work Study Program

- Implemented a data pre-processing pipeline which processes raw cytokine data and extracts integral features
- Built a feed-forward network in **PyTorch** that predicts the cytokine dynamics of T cells in response to antigens
- On a team of 4, showed two variables are sufficient to determine cytokine concentrations because our model predicted the correct output concentration with **0.01%** error using a bottleneck layer with 2 neurons

Research Assistant; Physics Education Group, University of Toronto May 2022 – Sept. 2022
Supervisor: Prof. Carolyn Sealfon

- Created a dataset of ∼**11 000** sentences from student feedback which labels whether they contain suggestions
- Compared the effectiveness of statistical and deep-learning classifiers at identifying suggestions using **scikit-learn** and **TensorFlow** respectively
- Demonstrated the efficacy of a BERT classifier at addressing this problem with it achieving an F_1 score of **0.823**

Industry Experience

ML Runtime Engineer; Cerebras Systems May 2024 – Present

- Implemented a runtime virtual memory system in **C++** with a team of 3 which pre-emptively loads data before it is accessed, allowing off-chip memory to be used for the first time with only a **10%** performance penalty
- Added support for network storage in the paging system with remote direct memory access, providing **100** GB/s read and write speeds with **10** μs latency to multiple remote servers
- Enabled the ability log and replay the network operations, decreasing the time to recreate stalls and timeouts by over **80%**, and setup unit tests to automatically catch breakages and performance regressions in the network layer
- Improved the throughput of the network layer by **6%** when transferring data by implementing best practices for remote direct memory access and reducing setup overhead
- Determined the cable and port mapping for a one node cluster which will be used for all future deployments

Awards & Scholarships

Fields Undergraduate Summer Research Program ($3 800), Fields Institute	June – Aug. 2024
Louis Savlov Scholarship in Sciences And Humanities ($1 000), University of Toronto	Nov. 2023 – Jan. 2025
Dean's List Scholar, University of Toronto	Jan. 2022 – June 2025
Second Malcom Wallace Scholarship ($5 000), University of Toronto	Sept. 2021 – Oct. 2024
University of Toronto Scholar ($7 500), University of Toronto	Sept. 2021
B.C. Achievement Scholarship ($1 250), Government of British Columbia	Aug. 2021
District/Authority Scholarship ($1 250), Government of British Columbia	Aug. 2021

Student Leadership

Director of Internal Relations; Computer Science Student Union, University of Toronto Apr. 2023 – Apr. 2024

- Organized orientation for the ~**500** undergraduate students entering the computer science stream
- Planned **20+** events in collaboration with various partners in industry (such as AMD and Google) or student organizations (such as UTMIST ↗ and WiCS ↗)
- Hosted **5+** talks with professors in the Department of Computer Science at the University of Toronto

First-Year Academic Officer; Math Union, University of Toronto Sept. 2021 – Apr. 2022

- Facilitated discussions between **20** mentor-mentee pairs in the *First-Year Mentorship Program* by providing guidance to the upper-year mentors
- Organized "Coffee and Chat" events which allowed for informal discussions between students and math professors

Registered Study Group Leader; Sidney Smith Commons, University of Toronto Sept. 2021 –April 2022

- Led study groups for *Foundations of Computer Science I* and *Enriched Introduction to the Theory of Computation*
- Headed weekly meetings for first-year students that reviewed content covered in the previous week's lecture
- Developed example problems to clarify and reinforce important concepts through group discussion

Projects

Student Response Classifier Mar. 2023 – Apr. 2023

- Developed a 3-parameter logistic item response theory classifier in **PyTorch**, using alternating gradient descent for training, to predict the correctness of student answers to multiple-choice questions
- Obtained an accuracy of **72%** on the *NeurIPS 2020 Education Challenge* dataset (within 5% of the best solution)

Image Classifier ↗ Dec. 2022 – Jan. 2023

- Implemented a softmax classifier with stochastic gradient descent (SGD) from scratch in **C++** using only the linear algebra library **Eigen3**
- Achieved **92%** accuracy on the *MNIST* dataset of handwritten digits (within 2% of the top classifier using SGD)
- Built in the ability to save trained weights, perform batch training, and track the training error in real-time

Image Restoration with Convolutional Neural Networks ↗ Sept. 2020 – June 2021

- Combined the models RIDNet and DeepDeblur using **PyTorch** to determine the ability of convolutional neural networks to deblur and denoise images
- Artificially generated a dataset of **5 000** noisy, blurred images using a Poisson-Gaussian noise model
- Discovered that integrating the two models offers marginal improvements over their individual performance

Technical Skills

Languages	C++, Rust, Python, C, C#, Bash, Java
Frameworks	Catch2, PyTest, PyTorch, TensorFlow, OpenCV, scikit-learn, NumPy, Pandas, SciPy, Eigen3
Tools	Git, Linux, Unix, tmux, CMake, Docker, Anaconda, Google Colab, QEMU, Jupyter, WSL, Slurm

Selected Coursework

Code	Title	Term
CSC2306*	High Performance Scientific Computing	Winter 2025
CSC2525*	Research Topics in Database Management	Winter 2025
CSC2234†	Database System Technology	Fall 2025
CSC2235*	Cloud-native Data Management Systems	Fall 2025
CSC2221*	Introduction to the Theory of Distributed Computing	Fall 2024
CSC324	Principles of Programming Languages	Winter 2024
CSC412†	Probabilistic Learning and Reasoning	Winter 2024
CSC413†	Neural Networks and Deep Learning	Winter 2024
CSC473	Advanced Algorithm Design	Winter 2024
MAT357	Real Analysis I	Winter 2024
APM462	Nonlinear Optimization	Fall 2023
CSC369	Operating Systems	Fall 2023
CSC420	Introduction to Image Understanding	Fall 2023
MAT354	Complex Analysis I	Fall 2023
MAT377	Mathematical Probability	Fall 2023
MAT327	Introduction to Topology	Summer 2023
CSC373	Algorithm Design, Analysis and Complexity	Winter 2023
CSC384	Introduction to Artificial Intelligence	Winter 2023
CSC438	Computability and Logic	Winter 2023
CSC463	Computational Complexity and Computability	Fall 2022
MAT344	Introduction to Combinatorics	Summer 2022

*Graduate course

†Cross-listed graduate course