

# Yu Bo Gao

ybgao@cs.toronto.edu

## EDUCATION

---

**University of Toronto** Sept 2022 -  
Ph.D. Student, Computer Science GPA 4.0

**University of Toronto** Sept 2018 - May 2022  
Honours Bachelor of Science GPA 4.00  
Computer Science Specialist, Mathematics Major

## EXPERIENCE

---

**Research Engineer**, CentML Inc. *October 2022 - Present*

- CentML accelerates Machine Learning workloads by optimizing models to utilize hardware accelerators, like GPUs or TPUs, more efficiently and without affecting model accuracy.
- Lead the research and development of CentML DeepView, an interactive profiler for deep learning training workloads with predictive capabilities.

**Research Assistant**, University of Toronto *May 2020 - August 2022*

- Supervised by Prof. Gennady Pekhimenko, funded by UTEA.
- Produced GPU profiles and performance analysis for neural networks which measured resource utilization and memory breakdown.
- Collaborated on the Habitat project with performance modelling for the USENIX ATC' 21 paper.

**Research Assistant**, University of Toronto *Sept 2021 - April 2022*

- Supervised by Prof. Maryam Mehri Dehnavi as part of an undergraduate research course.
- Worked with graduate students, studied and evaluated existing GPU kernels for sparse matrix multiplication (SpMM).

**Software Engineering Intern**, Amazon Web Services *May 2021 - Aug 2021*

- Worked at AWS Neuron on performance modelling for Amazon's machine learning accelerator.

**ML Software Developer**, Lexivalley Inc. *Jun 2019 - May 2020*

- Implemented depth-sensing model with TensorFlow after reading related literature.
- Adapted the model to a different environment by programmatically producing a synthetic dataset with Blender.

**Team Member**, aUToronto (University of Toronto Autodrive Team) *Feb 2019 - Jun 2019*

- Member of the mapping and localization subteam.
- Developed software for systematically detecting and adding features for multi-lane traffic maps including different types of intersections, stop lines, etc.

## PUBLICATIONS

---

Yu Bo Gao, Maryam Haghifam, Christina Giannoula, Renbo Tu, Gennady Pekhimenko, Nandita Vijaykumar  
*Proteus: Preserving Model Confidentiality during Graph Optimizations*  
The Seventh Annual Conference on Machine Learning and Systems (MLSys24). May 2024.

Geoffrey X. Yu, Yu Bo Gao, Pavel Golikov, Gennady Pekhimenko  
*A Runtime-Based Computational Performance Predictor for Deep Neural Network Training*  
USENIX Annual Technical Conference (ATC21). July 2021.

## ACADEMIC ACTIVITIES

---

- Conference Reviewer**, MLSys 2025 *Spring 2025*
- Teaching Assistant**, University of Toronto  
CSC263 - Data Structures and Analysis *Winter 2020, 2021*  
Held office hours before assessments, graded problem sets and exams.
- Volunteer Note-taker**, University of Toronto Accessibility Services *Sept 2018 - Jun 2019*
- Bayview Competitive Programming Club** *Sept 2016 - June 2018*  
Founded the club with the goal of teaching topics in algorithms and data structures used in competitive programming, and to prepare the school team for regional competitions.

## AWARDS AND SCHOLARSHIPS

---

- Ontario Graduate Scholarship**, University of Toronto *June 2024*  
The Queen Elizabeth II Graduate Scholarship in Science and Technology (QEII—GSST) program is designed to encourage excellence in graduate studies in science and technology.
- Canadian Graduate Scholarship – Masters**, University of Toronto *May 2023*  
The objective of the Canada Graduate Scholarships-Masters (CGS M) Program is to help develop research skills and assist in the training of highly qualified personnel by supporting students who demonstrate a high standard of achievement in undergraduate and early graduate studies.
- Wolfond Scholarship in Wireless Information Technology**, University of Toronto *2022–2023*  
Awarded to graduate students who are pursuing research in areas related to systems, wireless, networks, HCI and digital media. Awards to be given based on academic merit.
- McNab Undergraduate In-Course Scholarship**, University of Toronto *Fall 2022*  
Recognizes academic achievement.
- University of Toronto Excellence Award**, University of Toronto *Summer 2020*  
Funds undergraduate students with opportunities to conduct summer research projects with a professor.
- Dorothy Walters Scholarship**, University of Toronto *2019, 2020*  
This scholarship is awarded to outstanding students with a minimum cumulative grade point average of at least 3.50.
- Dean's List Scholar**, University of Toronto *2018 – 2022*  
Given to degree students in the Faculty having a Cumulative Grade Point Average of 3.50 or higher.
- Principal's Admission Scholarship**, University of Toronto *2018*  
Awarded during admission to the university.
- Bronze Medal**, Canadian Computing Olympiad, University of Waterloo *2016, 2017*  
Ranked top-25 in the Canadian Computing Competition amongst Canadian high school students.

## SELECTED PROJECTS

---

- Efficient Sparse Matrix Products for TPUs**, CSC2224 *Apr 2023*  
Accelerates TPU sparse matrix-vector products by up to  $8\times$  in software using a combination of (a) diagonal extraction and (b) block extraction. Different from hardware implementations of SpMV, this work functions during compile-time and works with TPU v3/4.
- Rewind**, UoTHacks VIII *Feb 2021*  
An intelligent, collaborative and interactive web canvas with built in voice chat that maintains a list of

live-updated keywords that summarize the voice chat history. Featured in The Varsity, the University of Toronto student newspaper [here](#). 4'th place winner; Best use of Google Cloud.

**Memoritis**, Hack The North 2019

*Sept 2019*

Created a platform that analyzes topics of user-uploaded videos and forms a graph between similar videos. Uses word2vec and the Google Video Intelligence API.

**Distributed Compiler Collection**, ETHUofT 2019

*Mar 2019*

Created a platform using Blockchain to establish trust from source code to compiled binaries by distributing the verification process.

**Circular**, MHacks X (University of Michigan)

*Sept 2017*

Implemented live recognition and simulation of hand-drawn circuits from webcam input with OpenCV.

**IdeaShare**, Hack The North 2016

*Sept 2016*

Implemented web-based idea sharing platform using NLP and graph algorithms. Responsible for implementing the NLP logic (with TextRazor NLP API) and the bipartite matching algorithm. Awarded top-12 winners.

## SKILLS

---

### Programming Languages

Proficient: Python, Java, C,  $\LaTeX$

Intermediate: JavaScript/node.js, C++, C#, CUDA, Verilog

Introductory: Haskell

### Frameworks, Tools and APIs

Python: PyTorch, TensorFlow, Keras, OpenCV, Pandas, Matplotlib

Other: LLVM

### Other

Proficient in UNIX-like operating systems, including GNU/Linux.