Yufei (Celina) Chen

ycelina.chen@mail.utoronto.ca
Website: celina-chen.github.io/
LinkedIn: linkedin.com/in/celina-chen-477301251
Phone: (825) 288-9098

Education

University of Toronto, Toronto, ON *Sept 2023 – May 2026 (expected)*
B.Sc. Honours, Computer Science Specialist
Relevant Coursework: Machine Learning, Deep Learning, Probability, Visual Computing, Algorithms, Analysis, Algebra, Multivariable Calculus

Awards

- **Schulich Leader Scholarship** (2023)
  Canada's most prestigious undergraduate STEM scholarship; awarded to 100 students nationwide across 20 partner universities.
  Value: $100,000.
- **Alberta High School Mathematics Competition Part II — 13th Place** (2023)
  Provincial invitation-only mathematics contest.
- **Canadian Senior Mathematics Contest — Top 1.2% Internationally** (2022)
  International mathematics contest administered by the University of Waterloo.

Research Experience

Student Researcher (Differential Privacy & Unlearning) *May 2025 – Present*
Vector Institute / CleverHans Lab (Supervisor: Prof. Nicolas Papernot)

Research Direction 1: Differential Privacy (DP) & Data-Mixing for Private Fine-Tuning

- Designing and evaluating **DP data-mixing objectives** to improve private fine-tuning performance in both vision and language models.
- Developed mixing strategies that **reduced validation loss** and **increased task precision** on benchmark datasets including **CIFAR-10** and **AIMI medical imaging**.
- Implementing DP training pipelines using **PyTorch, JAX, and Opacus** for low-data and high-sensitivity scenarios.
- Leading experimental work for a first-author manuscript in preparation for NeuRIPs 2026.

Research Direction 2: Exact Machine Unlearning

- Investigating exact unlearning in settings where data is scarce, long-tailed, or unevenly distributed, with a focus on ensuring models behave as if specific data points were never used.
- Examining how unlearning pipelines such as SISA, which rely on sharding or distillation, can underrepresent rare or low-frequency examples and destabilize updates in tail regions.
- Designing approaches that apply controlled data mixing and **learned auxiliary reweighting** to recover tail information during unlearning while maintaining computational efficiency.
- Contributing to method development, experimental evaluation, and analysis for paper to a top security conference.

Undergraduate Researcher (Neural Representation & Interpretability) *Sept 2024 – Apr 2025*
Bernhardt-Walther Lab, University of Toronto

- Investigated mid-level feature selectivity in CNNs, analyzing biological parallels in visual tuning.
- Analyzed first-layer feature maps of CNNs and showed that filters specialize in detecting different edge orientations and spatial frequencies, paralleling orientation-selective responses in early visual cortex.
- Produced internal research report; presented results in lab meetings and end of the year ROP Poster Fair

a) Research in Progress

Optimizing Public Data Mixtures for Differentially Private Learning
Y. Chen*, A. Thudi, N. Papernot.
Ongoing research on data-mixing methods to improve private learning under differential privacy constraints.

Improving Exact Machine Unlearning with Tail-Preserving Data Mixing
N. Jia, Y. Chen*, N. Papernot, et al.
Ongoing collaborative research on stabilizing exact unlearning under long-tailed and low-data regimes.

b) Non-Peer-Reviewed Work (Preprints / Technical Reports)

Neural Network Extraction Report — Link to Report
Internal technical report, Bernhardt-Walther Lab (2025).

c) Research Presentations

Poster: Neural Network Alignment with Mid-Level Vision —Link to Poster
Research Opportunities Program (ROP) Showcase, University of Toronto, Apr 2025. *Lightning talk and poster session.*

Academic & Technical Projects

Representation Geometry in Model Merging Link to Report *Fall 2025*
Course: Deep Learning: Theory & Data Science

- Studied model merging through the lens of feature geometry, motivated by Linear Mode Connectivity, superposition, and recent work on layerwise feature connectivity.
- Designed a synthetic autoencoder-based reconstruction task where two models are trained with different feature-importance masks, forcing them to specialize on different subsets of input dimensions.
- Proposed a feature-space-aligned merging (FSA-Merge) method that constructs shared layerwise subspaces via SVD on activations from both models and trains a merged model to stay within these shared subspaces using a projection loss (with optional task loss).
- Empirically compared naive weight interpolation, plain fine-tuning, and FSA-Merge, showing that the proposed method achieves lower reconstruction loss and faster convergence in challenging cases where source models diverge in their learned feature subspaces.

Personalized Student Performance Prediction — Link to Github *November 2024*

- Designed an advanced **autoencoder** with **PCA**-based dimensionality reduction, skip connections, and dropout, improving prediction accuracy on sparse student performance datasets.
- Optimized training pipeline, enhancing **model robustness and convergence speed**.
- Analyzed large-scale student performance data to identify key patterns and compiled findings into a report comparing model performances with actionable improvements.

ListenTTC — Link to Github *November 2023*

- Built a full-stack Flask web application for real-time audio monitoring on Toronto public transit routes.
- Implemented live speech-to-text transcription using **OpenAI Whisper**, custom audio processing, and the Multimedia Front End API.
- Integrated sentiment analysis with a **Hugging Face** model and automated high-risk incident alerts via the Twilio API.

Skills

Programming: Python, C, Java, R
ML Frameworks: PyTorch, TensorFlow, JAX, Opacus, Scikit-Learn
Tools: Git, LaTeX, Jupyter, OpenCV, NumPy, Pandas
Mathematical Foundations: Linear Algebra, Probability, Optimization, Statistics