Gavin Pu

gpu@cs.toronto.edu (339) 927-8425
https://www.linkedin.com/in/gavin-pu/
https://www.cs.toronto.edu/~gpu/

PROFESSIONAL PROFILE

- MSc in Applied Computing candidate with a cross-disciplinary background in computer science, AI and data science, and bioinformatics
- Experienced in developing scalable ML pipelines, architecting high-performance database system infrastructure, and performing applied research in ethical AI and computational biology

EDUCATION

MSc in Applied Computing Sep 2025 – Dec 2026
University of Toronto, Department of Computer Science
Concentration in Computer Science
- Courses: Deep Learning: Theory and Data Science, Software Engineering for ML, Ethical Aspects of AI, Database System Technology

Honors BSc Sep 2021 – Apr 2025
University of Toronto, Faculty of Arts and Science
Computer Science Specialist, Bioinformatics and Computational Biology Specialist, Statistics Minor
- Courses: Neural Networks, Statistical ML, Algorithm Design, Computer Networks, Operating Systems, Applied Bioinformatics

EXPERIENCE

Lin Lab, Department of Cell and Systems Biology, University of Toronto May 2023 – Present
Neural Data Engineer
- Design ML pipelines incorporating clustering and dimensionality reduction to visualize neural activity from imaging data
- Automate non-parametric hypothesis tests across 30 movement variables to determine effects of nanoparticle exposure
- Optimize movement trajectory calculations by vectorizing functions, reducing compute time from minutes to seconds

Department of Computer Science, University of Toronto Sep 2025 – Present
Teaching Assistant
- Organize weekly 3-hour hardware labs to evaluate student skills in logic design, communication, and presentation
- Respond to student concerns in real time on the online Q&A platform Piazza

Department of Statistical Sciences, University of Toronto Sep 2024 – Present
Teaching Assistant
- Lead classes of up to 60 first- and second-year undergraduates on Python, R, Jupyter Notebook, and LLMs
- Conduct weekly tutorials and office hours to foster independent problem solving

PROJECTS

Stable Diffusion Model Collapse Nov 2025 – Dec 2025
- Evaluated the impact of recursive fine-tuning on Stable Diffusion by iteratively training 20 generations of models on synthetic outputs
- Orchestrated multi-generational training pipelines on NVIDIA A100 GPUs via Google Colab
- Quantified drift of generated tensors by calculating pairwise distances using a three-dimensional generalization of the Frobenius norm

Key-Value Store Sep 2025 – Dec 2025
- Architected a persistent key-value store in C++ optimized to handle 1GB+ datasets with high-throughput point queries and ranged scans
- Utilized an LSM-tree design with B-tree-based storage layers for SSTs and support for Dostoevsky-based leveling and tiering policies
- Engineered advanced compaction and bulk-loading features, including a min-heap-based external sort-merge algorithm

PUBLICATIONS

* Indicates co-first authorship.

Farooqi, D.*, **Pu, G.***, Paudel, S., Sultana, S., & Ahmed, S.I. Job-Anxiety in Post-Secondary Computer Science Students Caused by Artificial Intelligence. arXiv preprint (2026).

Shen, P.Y., Wu, J., **Pu, G.**, Huang, K., & Lin, Q. Altered locomotion and anxiety after exposure to SiO_2 nanoparticles in larval zebrafish. *Scientific Reports* 15, 18229 (2025).

TECHNICAL SKILLS

- **Machine Learning:** PyTorch, Hugging Face, scikit-learn
- **Languages:** Python, C/C++, Java, JavaScript, HTML/CSS, Shell, SQL, R, LaTeX
- **Libraries and Frameworks:** NumPy, Matplotlib, pandas, SciPy, React, JavaFX, Django