

Alston Lo

 [alstonlo.github.io](https://github.com/alstonlo) |  [alstonlo](https://github.com/alstonlo) |  [alston-lo](https://www.linkedin.com/in/alston-lo)

 alston.lo@mail.utoronto.ca

EDUCATION

Honours Bachelor of Science, *University of Toronto (UofT)*
Math and Computer Science | cGPA: 4.00/4.00, Average: 96%

Sep. 2019 – Jun. 2024

PUBLICATIONS

- [1] **A. Lo**, R. Pollice, A. Nigam, A. D. White, M. Krenn, and A. Aspuru-Guzik, “Recent advances in the self-referencing embedded strings (SELFIES) library,” *Digital Discovery*, vol. 2, pp. 897–908, 4 2023. DOI: [10.1039/D3DD00044C](https://doi.org/10.1039/D3DD00044C).
- [2] J. Bae, N. H. Ng, **A. Lo**, M. Ghassemi, and R. B. Grosse, “If influence functions are the answer, then what is the question?” In *Advances in Neural Information Processing Systems*, 2022. [Online]. Available: <https://openreview.net/forum?id=hzbguA9zMJ>.
- [3] M. Krenn *et al.*, “SELFIES and the future of molecular string representations,” *Patterns*, vol. 3, no. 10, p. 100588, 2022. DOI: [10.1016/j.patter.2022.100588](https://doi.org/10.1016/j.patter.2022.100588).
- [4] W. D. Jeong*, S. D. Jeong*, K. R. Wang*, and **A. Lo***, “siRNA-mediated silencing of antifungal resistance genes: A research protocol,” *Undergraduate Research in Natural and Clinical Science and Technology Journal*, vol. 5, no. 8, pp. 1–7, 2021. DOI: [10.26685/urncst.292](https://doi.org/10.26685/urncst.292).

PREPRINTS

- [1] A. Cheng*, **A. Lo***, S. Miret, B. Pate, and A. Aspuru-Guzik, “Reflection-equivariant diffusion for 3D structure determination from isotopologue rotational spectra in natural abundance,” 2023. arXiv: [2310.11609](https://arxiv.org/abs/2310.11609),
presented at MoML 2023, Accelerate 2023, and NeurIPS AI4Mat 2023 (spotlight, upcoming).
- [2] A. Celaj *et al.*, “An RNA foundation model enables discovery of disease mechanisms and candidate therapeutics,” *bioRxiv*, 2023. DOI: [10.1101/2023.09.20.558508](https://doi.org/10.1101/2023.09.20.558508),

presented at Oligonucleotide Therapeutics Society 2023 Annual Meeting.

* Equal Contribution

AWARDS & HONORS

| | |
|---|-----------------------|
| Gold Medal, Best Model Award, and Nominations for Best Climate Crisis Project, Wiki, Integrated Human Practices, Entrepreneurship, and Presentation , iGEM | Oct. 2023 |
| Qualcomm’s Most Popular Project , EigenAI Conference, UTMIST | Sep. 2023 |
| The James Scott Scholarship , UofT (\$500) | Feb. 2023 |
| Dean’s List Scholar , UofT | Dec. 2020 – Dec. 2022 |
| Gold Medal and Nomination for Best Conservation Project , iGEM | Oct. 2022 |
| The Dickinson-Cartwright 3t0 Scholarship , UofT (\$500) | Nov. 2021 |
| UofT In-Course Scholarships , UofT (\$1500) | Sep. 2021 |
| Undergraduate Student Research Award , NSERC (\$7500) | May 2021 |
| The Honourable Ray Lawson Scholarships , UofT (\$500) | Dec. 2020 |
| The Joseph Alfred Whealy Scholarship , UofT (\$1500) | Sep. 2020 |
| 2nd Place Research Proposal , Scinapse Undergraduate Case Competition | May 2020 |
| The Elliott Scholarships , UofT (\$2000) | Sep. 2019 |
| UofT Scholars Program Scholarship , UofT (\$7500) | Sep. 2019 |
| 7th/3135 Place, Scholar with Distinction , UofT National Biology Competition (\$3000 scholarship) | Apr. 2019 |

EXPERIENCE

Research Intern, *Deep Genomics, Toronto*
Full-time until May 2023, part-time afterwards

May 2022 – Present

- Leveraged state-space models on long-range genomics tasks, achieving competitive performance to previous internal models, with only 17% as many parameters and 2.5× faster training speed
- Designed and implemented a tool that interfaces internal datasets with an accessible and reproducible API for training, supporting multiple ML frameworks

- Developed large genomics language models (200M+ parameters) in PyTorch and TensorFlow, for sequences with >100K tokens, and conducted distributed training on TPU pods and GPU clusters

Machine Learning Research Assistant, *The Matter Lab, UofT* May 2020 – Present
Supervisors: Austin Cheng, Mario Krenn, and Prof. Alán Aspuru-Guzik

- Main developer for v1 onwards of their **selfies** library (see publications and projects)
- Trained variational autoencoders and applied disentanglement and dimensionality reduction techniques to investigate whether scientific insight can be extracted from the structured latent representations of molecules
- Formulated and researched equivariant diffusion models for *de novo* 3D-structure determination of molecules from their rotational spectra and crystal structure elucidation

Machine Learning Research Intern, *Vector Institute, UofT* May 2021 – Jan. 2023
Supervisors: Juhan Bae and Prof. Roger Grosse, NSERC-funded

- Designed and implemented novel objectives for the efficient and scalable approximation of influence functions, which measure how training point(s) impact a deep model's predictions
- Assisted with empirical and theoretical analyses providing novel insights into influence functions

Researcher, *Machine Intelligence Student Team, UofT*

Project co-director | [poster](#) (presented at EigenAI conference) | [samples](#) Sep. 2022 – Apr. 2023

- Developed a hybrid state-space Transformer diffusion model to compose classical piano performances, and trained the model on the MAESTRO dataset using mixed-precision and DDP parallelism

Project developer Sep. 2023 – Present

- Formulating empirical analyses and improvements upon hashing approaches in neural graphics primitives

Dry Lab Researcher, *iGEM Toronto Club, UofT*

- All work is done in a collaborative student-led effort and is presented at the International Genetically Engineered Machine (iGEM) competition

Team member | [homepage](#) Apr. 2023 – Oct. 2023

- Constructed genome-scale metabolic models of methylotrophic bacteria and formulated flux balance analysis algorithms to discover gene knockout, overexpression, and addition targets that improve methanol consumption

Team member | [homepage](#) Apr. 2022 – Oct. 2022

- Designed LAMP primers *in silico* for the rapid cell-free diagnosis of oak wilt fungal pathogen, using primer discovery and analysis tools such as PrimerExplorer, NEB Primer Design Tools, GLAPD, and Primer3
- Analyzed 48 oak wilt fungi genomic sequences compiled from the GenBank NCBI database to find regions of the oak wilt genome amenable for LAMP primer design

Team member (2020) and lead (2021) Apr. 2020 – Jan. 2022

- Researched sequence- and structured-based ML techniques to design novel and improved plastic-degrading enzymes, with respect to thermostability and catalytic activity

PROJECTS

torch-influence (50+ ★, 5 🗨️) Jul. 2022

- An open-source library that implements influence functions in PyTorch
- Designed a simple and minimal API that researchers can easily adapt into existing workflows, while also fixing a few mathematical errors found in existing implementations of influence functions

selfies>=1.0.0 (550+ ★, 110+ 🗨️) Jun. 2020 – Aug. 2022

- An open-source library that implements SELF-ReferencIng Embedded Strings (SELFIES), which is a 100% robust string representation framework for molecules. The library has accrued over 300 stars since my joining
- Generalized v0 to a larger class of molecules and user-defined constraints
- Boosted conversion efficiency between SELFIES and SMILES by over 10-fold by formulating and implementing a novel translation algorithm
- Enhanced library quality-of-life by writing comprehensive documentation and tutorials hosted on ReadTheDocs, devising exhaustive unit-tests, and setting up CI workflows using tox

SELECTED COURSEWORK

MGY441, Bioinformatics Dec. 2023

MAT347, Groups, Rings and Fields Dec. 2023

| | |
|---|-----------|
| CSC412 [†] , Probabilistic Machine Learning | Apr. 2022 |
| CSC413 [†] , Neural Networks and Deep Learning | Apr. 2022 |
| CSC420 , Introduction to Image Understanding | Apr. 2022 |
| APM462 , Nonlinear Optimization | Apr. 2022 |
| MAT357 , Foundations of Real Analysis | Apr. 2022 |
| MAT1841 [*] , Mathematics of Massive Data Analysis: Fundamentals and Applications | Dec. 2021 |
| MAT327 , Introduction to Topology | Dec. 2021 |
| CSC498 , Introduction to Reinforcement Learning | Dec. 2021 |
| CSC265 , Enriched Data Structures and Analysis | Dec. 2020 |

* Graduate, [†] Cross-listed Graduate

TECHNICAL SKILLS

Languages: Python, Java, C, R, Bash

Libraries: PyTorch, TensorFlow, JAX, NumPy, SciPy, scikit-learn, pandas, Matplotlib, RDKit, Biopython, WandB, PyTorch Lightning, Deep Graph Library, PyTorch Geometric, Flax, Haiku

Tools: Git, Anaconda, Slurm, Google Cloud Platform, PyCharm

Training Methods: GPU, TPU, distributed training, mixed-precision