

## EDUCATION

---

- University of California, Berkeley** Berkeley, USA  
*PhD Student, Department of Electrical Engineering and Computer Science* *Aug 2021 - Present*
- University of Toronto** Toronto, Canada  
*Masters in Computer Science* *Jan 2019 – May 2020*  
– Thesis: *Contrastive Learning of Protein Representations by Mutual Information Maximization*
- University of Waterloo** Waterloo, Canada  
*Bachelors of Science, Honours Science, Bioinformatics Option* *Sept 2014 – May 2018*  
– Thesis: *Interpreting Convolutional Neural Networks for Discovering Regulatory Motifs of Femur Growth*

## RESEARCH EXPERIENCE

---

- UC Berkeley/Berkeley AI Research** Berkeley, USA  
*PhD Student — Advisor: Pieter Abbeel* *July 2021 – Present*  
– Developing methods for ML-guided protein design.
- Google Brain** Mountain View, USA  
*Student Researcher — Host: Andreea Gane* *May 2022 – Dec 2022*  
– Understanding protein functions using large language models.
- Insitro** South San Francisco, USA  
*Machine Learning Engineer III* *July 2020 – July 2021*  
– Engineering representations of microscopy image phenotypes for drug discovery.  
– Supporting ML-guided small molecule design through chemoinformatic analyses and developing novel representation learning strategies.
- University of Toronto/Vector Institute** Toronto, Canada  
*Masters Student — Advisors: Alan Moses, Marzyeh Ghassemi* *Jan 2019 – May 2020*  
– Developed self-supervised representations for proteins using contrastive mutual information maximization (*MLCB 2020, PLOS Computational Biology*).  
– Benchmarked self-supervised computer vision methods for microscopy images to address generalization challenges to natural covariate shifts (*NeurIPS 2019*).  
– Quantitative and qualitative evaluation of bias in contextual word embeddings of clinical notes (*Spotlight, ACM CHIL 2020*).
- Stanford University** Stanford, USA  
*Visiting Student Researcher* *Sept 2019 – Jan 2020*  
– Using domain adaptation methods to enable transcription factor binding prediction when evaluating on holdout cell lines.
- Harvard Medical School/Boston Children’s Hospital** Boston, USA  
*Intern, Research Computing — Advisor: Piotr Sliz* *Jun 2018 – Jan 2019*  
– Understanding genotype-phenotype relationships in childhood epilepsy. Interpreted important model features to seek novel disease-associated variants from whole exome (WES) data.
- University of Waterloo** Waterloo, Canada  
*Undergraduate Thesis Student — Advisor: Andrew Doxey* *Sept 2017 – May 2018*  
– Prediction of accessible chromatin regions for femur growth regulation. Reconstructed convolutional filters as a position-weighted matrix (PWM) with statistical matches in JASPAR, a database of known motifs.
- École polytechnique fédérale de Lausanne** Lausanne, Switzerland  
*Research Intern — Advisor: Matteo Dal Peraro* *Jun 2017 - Sept 2017*  
– Used molecular dynamics (MD) and GROMACS to simulate enzyme-membrane interaction mechanisms of NDM-1, an enzyme which confers antibiotic resistance.

## AWARDS

<b>D. E. Shaw Graduate &amp; Postdoc Women's Fellowship</b> – Fellowship program for graduate & postdoc women in computational drug discovery	2022
<b>Paula Hawthorn Fellowship</b> – UC Berkeley Computing, Data Science and Society departmental fellowship.	2021
<b>NSERC Postgraduate Scholarships – Doctoral program (PGS D) Award</b> – Federal doctoral scholarship tenurable abroad, selected in the Committee for Computing Sciences.	2020
<b>NSERC Canada Graduate Scholarships – Doctoral (CGS D) Award</b> – Federal doctoral scholarship tenurable only at a Canadian institution [DECLINED].	2020
<b>NSERC Michael Smith Foreign Supplement</b> – Supports high-calibre Canadian graduate students in pursuing research abroad.	2019
<b>Alexander Graham Bell Canada Graduate Scholarships – Master's (CGS M) Award</b> – Federal research scholarship for high-calibre Master's research students.	2018
<b>EPFL Scholarship of Excellence in Research</b> – Sponsors students for research internship at EPFL.	2017
<b>President's Scholarship of Distinction, Arebi Family Science Scholarship</b> – Entrance scholarships, University of Waterloo.	2014
<b>Royal Conservatory of Music (RCM)</b> – ARCT Performer's Diploma in Piano.	2013

## PREPRINTS AND PUBLICATIONS

- AX Lu, **AX Lu**, I Pritisanac, T Zarin, JD Forman-Kay, AM Moses. Discovering molecular features of intrinsically disordered regions by using evolution for contrastive learning. *PLOS Computational Biology*.
- S Kolli, **AX Lu**, X Geng, A Kumar, S Levine. Data-Driven Optimization for Protein Design: Workflows, Algorithms and Metrics. *ICLR 2022 Workshop on Machine Learning for Drug Discovery workshop (MLDD)*.
- C Dallago, K Schütze, M Heinzinger, T Olenyi, M Littmann, **AX Lu**, KK Yang, S Min, S Yoon, JT Morton, B Rost. Using protein sequence representations from deep learning to visualize and predict protein sets. *Current Protocols*.
- AX Lu**, H Zhang, M Ghassemi, AM Moses. Self-supervised contrastive learning of protein representations by mutual information maximization. *Machine Learning for Computational Biology (MLCB) 2020*.
- AX Lu**, AX Lu, AM Moses. Evolution Is All You Need: Phylogenetic Augmentation for Contrastive Learning. *Machine Learning for Computational Biology (MLCB) 2020*.
- H Zhang\*, **AX Lu\***, M Abdalla, M McDermott, M Ghassemi. Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings. *Spotlight, ACM Conference on Health, Inference, and Learning (CHIL) 2020*. \*Equal Contribution.
- AX Lu, **AX Lu**, W Schormann, M Ghassemi, DW Andrews, AM Moses. The Cells Out of Sample (COOS) dataset and benchmarks for measuring out-of-sample generalization of image classifiers. *Neural Information Processing Systems (NeurIPS) 2019*.
- M Abdalla, H Zhang, **AX Lu**, I Chen, M Ghassemi. Quantifying Fairness in a Multi-Group Setting and its Impact in the Clinical Setting. *NeurIPS 2019 Workshop on Fair ML for Health*.
- AM Moses, AX Lu, **AX Lu**, M Ghassemi. Transfer Learning vs. Batch Effects: what can we expect from neural networks in computational biology? *Machine Learning for Computational Biology (MLCB) 2019*.
- AX Lu, **AX Lu**, AM Moses. Paired Cell Inpainting: A Multiple-Instance Extension of Self-Supervised Learning for Bioimage Analysis. *ICML 2019 Workshop on Self-Supervised Learning*.
- J Ban, M Tadrous, **AX Lu**, EA Cicinelli, SM Cadarette. Diffusion of indirect comparison meta-analytic methods to study drugs: a systematic review and co-authorship network analysis. *BMJ Open*.

## SERVICE

Co-organizer, ML Protein Engineering Seminar Series – Biweekly research seminars for the broader ML for protein engineering community.	2022 – Present
Research to the People – Non-profit connecting patients of rare genomic diseases to the academic community to collaborative research initiatives.	2018 – 2020

Waterloo Residence Don	2016 – 2018
– Leader for the Velocity Residence (spin-off program of the Velocity start-up incubator) and first-year residences.	
Volunteer, Tosamaganga Hospital	2016
– Supported operations at a rural Tanzanian hospital.	

## REVIEWING

---

Machine Learning for Health (ML4H)	2020, 2021, 2022
Machine Learning for Computational Biology (MLCB)	2021
ICLR Workshop on AI4Science	2022
ICML Workshop on AI4Science	2022
NeurIPS Workshop on AI for Science	2021, 2022
NeurIPS Workshop on ML for Structural Biology	2022
NeurIPS Workshop on Distribution Shifts	2021, 2022
NeurIPS Workshop on Robustness in Sequence Modelling	2022

## TEACHING

---

Teaching Assistant, Genetics (BIOL 239)	2016
---	------

## MENTORSHIP

---

Sathvik Kolli	Undergrad, 5th year MS
---------------	------------------------