

## EDUCATION

---

<b>University of California, Berkeley</b> <i>PhD Student, Department of Electrical Engineering and Computer Science</i>	Berkeley, USA <i>Aug 2021 - Dec 2025</i>
– Thesis: <i>Generative Models for Real-World Drug Discovery</i>	
<b>University of Toronto</b> <i>Masters in Computer Science</i>	Toronto, Canada <i>Jan 2019 – May 2020</i>

– Thesis: <i>Contrastive Learning of Protein Representations by Mutual Information Maximization</i>	
<b>University of Waterloo</b> <i>Bachelors of Science, Honours Science, Bioinformatics Option</i>	Waterloo, Canada <i>Sept 2014 – May 2018</i>
– Thesis: <i>Interpreting Convolutional Neural Networks for Discovering Regulatory Motifs of Femur Growth</i>	

## RESEARCH EXPERIENCE

---

<b>Isomorphic Labs</b> <i>Senior Research Scientist</i>	London, UK <i>July 2025 – Present</i>
– Developing post-AlphaFold3 models for in-house drug discovery pipeline on the Pretraining team.	
<b>Prescient Design (Genentech)</b> <i>Data Scientist II (Part-Time, concurrent employment with PhD enrollment.)</i>	New York / South San Francisco, USA <i>June 2023 – June 2025</i>
– Led the development of PLAID and CHEAP, methods for prompt-conditioned generation of all-atom protein structures. – Scaled distributed training and dataloading pipelines for a 2 billion parameter model. – Explored strategies for guided diffusion for proteins using differentiable biophysical functions.	
<b>UC Berkeley/Berkeley AI Research</b> <i>PhD Student — Advisor: Dr. Pieter Abbeel</i>	Berkeley, USA <i>July 2021 – December 2025</i>
– Various projects in generative modeling, model-based optimization, interpretability, biosafety, and inference infrastructure, with applications to AI for science. – Provided technical mentorship and/or oversaw projects for undergraduate and Masters students.	
<b>Google Brain</b> <i>Student Researcher (full-time May-Aug, part-time Aug-Dec.) — Host: Dr. Andreea Gane</i>	Mountain View, USA <i>May 2022 – Dec 2022</i>
– Applied information retrieval techniques to protein function annotation.	
<b>Insitro</b> <i>Machine Learning Engineer III</i>	South San Francisco, USA <i>July 2020 – July 2021</i>
– Developed phenotype learning methods from microscopy images, and equivariant representation learning for small molecule design.	
<b>University of Toronto/Vector Institute</b> <i>Masters Student — Advisors: Drs. Alan Moses, Marzyeh Ghassemi</i>	Toronto, Canada <i>Jan 2019 – May 2020</i>
– Self-supervised representations for proteins using contrastive mutual information maximization. – Benchmarking self-supervised computer vision methods for microscopy images to address generalization challenges to natural covariate shifts. – Quantitative and qualitative evaluation of bias in contextual word embeddings of clinical notes.	
<b>Stanford University</b> <i>Visiting Student Researcher — Advisor: Dr. Anshul Kundaje</i>	Stanford, USA <i>Sept 2019 – Jan 2020</i>
– Developed methods for cross-cell transcription factor binding prediction via domain adaptation methods.	
<b>Harvard Medical School/Boston Children's Hospital</b> <i>Intern, Research Computing — Advisor: Dr. Piotr Sliz</i>	Boston, USA <i>Jun 2018 – Jan 2019</i>

- Developing interpretable models for discovering disease-associated variants from whole exome (WES) data.

### University of Waterloo

*Undergraduate Thesis Student — Advisor: Dr. Andrew Doxey*

Waterloo, Canada

*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

*Research Intern — Advisor: Dr. Matteo Dal Peraro*

Lausanne, Switzerland

*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

### D. E. Shaw Graduate & Postdoc Women's Fellowship

2022

- Fellowship program for graduate & postdoc women in computational drug discovery

### Paula Hawthorn Fellowship

2021

- UC Berkeley Computing, Data Science and Society departmental fellowship.

### NSERC Postgraduate Scholarships – Doctoral program (PGS D) Award

2020

- Federal doctoral scholarship tenurable abroad, selected in the Committee for Computing Sciences.

### NSERC Canada Graduate Scholarships – Doctoral (CGS D) Award

2020

- Federal doctoral scholarship tenurable only at a Canadian institution [DECLINED].

### NSERC Michael Smith Foreign Supplement

2019

- Supports high-calibre Canadian graduate students in pursuing research abroad.

### Alexander Graham Bell Canada Graduate Scholarships – Master's (CGS M) Award

2018

- Federal research scholarship for high-calibre Master's research students.

### EPFL Scholarship of Excellence in Research

2017

- Sponsors students for research internship at EPFL.

### President's Scholarship of Distinction, Arebi Family Science Scholarship

2014

- Entrance scholarships, University of Waterloo.

### Royal Conservatory of Music (RCM)

2013

- ARCT Performer's Diploma in Piano.

## PREPRINTS AND PUBLICATIONS

(*To appear*) G Brixi\*, MG Durrant\*, J Ku\*, M Poli\*, ... **AX Lu**, PD Hsu†, BL Hie†. Genome modeling and design across all domains of life with Evo 2. *Nature*. \*†*Equal Contributions*.

(*Preprint*) **AX Lu**, W Yan, S Robinson, KK Yang, V Gligorijevic, K Cho, Richard B, P Abbeel, N Frey. All-Atom Protein Generation with Latent Diffusion. *Oral, Workshop on Machine Learning for Structural Biology (MLSB)*, 2024.

C Cordon, **AX Lu**, P Abbeel. Protein Language Model Fitness Is a Matter of Preference. *International Conference on Learning Representations (ICLR)*, 2025.

**AX Lu**, W Yan, KK Yang, V Gligorijevic, K Cho, P Abbeel, Richard B, N Frey. Tokenized and Continuous Embedding Compressions of Protein Sequence and Structure. *Cell Patterns*, 2025.

R Boger\*, **AX Lu**\*, S Chithrananda\*, K Yang, P Skopintsev, B Adler, E Wallace, P Yoon, P Abbeel, J Doudna. TOPH: Adapting A Contrastive Question-Answering Framework for Protein Search. *ICML Workshop on Computational Biology*, 2023. \**Equal Contributions*.

(*Preprint*) AJ Reddy, MH Herschl, S Kolli, **AX Lu**, X Geng, A Kumar, PD Hsu, S Levine, NM Ioannidis. Pretraining strategies for effective promoter-driven gene expression prediction. *bioRxiv*, 2023.

**AX Lu**, **AX Lu**, I Pritišanac, T Zarin, JD Forman-Kay, AM Moses. Discovering molecular features of intrinsically disordered regions by using evolution for contrastive learning. *PLOS Computational Biology*, 2022

S Kolli, **AX Lu**, X Geng, A Kumar, S Levine. Data-Driven Optimization for Protein Design: Workflows, Algorithms and Metrics. *ICLR Workshop on Machine Learning for Drug Discovery (MLDD)*, 2022.

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*, 2021.

**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 Contributions.

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 Workshop on Fair ML for Health*, 2019.

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 Workshop on Self-Supervised Learning*, 2019.

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*, 2018.

## SERVICE

---

Co-Organizer, ICML 2025 Workshop on Exploration in AI Today (EXAIT)	2025
– Workshop on advances in exploration theory and applications to robotics, LLM alignment, and AI for science.	
Board Member, Berkeley Women in Computer Science and Engineering	2023 – 2025
– Event and programming supporting Berkeley women in CS and engineering.	
Co-organizer, ML Protein Engineering Seminar Series	2022 – 2023
– Biweekly research seminars for the broader ML for protein engineering community.	
Core Team, Research to the People	2018 – 2020
– Non-profit connecting rare genomic disease patients to academic communities and industry sponsors for collaborative research.	
Waterloo Residence Don	2016 – 2018
– Residence manager for the Velocity Residence (spin-off program of the Velocity start-up incubator) and first-year residences.	
Volunteer, Tosamaganga Hospital	2016
– Supported operations and shadowed physicians at a rural Tanzanian hospital	

## INVITED TALKS

---

Delta Podcast	January 2026
In silico 002, IDEALondon	October 2025
Novonesis	May 2025
Latent Labs Journal Club	March 2025
Baker Lab, University of Washington	February 2025
Gingko Bioworks Journal Club	February 2025
EvolutionaryScale	January 2025
Microsoft Research, AI for Science	January 2025
Marks Group, Harvard Medical School	January 2025
Listgarten Group, UC Berkeley	November 2024
Stanford AI + Biomedicine Seminar	October 2024
ML Protein Engineering Seminar Series	October 2024
South Park Commons Demo Night	September 2024
Learning on Graphs and Geometry (LoGG) Seminar Series	August 2024
Guest Lecture, BIOE145 at UC Berkeley on AlphaFold2	April 2024

## REVIEWING

---

Nature	2023
Nature Communications	2025
Neural Information Processing Systems (NeurIPS)	2024 – 2025
International Conference on Learning Representations (ICLR)	2025 – 2026
International Conference on Machine Learning (ICML)	2025 – 2026
Artificial Intelligence and Statistics Conference (AISTATS)	2025
Machine Learning for Computational Biology (MLCB)	2021
Machine Learning for Health (ML4H)	2020 – 2024
ML for Structural Biology (MLSB)	2025
NeurIPS Workshop on ML for Structural Biology (MLSB)	2022 – 2024
NeurIPS Workshop on Generative AI for Biology	2023
NeurIPS Workshop on AI for Science	2021 – 2023
NeurIPS Workshop on Distribution Shifts	2021 – 2023
NeurIPS Workshop on Robustness in Sequence Modelling	2022
ICML Workshop on AI4Science	2022
ICLR Workshop on AI4Science	2022
ICLR Workshop on Generative and Experimental Perspectives for Biomolecular Design (GEM)	2024 – 2026

## TEACHING

---

BIOENG 145: Intro to Machine Learning in Computational Biology, UC Berkeley – <i>Teaching Assistant</i>	2024
BIOL 239: Genetics, University of Waterloo – <i>Teaching Assistant</i>	2016
IB math & chemistry – <i>Tutor</i>	2014 – 2016
Piano performance & music theory – <i>Tutor</i>	2010 – 2014

## MENTORSHIP

---

Cade Gordon	Undergrad, 5th year MS
Alishba Imran	Undergrad
Sathvik Kolli	Undergrad, 5th year MS