Amy X. Lu

amyxlu@cs.toronto.edu

http://amyxlu.github.io

github.com/amyxlu

I am a graduate student at the University of Toronto/Vector Institute for AI. I am interested in **developing biologically-principled machine learning methods**, with an emphasis on **self-supervised representation learning, generalizability, and biological interpretability**. I also serve on Research to the People, which organizes research collaborations of genomics/ML researchers for patients of understudied diseases.

Education

- **University of Waterloo** — Waterloo, Canada
  Bachelors of Science, Honours Science, Bioinformatics Option — *2014 – 2019*
- **University of Toronto, Vector Institute** — Toronto, Canada
  Master's in Computer Science — *Jan 2019 – Present*

Experience

- **Stanford University** — Palo Alto, USA
  Visiting Student Researcher — Domain Adaptation in Regulatory Genomics, Dr. Anshul Kundaje — *2019 – 2019*
  - **Domain adaptation, ChIP-seq data, interpretability:** Using domain adaptation methods to improve transcription factor binding prediction when evaluating for a different cell line.
- **University of Toronto/Vector Institute** — Toronto, Canada
  Masters Student — Representation learning in genomics, Dr. Alan Moses & Dr. Marzyeh Ghassemi — *2019 – 2020*
  - **Self-supervised learning, protein representation learning:** Applied methods in NLP and mutual information maximization to learn self-supervised representations of protein sequences.
  - **Generalizability, computer vision:** Benchmarked self-supervised computer vision methods in a microscopy image dataset with covariate shift to highlight generalization failures in machine learning. (*NeurIPS 2019*)
  - **Algorithmic fairness, clinical decision support:** Quantitative and qualitative evaluation of bias in contextual word embeddings on clinical notes; fairness definitions for multi-group settings (*Spotlight, ACM CHIL 2020*)
- **Harvard Medical School/Boston Children's Hospital** — Boston, USA
  Research Intern — Machine Learning in Clinical Genomics, Dr. Piotr Sliz — *2018 – 2019*
  - **Genotype-phenotype studies, clinical interpretability:** Understanding disease genotype-phenotype relationships using machine learning models. Interpreted important model features to seek novel disease-associated variants from whole exome (WES) data.
  - **Variant filtering, pathway analyses:** Applied standard filtering pipelines for false-positive variants. Explored classification from pathway- and variant-level features.
  - **Machine learning, dimensionality reduction:** Explored methods for capturing epistatic non-linearities and statistical dimensionality reduction.
- **University of Waterloo** — Waterloo, Canada
  Undergraduate Thesis Student — Deep Learning in Regulatory Genomics, Dr. Andrew Doxey — *2017 – 2018*
  - **Chromatin accessibility prediction, interpretability:** Trained a convolutional neural network to classify ATAC-seq accessible regions; reconstructed first-layer features as a position-weighted matrix (PWM) with statistical matches in JASPAR, a database of known motifs.
  - **Phylogenetics, metagenomic data mining:** Used various bioinformatics pipeline tools (HMMER, BLAST, etc.) to understand biochemical properties of potentially uncharacterized toxins in metagenomic data.
- **École polytechnique fédérale de Lausanne** — Lausanne, Switzerland
  Research Intern — Molecular Dynamics Simulations, Dr. Matteo Dal Parero — *2017*
  - **Molecular dynamics:** Used molecular dynamics (MD) and GROMACS to simulate enzyme-membrane mechanisms of bacterial "superbugs"
- **University of Toronto** — Toronto, Canada
  Research Intern — Data Visualization in Pharmacoepidemiology, Dr. Suzanne Cadarette — *2015 – 2017*
  - **Data visualization, pharmacoepidemiology:** Analysis and visualization of the social diffusion of methodological innovation in pharmacoepidemiology.

Awards

- **NSERC Postgraduate Scholarships – Doctoral program (PGS D) Award *[Declined]***: Federal doctoral scholarship, selected in the Committee for Computing Sciences ($63,000).
- **Alexander Graham Bell Canada Graduate Scholarships – Master's (CGS M) Award**: Federal research grant ($17,500).
- **NSERC Michael Smith Foreign Supplement**: "Supports high-calibre Canadian graduate students in pursuing research abroad" ($6,000).
- **Scholarship of Excellence in Research**: Sponsors students for research internship at EPFL (CHF 4,500).
- **University of Waterloo**: Various entrance awards totaling $6,000.
- **Royal Conservatory of Music (RCM)**: ARCT Performer's Diploma in Piano.

Publications and Conference Proceedings

- **Preprint: Lu AX***, Zhang H, Ghassemi M, Moses AM. Self-Supervised Contrastive Learning of Protein Representations By Mutual Information Maximization. ***bioRxiv***.
- **Accepted:** Zhang H*, **Lu AX***, Abdalla M, McDermott M, Ghassemi M. Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings. ***Spotlight Talk, ACM Conference on Health, Inference, and Learning (CHIL) 2020***. **Equal contribution.*
- **Published:** Lu AX, **Lu AXP**, Schormann W, Ghassemi M, Andrews DW, Moses AM. The Cells Out of Sample (COOS) dataset and benchmarks for measuring out-of-sample generalization of image classifiers. ***Poster, Neural Information Processing Systems (NeurIPS) 2019.***
- **Presented:** Moses A, Lu AX, **Lu AXP**, Ghassemi M. Transfer Learning vs. Batch Effects: what can we expect from neural networks in computational biology? ***Poster, Machine Learning for Computational Biology (MLCB) 2019.***
- **Published:** Ban J, Tadrous M, **Lu AX**, Cicinelli EA, Cadarette SM. Diffusion of indirect comparison meta-analytic methods to study drugs: a systematic review and co-authorship network analysis. ***BMJ Open.***

Workshop Papers and Posters

- **Presented: Lu AX***, Zhang H*, Abdalla M, McDermott M, Ghassemi M. Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings. Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings. ***Poster, NeurIPS 2019 Workshop on Fair ML for Health.*** **Equal contribution.*
- **Presented:** Zhang H*, **Lu AX***, Abdalla M, McDermott M, Ghassemi M. Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings. Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings. ***Extended Abstract, NeurIPS 2019 Workshop on Machine Learning for Healthcare.*** **Equal contribution.*
- **Presented:** Abdalla M, Zhang H, **Lu AX**, Chen I, Ghassemi M. Quantifying Fairness in a Multi-Group Setting and its Impact in the Clinical Setting. ***Poster, NeurIPS 2019 Workshop on Fair ML for Health.***
- **Presented:** Lu AX, **Lu AXP**, Moses A. Paired Cell Inpainting: A Multiple-Instance Extension of Self-Supervised Learning for Bioimage Analysis. ***Poster, ICML 2019 Workshop on Self-Supervised Learning.***
- **Presented: Lu AX**, Rockowitz S, Poduri A, Sliz P. From data to precision medicine: predictive machine learning models to uncover disease-associated variants. ***Lightning Talk/Poster, Harvard Medical School BCMP Retreat 2019.***
- **Submitted: Lu AX**, Tam ES. Effect of prophylactic brimonidine on subconjunctival hemorrhage in laser-assisted cataract surgery. ***Poster Abstract.***
- **Presented: Lu AX**, Consiglio GP, Cadarette SM. Dynamic Visualization in Co-Authorship Network Analysis. ***Poster, Leslie Dan Faculty of Pharmacy Undergraduate Research Symposium.***

Service and Activities

- **Research to the People** (*formerly SVAI*): Core Team of Research to the People, a non-profit connecting patients of rare genomic diseases to the medical/AI research community and industry partners through collaborative research initiatives.
- **Residence Don:** Organized events and established rapport with diverse students. Responded to mental health and conduct crises. Leader for the Velocity Residence, a spin-off for the Velocity start-up incubator.
- **Tosamaganga Hospital, Tanzania:** Supported operations at a rural Tanzanian hospital and shadowed surgical procedures. Expenses were covered by scholarships, fundraising, and part-time tutoring.

Teaching

- **Teaching Assistant, Genetics:** Taught weekly tutorial lectures for BIOL 239 at the University of Waterloo.
- **Piano, music theory:** Taught piano performance (up to RCM 7), ear training, RCM music history, and RCM Intermediate Rudiments.
- **Tutoring:** Tutor for IB Math and IB Chemistry.

Talks

- **Vector NLP Talks:** Quantifying and Removing Biases in Clinical Contextual Word Embeddings. *Co-presenter.*
- **Harvard Medical School BCMP Retreat 2019:** From data to precision medicine: predictive machine learning models to uncover disease-associated variants. *Lightning talk.*