

Amy X. Lu

<http://amyxlu.github.io>
github.com/amyxlu

amyxlu@stanford.edu
+1-650-776-6772
[linkedin.com/in/amyxlu](https://www.linkedin.com/in/amyxlu)

SUMMARY

I am a graduate student at the Vector Institute for Artificial Intelligence. 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 SVAI, which organizes research collaborations of genomics/AI researchers and industry partners 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 Degree in Computer Science 2019 – Present
- **Stanford University** Palo Alto, USA
Partial completion of Master's Thesis, Departments of Computer Science and Genetics 2019 – Present

EXPERIENCE

- **Stanford University** Palo Alto, USA
Visiting Student Researcher — Domain Adaptation in Regulatory Genomics, Dr. Anshul Kundaje 2019 – Present
 - **Domain adaptation, ChIP-seq analyses:** Using domain adaptation to improve classifier performance and learned genomic sequence features for transcription factor binding when evaluating in a different cellular context.
- **University of Toronto/Vector Institute** Toronto, Canada
Masters Student — Representation learning in genomics, Dr. Alan Moses & Dr. Marzyeh Ghassemi 2019 – Present
 - **Natural language processing, self-supervised learning:** Applying contextual word embedding models to learn generalizable and interpretable representations for genomic and proteomic sequences.
 - **Algorithmic fairness:** Quantitative and qualitative evaluation of bias in “black box” contextual clinical word embeddings, and exploring debiasing methods using adversarial gradient-reversal. (*NeurIPS 2019 ML for Healthcare workshop* – 34% acceptance rate; submitted to *FAT** 2020)
 - **Generalizability, computer vision:** Benchmarked a new “paired cell in-painting” method for out-of-sample generalization in microscopy against various common computer vision algorithms. (*NeurIPS 2019* – 21% acceptance rate; *ICML 2019 Self-Supervised Learning Workshop*)
- **Harvard Medical School/Boston Children's Hospital** Boston, USA
Research Intern — Machine Learning in Clinical Genomics, Dr. Piotr Sliz 2018 – Present
 - **Variant analysis, genotype-phenotype studies:** Applying standard filtering pipelines to a whole exome (WES) dataset of epilepsy patients.
 - **Machine learning, dimensionality reduction:** Building machine learning models to classify disease phenotype from high-dimensional exome variant data.
- **University of Waterloo** Waterloo, Canada
Undergraduate Thesis Student — Deep Learning in Regulatory Genomics, Dr. Andrew Doxey 2017 – 2018
 - **Convolutional neural networks in genomics:** Implemented a convolutional neural network which recognizes genomic enhancers regulating femur growth with 91% accuracy.
 - **Interpretability:** Distilled biological knowledge from the “black box” network by reconstructing first-layer features as position-weighted matrices (PWMs) of “motifs” (i.e. genomic patterns). All motifs found statistical matches in JASPAR (database of known motifs).
 - **Phylogenetics, metagenomic data mining:** Used various bioinformatics pipeline tools (HMMER, BLAST, etc.) to understand and discover potential novel toxin homologs in metagenomic databases.
 - **Self-directed learning:** Self-learned deep learning concepts and Python as a biology student.
- **École polytechnique fédérale de Lausanne** Lausanne, Switzerland
Research Intern — Structural Biology and MD Simulations, Dr. Matteo Dal Parero 2017

- **Molecular dynamics:** Contributed three potential findings on enzyme-membrane mechanisms of bacterial “superbugs” using computational simulations based on molecular dynamics (MD) principles.
- **Automating workflows:** Increased workflow efficiency by writing R, tcl/tk, and Csh scripts.

University of Toronto

Toronto, Canada

Research Intern — Data Visualization in Pharmacoepidemiology, Dr. Suzanne Cadarette

2015 – 2017

- **Pharmacoepidemiology:** Took initiative to redesign manual workflows. Contributed two co-authored publications, one poster, and figures for an international epidemiology conference within two months.
- **Data visualization:** Wrangled data using R and SQL to create animated and interactive co-authorship networks.

AWARDS

- **NSERC Michael Smith Foreign Supplement:** Supports high-calibre Canadian graduate students in pursuing research abroad. Valued at \$6,000.
- **Canada Graduate Scholarships-Master’s (CGS-M) Award:** NSERC federal research grant. Valued at \$17,500.
- **Scholarship of Excellence in Research:** One of 13 students sponsored for research at EPFL. Valued at CHF 4500.
- **University of Waterloo:** Various entrance awards totaling \$6000.
- **Associate of The Royal Conservatory (ARCT):** Performer’s Diploma in Piano (highest academic standing offered).

SERVICE AND ACTIVITIES

- **SVAI:** Core Team of SVAI, a non-profit connecting patients of rare genomic diseases to the medical/AI research community and industry partners through collaborative research initiatives.
- **Residence Don, Velocity Incubator:** Leader for the residence program of Canada’s most productive start-up incubator; fostered close relations with diverse students to support conduct and mental health crises.
- **Tosamaganga Hospital, Tanzania:** Supported operations at a rural Tanzanian hospital and shadowed surgical procedures. Expenses were covered by scholarships, fundraising, and part-time tutoring.

PUBLICATIONS AND CONFERENCE PROCEEDINGS

- **Accepted:** Lu AX, Lu AXP, Moses A. The Cells Out of Sample (COOS) dataset and benchmarks for measuring out-of-sample generalization of image classifiers. *NeurIPS 2019*.
- **Submitted:** Zhang H*, Lu AX*, Abdalla M, McDermott M, Ghassemi M. Hurtful Words: Quantifying Biases in Clinical Contextual Word Embeddings. *Submitted to ACM Conference on Fairness, Accountability, and Transparency (ACM FAT*)*. *Co-first authors.
- **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*.
- **Under revision:** Consiglio GP, Maclure M, Lu AX, Cicinelli EA, Ban JK, McCarthy L, Cadarette SM. Guidance documents for the application of Self-controlled Crossover Observational PharmacoEpidemiology (SCOPE) designs are needed: systematic review and animated co-authorship networks. *Submitted to Pharmacoepi Drug Saf*.

WORKSHOP PAPERS AND POSTER PRESENTATIONS

- **Accepted:** 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 presentation, NeurIPS 2019 Workshop on Fair ML for Health*. *Co-first authors.
- **Accepted:** 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*. *Co-first authors.
- **Accepted:** 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 presentation, 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 presentation, 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. *Poster and “Lightning Talk”, 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. Submitted to American Academy of Ophthalmology 2017*.

- **Presented:** Lu AX, Consiglio GP, Cadarette SM. Dynamic Visualization in Co-Authorship Network Analysis. *Poster presentation. Leslie Dan Faculty of Pharmacy Undergraduate Research Symposium.*
- **Presented:** McIlroy-Young R, Lu AX, Guenther N, Olarnyk A. A Network Analysis: Did the Arab Spring Impact the Academic Network of Tunisia Between 2010 and 2015? *Poster presentation. Knowledge Integration Symposium.*

TEACHING

- **Teaching Assistant, Genetics:** Taught weekly tutorials for BIOL 239. *University of Waterloo.*
- **Piano:** Taught piano performance (up to RCM 7), music history, and music theory throughout high school.
- **Tutoring:** Tutor for IB Calculus, IB Chemistry, and English.

HACKATHONS AND PROJECTS

- **Chatsense:** Mobile chat app using a sentiment analysis machine learning API on voice messages and mapping dominant emotions to emojis, designed for the autistic community. *Team project. Winner at YHacks at Yale.*
- **Alice:** Web login page using a trained facial recognition model to login to online banking without passwords, designed for dementia patients. *Team project. Winner at ConUHacks III.*