Amy X. Lu

amyxlu@berkeley.edu

http://amyxlu.github.io

github.com/amyxlu

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*
- **University of Waterloo** — Waterloo, Canada
  Bachelors of Science, Honours Science, Bioinformatics Option — *Sept 2014 – May 2018*

Research Experience

- **Google Brain** — Mountain View, USA
  Student Researcher — *May 2022 – Present*
  - **Protein sequence modelling:** Understanding protein functions with large-scale training pretraining.
- **UC Berkeley/Berkeley AI Research** — Berkeley, USA
  PhD Student — *July 2021 – Present*
  - **Biological sequence design:** Developing methods for exploring protein fitness landscapes for ML-guided protein design.
- **Insitro** — South San Francisco, USA
  Machine Learning Engineer III — *July 2020 – July 2021*
  - **Image profiling:** Engineered biologically interpretable representations of microscopy imaging phenotypes for drug discovery.
  - **Molecular design:** Supported ML-guided molecular design through representation learning and chemoinformatics analyses.
- **University of Toronto/Vector Institute** — Toronto, Canada
  Masters Student — *Jan 2019 – May 2020*
  - **Representation learning for proteins:** Developed a parameter-efficient representation for proteins using contrastive mutual information maximization (*MLCB 2020*).
  - **Generalizability to distribution shifts:** Benchmarked self-supervised computer vision methods in a microscopy image dataset with covariate shift to highlight generalization failures in machine learning (*NeurIPS 2019*).
  - **Medical note NLP, algorithmic fairness:** Quantitative and qualitative evaluation of bias in contextual word embeddings of clinical notes; fairness definitions for multi-group settings (*Spotlight, ACM CHIL 2020*).
- **Stanford University** — Stanford, USA
  Visiting Student Researcher, Kundaje Lab — *Sept 2019 – Jan 2020*
  - **Regulatory genomics:** 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
  Research Intern — *Jun 2018 – Jan 2019*
  - **Exome variant prioritization:** 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 — *Sept 2017 – May 2018*
  - **Chromatin accessibility:** 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 — *Jun 2017 - Sept 2017*
  - **Molecular dynamics:** Used molecular dynamics (MD) and GROMACS to simulate enzyme-membrane mechanisms of NDM-1, an enzyme which confers antibiotic resistance.

Preprints and Publications

- AX Lu, **AX Lu**, I Pritianac, 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.***
- C Dallago, K Schütze, M Heinzinger, T Olenyi, M Littmann, **AX Lu**, KK Yang, S Min, S Yoon, B Rost. Streamlining value of protein embeddings through `bio_embeddings`. ***NeurIPS 2020 Workshop on Learning Meaningful Representations of Life (LMRL).***
- **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.***

Awards

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

Service and Activities

- **Reviewing & Program Committee:** Machine Learning for Health *(2020, 2021)*; Machine Learning for Computational Biology *(2021)*; NeurIPS Workshop on AI for Science *(2021)*; NeurIPS Workshop on Distribution Shifts *(2021)*.
- **Research to the People, Core Team** (*formerly SVAI*): Non-profit connecting patients of rare genomic diseases to the medical/AI research community and industry partners through collaborative research initiatives.
- **Tosamaganga Hospital:** Supported operations at a rural Tanzanian hospital.
- **Residence Don:** Organized events, responded to crises, and established rapport with diverse students. Leader for the Velocity Residence, a spin-off program of the Velocity start-up incubator.