Victoria Lin

PERSONAL INFORMATION

Phone: +1 (408) 318-4905
Email: victoria@stat.cmu.edu
Citizenship: United States, Canada

Website
Google Scholar

EDUCATION

Carnegie Mellon University 2020 - 2026 (expected)
Ph.D., Statistics and Machine Learning
Advisors: Louis-Philippe Morency and Eli Ben-Michael

Carnegie Mellon University 2020 - 2021
M.S., Statistics

Carnegie Mellon University 2018 - 2020
M.S., Computational Data Science

Harvard University 2013 - 2017
A.B. (cum laude), Statistics and Molecular & Cellular Biology

AWARDS

Meta Research PhD Fellowship	2023 -
Two Sigma Diversity PhD Fellowship Finalist	2023
Best Paper Award Nominee, ICMI 2020	2020
Best Paper Award, AffCon Workshop at AAAI 2020	2020
Research Fellowship, Harvard College Program for Research in Science and Engineering	2016
Research Grant, Harvard College Pechet Family Research Fund	2016

EXPERIENCE

Carnegie Mellon University 2019 -
Graduate Research Assistant (Advisors: Louis-Philippe Morency and Eli Ben-Michael)

I work on "causal NLP," or causal inference for natural language data and language models. Some technical problems I find interesting include using causal principles to improve language models, learning causal effects from high-dimensional unstructured data like language, and learning language representations for valid causal inference.

Previously, I worked on affective computing and multimodal machine learning (with Louis-Philippe Morency and Jeffrey Girard) and mental health applications of causal inference (with Edward Kennedy).

Microsoft Research & Microsoft Experiences+Devices 2022
JEM Research Intern (Hosts: Srinagesh Sharma and Dimitrios Dimitriadis)

Harvard School of Public Health 2017 - 2018
Research Assistant (Advisor: Miguel Hernán)

Broad Institute of MIT and Harvard 2016 - 2017
Research Assistant (Advisor: Steven McCarroll)

Publications

* denotes equal contribution

Preprints

Victoria Lin, Louis-Philippe Morency, Eli Ben-Michael. "Isolated Causal Effects of Natural Language." *arXiv 2410.14812. Under review.* [PDF, code]

Peer-Reviewed Publications

Victoria Lin, Eli Ben-Michael, Louis-Philippe Morency. "Optimizing Language Models for Human Preferences is a Causal Inference Problem." [PDF, code]
UAI 2024.

Victoria Lin, Louis-Philippe Morency, Eli Ben-Michael. "Text-Transport: Toward Learning Causal Effects of Natural Language." [PDF, code]
EMNLP 2023.

Victoria Lin, Louis-Philippe Morency, Dimitrios Dimitriadis, Srinagesh Sharma. "Counterfactual Augmentation for Multimodal Learning Under Presentation Bias." [PDF, code]
EMNLP Findings 2023.

Victoria Lin, Louis-Philippe Morency. "SenteCon: Leveraging Lexicons to Learn Human-Interpretable Language Representations." [PDF, code]
ACL Findings 2023.

Aneesha Sampath, **Victoria Lin**, Louis-Philippe Morency. "SeedBERT: Recovering Annotator Rating Distributions from an Aggregated Label." [PDF]
UDM Workshop at AAAI 2023.

Victoria Lin*, Jeffrey Girard*, Michael Sayette, Louis-Philippe Morency. "Toward Multimodal Modeling of Emotional Expressiveness." [PDF, code]
ICMI 2020. Nominated for Best Paper Award.

Sean McGrath*, **Victoria Lin***, Zilu Zhang, Lucia Petito, Roger Logan, Miguel Hernán, Jessica Young. "gfoRmula: An R Package for Estimating the Effects of Sustained Treatment Strategies via the Parametric g-formula." [PDF, code]
Patterns 1(3), 100008.

Victoria Lin, Jeffrey Girard, Louis-Philippe Morency. "Context-Dependent Models for Predicting and Characterizing Facial Expressiveness." [PDF]
AffCon Workshop at AAAI 2020. Best Paper Award.

Leslie Tong, Seo Yeon Yoon, Yaisa Andrews-Zwilling, Alyssa Yang, **Victoria Lin**, Hanci Lei, Yadong Huang. "Enhancing GABA Signaling during Middle Adulthood Prevents Age-Dependent GABAergic Interneuron Decline and Learning and Memory Deficits in ApoE4 Mice." [PDF]
Journal of Neuroscience 36(7), 2316-2322.

Software Libraries

`sentecon`: Python library for interpretable language representations.
`gfoRmula`: R package for longitudinal causal effect estimation with the g-formula.

Presentations

Invited Talks

Carnegie Mellon University DeGroot Research Workshop — 2023
Counterfactual Augmentation for Learning Under Presentation Bias.

Contributed Talks

American Causal Inference Conference (ACIC) — 2023
Generalizing Text Experiments to Real-World Contexts with Large Language Models.

International Conference on Multimodal Interaction (ICMI) — 2020
Toward Multimodal Modeling of Emotional Expressiveness.

AffCon Workshop at the AAAI Conference on Artificial Intelligence — 2020
Context-Dependent Models for Predicting and Characterizing Facial Expressiveness.

Reviewing

UAI (2025), ACL (2023), EMNLP (2022, 2023)

Teaching

Carnegie Mellon University *Graduate Teaching Assistant*

36-402 Advanced Methods for Data Analysis	Spring 2023
36-468 Text Analysis	Fall 2022
36-462 Methods of Statistical Learning	Spring 2022
36-309 Experimental Design for Behavioral & Social Sciences	Fall 2021
36-401 Modern Regression	Fall 2020
11-631 Data Science Seminar (*also as guest lecturer*)	Fall 2019

Harvard University *Undergraduate Teaching Fellow*

CS109A Data Science	Fall 2016

Mentorship

Aneesha Sampath (CMU B.S. → University of Michigan Ph.D.)

Skills

Programming & Frameworks: Python, PyTorch, R, SQL, Java, C/C++, Bash, Git, LaTeX
Languages: English (native), Mandarin Chinese (heritage)