## Megha Srivastava

## megha@cs.stanford.edu • www.cs.stanford.edu/~megha 735 Campus Dr., Stanford, CA 94305

EDUCATION Stanford University • Stanford, CA Fall 2020 - present PhD in Computer Science, co-advised by Dorsa Sadigh & Dan Boneh Supported by the NSF Graduate Research Fellowship (2018 - 2023) and IBM PhD Fellowship (2023-2024) Research Areas: human-AI interaction, reliable machine learning Fall 2023 - Winter 2024 Massachusetts Institute of Technology • Cambridge, MA Visiting PhD student in CSAIL, hosted by Jacob Andreas Stanford University • Stanford, CA Fall 2014 – Spring 2019 BS in Computer Science with Honors and Distinction, Minor in Creative Writing MS in Computer Science (Artificial Intelligence), advised by Percy Liang & Tatsunori Hashimoto Bing Overseas Studies Program at Oxford University (Logic & Computability) in Spring 2017 WORK EXPERIENCE AI Resident – Microsoft Research Fall 2019 - Fall 2020 Redmond, WA • Model compatibility and human-AI collaboration with the Adaptive Systems & Interaction team. **Research Intern – Google Research** Summer 2019 Los Angeles, CA • Holistic image understanding with the Research & Machine Intelligence team. Research Intern – ETH Zürich Learning & Adaptive Systems Group Summer 2018 Zürich, Switzerland • Adaptive questioning and human subjectivity of fairness in machine learning (Andreas Krause & Hoda Heidari) • Supported by the ETH Zürich Student Summer Research Fellowship. Research Assistant - Stanford Natural Language Processing Group Summer 2017 – Spring 2019 Stanford, CA • Distributionally robust optimization and bias in machine learning systems (Percy Liang & Tatsu Hashimoto) • Interactive data collection to build natural language interfaces for plotting tools (CURIS Summer Project) Research Assistant - Vision & Perception Neuroscience Group Summer 2016 – Spring 2017 Stanford, CA • Generalization and perceptual invariances in both human and artificial vision models (Kalanit Grill-Spector) • Supported by the Bio-X Undergraduate Research Award. Awards & Fellowships • Selected as a Rising Star in Machine Learning (University of Maryland), 2023 • IBM PhD Fellowship, 2023 • Selected for WIN6 at Banff International Research Station, 2023 Machine learning & arithmetic group led by Kristin Lauter and Rachel Newton • Women in National Security Scholar, 2023 Project with Gordian Knot Center on developing AI Literacy • American Association for the Advancement of Science (AAAS) Mass Media Fellowship finalist, 2019 • International Conference on Machine Learning (ICML) Best Paper Runner-Up Award, 2018 • National Science Foundation Graduate Research Fellowship, 2018 • Ben Wegbreit Prize for Best Undergraduate Honors Thesis in Computer Science (Stanford), 2018 • Phi Beta Kappa, 2018 • Lunsford Award for Oral Presentation of Research Nominee (Stanford), 2016 ACTIVITIES & SERVICE • Stanford AI Lab Blog Editor, 2020-present • Organizer for Women in Machine Learning (WiML) 2023, co-located with NeurIPS 2023 Student Program & Travel Funding Chair • Reviewer for NeurIPS 2020, ICLR 2021, ICML 2021, ICML 2022, ICLR 2023, CoRL 2023, NeurIPS 2023

- MS Computer Science Social Committee, 2018 2019
- Stanford Women in Computer Science (Vice President), 2017 2018
- Course Assistant for CS 221: Introduction to Artificial Intelligence, 2018
- Photography showcased at Stanford Art of Science Exhibition, Frost Music Festival, dnjGallery at Bergamot

## PUBLICATIONS

Neil Perry\*, Megha Srivastava\*, Deepak Kumar, Dan Boneh. "Do Users Write More Insecure Code with AI Assistants?" ACM Conference on Computer and Communications Security (CCS), 2023

Mina Lee, **Megha Srivastava**, Amelia Hardy, John Thickstun, Esin Durmus, Ashwin Paranjape, Ines Gerard-Ursin, Xiang Lisa Li, Faisal Ladhak, Frieda Rong, Rose E. Wang, Minae Kwon, Joon Sung Park, Hancheng Cao, Tony Lee, Rishi Bommasani, Michael Bernstein, Percy Liang. *"Evaluating Human-Language Model Interaction"* Transactions on Machine Learning Research (TMLR), 2023.

Megha Srivastava, Noah Goodman, Dorsa Sadigh. "Generating Language Corrections for Teaching Physical Control Tasks" Proceedings of the 40th International Conference on Machine Learning (ICML), 2023.

Megha Srivastava, Erdem Biyik, Suvir Mirchandani, Noah Goodman, Dorsa Sadigh. "Assistive Teaching of Motor Control Tasks to Humans" Advances in Neural Information Processing Systems 35 (NeurIPS), 2022.

Siddharth Karamcheti<sup>\*</sup>, **Megha Srivastava**<sup>\*</sup>, Percy Liang, Dorsa Sadigh. "*LILA: Language-Informed Latent Actions*" Proceedings of the 5th Conference on Robot Learning (CoRL), 2021.

Megha Srivastava and Noah Goodman. "Question Generation for Adaptive Education" Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL), 2021.

Megha Srivastava, Besmira Nushi, Ece Kamar, Shital Shah, Eric Horvitz. "An Empirical Analysis of Backward Compatibility in Machine Learning Systems" Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), 2020.

Megha Srivastava, Tatsunori Hashimoto, Percy Liang. "Robustness to Spurious Correlations via Human Annotations" Proceedings of the 37th International Conference on Machine Learning (ICML), 2020.

Megha Srivastava, Hoda Heidari, Andreas Krause. "Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning."Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), 2019.

Tatsunori Hashimoto, Megha Srivastava, Hongseok Namkoong, Percy Liang. "Fairness Without Demographics in Repeated Loss Minimization." Proceedings of the 35th International Conference on Machine Learning (ICML), 2018. Best Paper Runner-Up Award

Megha Srivastava. "The Computational and Aesthetic Foundations of Artificial Empathy." Intersect: The Stanford Journal of Science, Technology and Society 10, no. 1 (2016).

Megha Srivastava, David Remus, Kalanit Grill-Spector. "The Role of Learning in Complex Object Recognition and Discrimination Across Spatial Transformations: An Experimental Comparison of Artificial CNNs and Human Subjects." Women in Machine Learning Workshop (WiML), 2017 and Vision Sciences Society (VSS), 2017.

## INVITED TALKS

"Challenges in Human-AI Interaction for Information-Seeking Tasks" Rising Stars in Machine Learning Workshop, November 2023.

"Do Users Write More Insecure Code with AI Assistants?" University of Maryland College Park (Guest Lecture for Large Language Models, Security, and Privacy seminar), October 2023.

"Robustness to Spurious Correlations via Human Annotations" Two Sigma PhD Symposium, June 2023.

"Assistive Teaching of Motor Control Tasks to Humans" Stanford Collaborative Haptics in Robotics and Medicine Lab, May 2023.

"Do Users Write More Insecure Code with AI Assistants?" UC Berkeley Security Seminar, May 2023.

"Assistive Teaching of Motor Control Tasks to Humans" SystemX Alliance Fall Conference, November 2022.

"Assistive Teaching of Motor Control Tasks to Humans" Simons Institute Workshop on AI & Humanity, July 2022.

"Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning" Oxford University Algorithms at Work Reading Group, May 2021.

"Fairness & Robustness with Missing Information" Stanford Causality & Cognition Lab, December 2020.

"Mathematical Notions vs. Human Perception of Fairness: A Descriptive Approach to Fairness for Machine Learning" Microsoft Research AI & Society Reading Group, October 2019.