

Appointment

2025- **Assistant Professor**, Cornell University
Department of Information Science and Cornell Tech
Field Membership in Computer Science, Data Science

Education and Training

2024-2025 **Postdoctoral Fellow**, Stanford University
Advisors: Sanmi Koyejo, Daniel E. Ho
Affiliations: HAI, RegLab

2019-2024 **Ph.D. Computer Science**, Princeton University
M.A. Computer Science
Dissertation: Operationalizing Responsible Machine Learning: From Equality Towards Equity
Committee: Olga Russakovsky (advisor), Arvind Narayanan, Solon Barocas, Janet Vertesi, Aleksandra Korolova

2015-2019 **B.S. Electrical Engineering and Computer Science, Philosophy minor**, UC Berkeley
Advisors: Pieter Abbeel, Aviv Tamar

Fellowships and Awards

Center for Democracy & Technology Non-Resident Fellow: 2026-2027

Microsoft AI & Society Fellowship Sociotechnical Approaches to Measuring Harms Caused by AI Systems: 2024, 2025

Machine Learning Rising Stars 2024 (Academic career workshop hosted at University of Maryland)

Siebel Scholar Class of 2024 (Awarded annually for academic excellence and demonstrated leadership to over 80 top students from the world's leading graduate schools)

EECS Rising Stars 2023 (Intensive academic career workshop hosted at Georgia Tech)

National Science Foundation Graduate Research Fellowship Program (NSF GRFP)

National Defense Science and Engineering Graduate Fellowship Program (NDSEG) (declined)

Mark D. Weiser Excellence in Computing Scholarship (Merit-based award for 1-2 students in Berkeley EECS)

Regents and Chancellors' Scholar (top 2% of incoming class at UC Berkeley)

Berkeley EECS Honors Program, concentration in Philosophy

Phi Beta Kappa (Academic Honor Society)

Eta Kappa Nu (Electrical and Computer Engineering Honors Society)

Tau Beta Pi (Engineering Honors Society)

Journal and Conference Publications

A. Olteanu, S. L. Blodgett, A. Balayn, A. Wang, F. Diaz, F. Calmon, M. Mitchell, M. Ekstrand, R. Binns, S. Barocas. Rigor in AI: Doing Rigorous AI Work Requires a Broader, Responsible AI-Informed Conception of Rigor. *Annual Conference on Neural Information Processing Systems (NeurIPS) Position Paper* 2025 (40/496, 8%)

A. Wang, D. Ho, S. Koyejo. The Inadequacy of Offline LLM Evaluations: A Need to Account for Personalization in Model Behavior. *Patterns* 2025

V. Raman, J. Shen, A. Zhang, L. Gailmard, R. Bommasani, D. Ho*, A. Wang*. Disclosure and Evaluation as Fairness Interventions for General-Purpose AI. *AAAI/ACM Conference on AI, Ethics, and Society (AIES)* 2025 (239/749, 32%)

A. Wang, M. Phan, D. Ho, S. Koyejo. Fairness through Difference Awareness: Measuring Desired Group Discrimination in LLMs. *Proceedings of the Association for Computational Linguistics (ACL)* 2025 - *Best paper* (4/8360 for best paper; 1699/8360, 29% acceptance in main conference overall)

A. Wang. Identities are not Interchangeable: the Harms of Generalizing in Fair ML. *ACM Conference on Fairness, Accountability and Transparency (FAccT)* 2025 (218/812, 27% acceptance)

A. Wang, X. Bai, S. Barocas, S. L. Blodgett. Measuring Machine Learning Harms from Stereotypes Requires Understanding Who is Being Harmed by Which Errors in What Ways. *ACM Conference on Fairness, Accountability and Transparency (FAccT)* 2025 (218/812, 27% acceptance)

L. Weidinger, I. D. Raji, H. Wallach, M. Mitchell, A. Wang, O. Salaudeen, R. Bommasani, D. Ganguli, S. Koyejo, W. Isaac. Toward an Evaluation Science for Generative AI Systems. *The Bridge* 2025

H. Wallach, M. Desai, A. F. Cooper, A. Wang, C. Atalla, S. Barocas, S. L. Blodgett, A. Chouldechova, E. Corvi, P. A. Dow, J. Garcia-Gathright, A. Olteanu, N. Pangakis, S. Reed, E. Sheng, D. Vann, J. W. Vaughan, M. Vogel, H. Washington, A. Z. Jacobs. Position: Evaluating Generative AI Systems Is a Social Science Measurement Challenge. *International Conference on Machine Learning (ICML) Position Paper* 2025

A. Wang, J. Morgenstern, J. P. Dickerson. Large Language Models that Replace Human Participants can Harmfully Misportray and Flatten Identity Groups. *Nature Machine Intelligence* 2025

X. Bai, A. Wang, I. Sucholutsky, T. L. Griffiths. Measuring Implicit Bias in Explicitly Unbiased Large Language Models. in *Proceedings of the National Academy of Sciences (PNAS)* 2025

A. Wang, A. Hertzmann, O. Russakovsky. Benchmark suites instead of leaderboards for evaluating AI fairness. *Patterns* 2024

A. Wang, T. Datta, J. P. Dickerson. Strategies for Increasing Corporate Responsible AI Prioritization. *AAAI/ACM Conference on AI, Ethics, and Society (AIES)* 2024 (150/468, 32% acceptance)

S. Engelmann, M. Z. Choksi, A. Wang, C. Fiesler. Visions of a Discipline: Analyzing Introductory AI Courses on YouTube. *ACM Conference on Fairness, Accountability and Transparency (FAccT)* 2024 (175/725, 24% acceptance)

A. Wang, O. Russakovsky. Overwriting Pretrained Bias with Finetuning Data. *International Conference on Computer Vision (ICCV)* 2023 - *Oral presentation* (152/8068, 2% acceptance for oral; 2160/8068 acceptance overall)

N. Meister*, D. Zhao*, A. Wang, V. Ramaswamy, R. Fong, O. Russakovsky. Gender Artifacts in Visual Datasets. *International Conference on Computer Vision (ICCV)* 2023 (2160/8068, 27% acceptance)

A. Wang*, S. Kapoor*, S. Barocas, A. Narayanan. Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy. *Journal of Responsible Computing (JRC)* 2023. Also at *ACM Conference on Fairness, Accountability and Transparency (FAccT)* 2023 (150/608, 25% acceptance)

J. Katzman*, A. Wang*, M. Scheuerman, S. L. Blodgett, K. Laird, H. Wallach, S. Barocas. Taxonomizing and Measuring Representational Harms: A Look at Image Tagging. *AAAI Conference on Artificial Intelligence (AAAI)* 2023 (1721/8777, 20% acceptance)

A. Mathur, A. Wang, C. Schwemmer, M. Hamin, B. M. Stewart, A. Narayanan. Manipulative Tactics are the Norm in Political Emails: Evidence from 100K emails from the 2020 U.S. Election Cycle. *Big Data & Society* 2023

A. Wang, V. V. Ramaswamy, O. Russakovsky. Towards Intersectionality in Machine Learning: Including More Identities, Handling Underrepresentation, and Performing Evaluation. *ACM Conference on Fairness, Accountability and Transparency (FAccT)* 2022 (179/711, 25% acceptance)

A. Wang, S. Barocas, K. Laird, H. Wallach. Measuring Representational Harms in Image Captioning. *ACM Conference on Fairness, Accountability and Transparency (FAccT)* 2022 (179/711, 25% acceptance)

A. Wang, A. Liu, R. Zhang, A. Kleiman, L. Kim, D. Zhao, I. Shirai, A. Narayanan, O. Russakovsky. REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets. *International Journal of Computer Vision (IJCV)* 2022

D. Zhao, A. Wang, O. Russakovsky. Understanding and Evaluating Racial Biases in Image Captioning. *International Conference on Computer Vision (ICCV)* 2021 (1617/6236, 26% acceptance)

A. Wang, O. Russakovsky. Directional Bias Amplification. *International Conference on Machine Learning (ICML)* 2021 (1184/5513, 21% acceptance)

A. Wang, A. Narayanan, O. Russakovsky. REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets. *European Conference on Computer Vision (ECCV)* 2020 - *Spotlight presentation* (160/5150, 3% acceptance for spotlight; 1360/5150 acceptance overall)

A. Wang, T. Kurutach, K. Liu, P. Abbeel, A. Tamar. Learning Robotic Manipulation through Visual Planning and Acting. *Robotics: Science and Systems (RSS)* 2019 (85/272, 31% acceptance)

Workshop Publications

Computer science workshop publications tend to be more “lightly” peer-reviewed

A. Chan*, C. T. Okolo*, Z. Turner*, A. Wang*. The Limits of Global Inclusion in AI Development. *AAAI 2021 Workshop on Reframing Diversity in AI* - *Spotlight presentation*

W. Wang, A. Wang, A. Tamar, X. Chen, P. Abbeel. Safer Classification by Synthesis. *NeurIPS 2017 Aligned AI Workshop*

Work Experience

- Summer 2023 **Arthur AI**, *Machine Learning Research Fellow*, New York City, NY
- Conduct research on corporate motivations for RAI and societal implications of large language models
 - Mentor: John P. Dickerson
- Summer 2021 **Microsoft Research**, *Research Intern*, Remote
- Measure representational harms in image captioning with MSR FATE (Fairness, Accountability, Transparency, and Ethics) and Cognitive Services teams
 - Mentors: Solon Barocas, Hanna Wallach, Lijuan Wang, Zhe Gan
- Summer 2019 **Google**, *Software Engineering Intern*, Mountain View, CA
- Worked on Google Shopping infrastructure team to build out new features for metric reporting
 - Contributed side project of supplementing filtering by different metrics for fairness evaluation
- Jan 2017 - **Archer (Technology Nonprofit, archerimpact.com)**, *Engineering Lead*, Berkeley, CA
- Oct 2018
- Use Node and React to build web app for conducting open source investigations based off user interviews
 - Visualize public data and create adjacency matrix scheme to manipulate entity connections using D3
 - Present products in D.C. and RightsCon 2018 in Toronto, receiving medal from U.S. Treasury
- Summer 2017 **Google**, *Engineering Practicum Intern*, Seattle, WA
- Worked on infrastructure team to improve internal streaming data processing system

Teaching

Cornell Tech

Spring 2026 **INFO 5390/CS 5382: Fairness in Machine Learning**

Fall 2025 **INFO/CS 6125: Non-Ideal Algorithmic Fairness**

Other

Summer 2023 **AI4ALL (AI for historically marginalized talent)**, *Responsible AI Curriculum Specialist*, Lead
- Spring 2024 creation of Responsible AI education content for college program

Spring 2021 **Fairness in Machine Learning (COS 534)**, *Teaching Assistant*, Princeton, NJ

Fall 2020 **Fairness in Visual Recognition (COS IW 08)**, *Teaching Assistant*, Princeton, NJ

Fall 2020 **Computer Vision (COS 429)**, *Teaching Assistant*, Princeton, NJ

July 2020 **AI4ALL**, *Instructor*, Princeton, NJ

July 2018 - **Introduction to Deep Learning DeCal**, *Student instructor for 200 person course*, Berkeley, CA
May 2019

July 2018 - **Robot Learning Lab Outreach**, *Lead*, Berkeley, CA
May 2019

Fall 2018 **Introduction to Machine Learning (CS189/289A)**, *Academic Intern*, Berkeley, CA

Service

Advisory Board: Patterns Journal

Conference Organizing Committee: Financial Support Chair at FAccT 2025, 2026; Ethics Chair of COLM 2026

Workshop Organizing Committee: Workshop at CVPR 2021 (Responsible Computer Vision), Workshop at ECCV 2022 (Responsible Computer Vision), Tutorial at FAccT 2022 (Fairness in Computer Vision: Datasets, Algorithms, and Implications), Workshop at NAACL 2025 (Language Models for Underserved Communities), Tutorial at FAccT 2025 (AI Measurement as a Stakeholder-Engaged Design Practice)

Reviewer: NeurIPS (2023, 2024, D&B 2025), ICML Ethics (2024, 2025, 2026), FAccT (2023, 2024, 2025), AIES (2022), CVPR (2022, 2023, 2024), ECCV (2022), Responsible Computer Vision Workshop (CVPR 2021, ECCV 2022), Regulatable ML Workshop (NeurIPS 2024), IJCV (2023), Science Advances (2024), Patterns (2025), Nature (2025), Big Data & Society (2025), PNAS (2025), Trends in Cognitive Sciences x 2 (2025), Nature Climate Change (2025), Journal of Responsible Computing (2025), Social Sciences and Humanities Research Council of Canada (2026), Science (2026), Nature Machine Intelligence (2026)

National Science Foundation (NSF) Panelist: 2025

Queer in AI: AI Policy

Stanford HAI: Coordinating panelist for ethics and society review of seed grant proposals

Talks and Panels

November 2025 **MIT AI4Society Seminar**, *More Discrimination for Fairness*

October 2025 **Microsoft Security GLEAM Series**, *Evaluating Generative AI is a Social Problem*

October 2025 **Digital Life Initiative**, *"Fairness" in AI: Gone Too Far, and Also Not Far Enough*

October 2025 **Microsoft's NYC Engineering + Research Responsible AI Mixer**, *"Fairness" in AI: Gone Too Far, and Also Not Far Enough*

August 2025 **Cisco Tech Talk**, *Evaluating Generative AI is a Social Problem*

June 2025 **CVPR Demographic Diversity Workshop**, *Demographic Diversity: Gone Too Far, And Also Not Far Enough*

- June 2025 **CVPR The Sixth Workshop on Fair, Data-efficient, and Trusted Computer Vision**, *Benchmark Suites to Move Beyond Simple Fairness Considerations*
- February 2025 **Berkeley Institute of Design Seminar**, *The Biases To Care About: ML Fairness that is "Just Right"*
- December 2024 **NeurIPS Workshop: Algorithmic Fairness through the Lens of Metrics and Evaluation**, *Fairness through Difference Awareness: Measuring Desired Group Discrimination in LLMs*
- October 2024 **University of Washington RAISE Seminar**, *Operationalizing Responsible Machine Learning: From Equality Towards Equity*
- September 2024 **Arthur AI Fest**, *LLM Representation of Personas*
- September 2024 **Nokia Bell Labs Responsible AI Seminar Series**, *Equality is not equity: Recognizing group differences in AI fairness*
- September 2024 **Brown University Critical Data and Machine Learning Studies Seminar**, *Against Predictive Optimization*
- March 2024 **Morgan State University National Symposium on Equitable AI**, *Equality is not Equity: Recognizing Group Differences in AI*
- March 2024 **Cornell Tech Seminar**, *Operationalizing Responsible Machine Learning: From Equality Towards Equity*
- March 2024 **Cornell University Information Science Colloquium**, *Operationalizing Responsible Machine Learning: From Equality Towards Equity*
- March 2024 **University of Southern California Computer Science Colloquium**, *Operationalizing Responsible Machine Learning: From Equality Towards Equity*
- January 2024 **Georgia Tech School of Computational Science and Engineering**, *Operationalizing Responsible Machine Learning: From Equality Towards Equity*
- November 2023 **University of Maryland's Values-Centered Artificial Intelligence (VCAI) Initiative**, *Sociotechnically Grounded Responsible Machine Learning*
- April 2023 **Yale University's Data (Re)Makes the World**, *Against predictive optimization: on the legitimacy of decision-making algorithms that optimize predictive accuracy*
- November 2022 **Princeton University's Data-Driven Social Science Initiative**, *Machine learning mistakes aligned with stereotypes are more harmful*
- September 2022 **Princeton University Psychology's Susan Fiske Lab**, *ML mistakes aligned with stereotypes are more harmful than mistakes which are not*
- August 2022 **AAAI/ACM Conference on AI, Ethics, and Society (AIES)**, *Student presentation: Fairness implications behind the technical assumptions of machine learning*
- May 2022 **University of Chicago Psychology's Perception and Judgment Lab**, *Stereotypes in machine learning*
- April 2022 **Canadian Parliament Hearing for Facial Recognition Technology before the Information, Privacy and Ethics Committee**, *Expert witness*
- June 2021 **CVPR Learning with Limited and Imperfect Data Workshop**, *Mitigating bias and privacy concerns in visual data (with advisor Prof. Olga Russakovsky)*
- June 2021 **CVPR Women in Computer Vision Workshop Keynote**, *Perception, interaction and fairness: key components of visual recognition (with advisor Prof. Olga Russakovsky)*
- April 2021 **Out in Tech Panel for Macy's Pride ERG**, *Panelist*
- June 2020 **CVPR's Seventh Workshop on Fine-Grained Visual Categorization**, *Revealing and mitigating biases in visual datasets (with advisor Prof. Olga Russakovsky)*

November **Berkeley CS10, Guest Lecture on AI**
2018

Outreach

- Fall 2020 - **Bias in AI Reading Group** - organize biweekly reading group centered on topics of bias and
Spring 2024 fairness in machine learning at Princeton University
- Fall 2021 - **Princeton Pre-Application Support Program** - coordinate department-wide program that
Fall 2023 matches graduate school computer science applicants to current students to review application
materials; grants application fee waivers to all participants (159 applicants in recent year)
- Fall 2020 - **RISE (Research Inclusion Social Event)** - co-organize monthly department social event with
Spring 2022 discussions centered around diversity and inclusion issues in computer science
- Fall 2021 - **Q'nnnections Mentor** - mentor for LGBTQ+ students at Princeton
Fall 2022
- Fall 2019 - **Society of Women Engineers** - mentor for undergraduates
Spring 2022
- Fall 2019 - **Out in Tech Mentor** - mentor for LGBTQ+ youth ages 17-24 interested in technology
Fall 2022
- Fall 2020 **JuST (Technology for a Just Society)** - moderate anti-racist reading group discussions