

Education

- 2019-2024 **Ph.D. Computer Science**, Princeton University
(expected) **M.A. Computer Science**
Advisor: Olga Russakovsky
Research Interests: machine learning fairness, algorithmic bias
- 2015-2019 **B.S. Electrical Engineering and Computer Science**, *Philosophy minor*, UC Berkeley
Advisors: Pieter Abbeel, Aviv Tamar
Major GPA: 3.96/4.00

Awards

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

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

Member of Phi Beta Kappa (Academic Honor Society)

Member of Eta Kappa Nu (Electrical and Computer Engineering Honors Society)

Member of Tau Beta Pi (Engineering Honors Society)

Journal and Conference Publications

Overwriting Pretrained Bias with Finetuning Data

A. Wang, O. Russakovsky. *ICCV 2023* - *Oral presentation* (2160/8068, 27% acceptance – 152/8068 for Oral)

Gender Artifacts in Visual Datasets

N. Meister*, D. Zhao*, A. Wang, V. Ramaswamy, R. Fong, O. Russakovsky. *ICCV 2023* (2160/8068, 27% acceptance)

Against Predictive Optimization: On the Legitimacy of Decision-Making Algorithms that Optimize Predictive Accuracy

A. Wang*, S. Kapoor*, S. Barocas, A. Narayanan. *FAccT 2023* (150/608, 25% acceptance)

Taxonomizing and Measuring Representational Harms: A Look at Image Tagging

J. Katzman*, A. Wang*, M. Scheuerman, S. L. Blodgett, K. Laird, H. Wallach, S. Barocas. *AAAI 2023* (1721/8777, 20% acceptance)

Manipulative Tactics are the Norm in Political Emails: Evidence from 100K emails from the 2020 U.S. Election Cycle

A. Mathur, A. Wang, C. Schwemmer, M. Hamin, B. M. Stewart, A. Narayanan. *Big Data & Society 2023*

Towards Intersectionality in Machine Learning: Including More Identities, Handling Underrepresentation, and Performing Evaluation

A. Wang, V. V. Ramaswamy, O. Russakovsky. *FAccT 2022* (179/711, 25% acceptance)

Measuring Representational Harms in Image Captioning

A. Wang, S. Barocas, K. Laird, H. Wallach. *FAccT 2022* (179/711, 25% acceptance)

REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets

A. Wang, A. Liu, R. Zhang, A. Kleiman, L. Kim, D. Zhao, I. Shirai, A. Narayanan, O. Russakovsky. *IJCV 2022*

Understanding and Evaluating Racial Biases in Image Captioning

D. Zhao, A. Wang, O. Russakovsky. *ICCV 2021* (1617/6236, 26% acceptance)

Directional Bias Amplification

A. Wang, O. Russakovsky. *ICML 2021* (1184/5513, 21% acceptance)

REVISE: A Tool for Measuring and Mitigating Bias in Visual Datasets

A. Wang, A. Narayanan, O. Russakovsky. *ECCV 2020* - *Spotlight presentation* (1360/5150, 26% acceptance – 160/5150 for spotlight)

Learning Robotic Manipulation through Visual Planning and Acting

A. Wang, T. Kurutach, K. Liu, P. Abbeel, A. Tamar. *RSS 2019* (85/272, 31% acceptance)

Workshop Publications

The Limits of Global Inclusion in AI Development

A. Chan*, C. T. Okolo*, Z. Turner*, A. Wang*. *AAAI 2021 Workshop on Reframing Diversity in AI* - *Spotlight presentation*

Safer Classification by Synthesis

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

Work Experience

- Summer 2023 **Arthur AI**, *Machine Learning Research Fellow*, New York City, NY
- Conduct research on societal implications of large language models
 - Mentor: John 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

- Summer 2023 **AI4ALL Responsible AI Curriculum Specialist** - create teaching content for Responsible AI component of AI4ALL (AI for historically marginalized talent) college program
- present
- 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 (AI for Historically Marginalized Talent)**, *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

Talks and Panels

- 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 Institute**, *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 **Hearing for Facial Recognition Technology before the Information, Privacy and Ethics Committee of the Canadian Parliament**, *Expert witness*
- June 2021 **CVPR Learning with Limited and Imperfect Data Workshop**, *Mitigating bias and privacy concerns in visual data with Olga Russakovsky*
- June 2021 **CVPR Women in Computer Vision Workshop Keynote**, *Perception, interaction and fairness: key components of visual recognition with 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 Olga Russakovsky*
- November 2018 **Berkeley CS10**, *Guest Lecture on AI*

Leadership

- Fall 2021 - present **Princeton Pre-Application Support Program** - coordinate department-wide program that matches graduate school computer science applicants to current students to help review applicants' application materials; grants application fee waivers to all participants (159 applicants in most recent year)
- Fall 2020 - present **Bias in AI Reading Group** - organize biweekly reading group centered on topics of bias and fairness in machine learning at Princeton University
- Fall 2020 - Spring 2022 **RISE (Research Inclusion Social Event)** - co-organize monthly social event with discussions centered around diversity and inclusion issues in computer science

Organizing

Fairness in Computer Vision: Datasets, Algorithms, and Implications Tutorial at FAccT 2022

Responsible Computer Vision Workshop at ECCV 2022

Responsible Computer Vision Workshop at CVPR 2021

Reviewing

NeurIPS 2023

FAccT 2023

AIES 2022

CVPR 2022, 2023

ECCV 2022

Responsible Computer Vision Workshop at CVPR 2021, ECCV 2022

International Journal of Computer Vision (IJCV)

Outreach

Some under "Leadership"

Fall 2021 - **Q'nnnections Mentor** - mentor for LGBTQ+ students at Princeton
Fall 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

Fall 2019 - **Society of Women Engineers** - mentor for undergraduates
Spring 2020

Research Mentorship

Non-research mentorship under "Outreach"

Allison Chen, PhD student at Princeton Computer Science

Xinran Liang, PhD student at Princeton Computer Science

Dora Zhao, Undergraduate and Master's student at Princeton, Current: PhD student at Stanford Computer Science

Nicole Meister, Undergraduate student at Princeton, Current: PhD student at Stanford Electrical Engineering

Kara Liu, Undergraduate student at UC Berkeley, Current: PhD student at Stanford Computer Science

Anat Kleiman, Master's student at Princeton, Current: PhD student at Harvard Computer Science

Alexander Liu, Undergraduate student at Princeton

Ryan Zhang, Undergraduate student at Princeton

Leslie Kim, Undergraduate student at Princeton

Iroha Shirai, Undergraduate student at Princeton

Frelicia Tucker, Undergraduate student at Princeton