Sarah H. Cen

Assistant Professor
Carnegie Mellon University
Department of Electrical & Computer Engineering
Department of Engineering & Public Policy

Email: sarahcen@andrew.cmu.edu
Website: https://shcen.github.io/

PROFESSIONAL	**Carnegie Mellon University**, Assistant Professor	2026-
	Department of Electrical and Computer Engineering	
	Department of Engineering and Public Policy	
	Stanford University, Postdoctoral Scholar	2025-2026
	Institute for Human-Centered Artificial Intelligence (HAI)	
	Department of Computer Science	
	Stanford Law School	
	Advised by Professor Daniel E. Ho and Professor Percy Liang	
	DeepMind, Research Intern	2021
	Princeton University, Research Assistant	2016
	MIT Lincoln Labs, Research Intern	2015
	University of Pennsylvania, Research Intern	2014
	Wattvision, Software Engineer	2013

EDUCATION	**Massachusetts Institute of Technology, Cambridge, MA**	2018-2024
	Ph.D. in Electrical Engineering and Computer Science	
	Advised by Professor Aleksander Mądry and Professor Devavrat Shah	
	Thesis: "Paths to AI Accountability: Design, Measurement, and the Law"	
	University of Oxford, Oxford, UK	2016-2018
	M.Sc. by Research in Engineering Science	
	Advised by Professor Paul Newman	
	Thesis: "Radar-only Ego-Motion Estimation and Localization"	
	Princeton University, Princeton, NJ	2012-2016
	B.S.E. in Mechanical Engineering	
	Certificate in Computer Science & Robotics	
	Advised by Professor Naomi Leonard	
	Thesis: "Leader selection in dynamic networks"	

SELECTED AWARDS & FELLOWSHIPS		
	Thriving Star, MIT EECS	2022
	Best Student Talk, LIDS Conference	2022
	Hugh Hampton Young Memorial Fund Fellowship, MIT fellowship	2020
	Chyn Duog Shiah Memorial Fellowship, MIT fellowship	2020
	Edwin S. Webster Fellowship, MIT EECS fellowship	2018
	Ida M. Green Fellowship, MIT fellowship	2018

Daniel M. Sachs Scholarship, Oxford scholarship given to one student   2016
MacCracken Senior Thesis Award, top thesis in Princeton SEAS   2016
McKinzie Senior Thesis Prize, top thesis in Princeton Mechanical Eng.   2016
Sau-Hai Lam Prize, top student in Mechanical & Aerospace Eng.   2016
Phi Beta Kappa, Sigma Xi   2016
Tau Beta Pi   2015
Shapiro Prize for Academic Excellence, top 3 percent of Princeton class   2014

PUBLICATIONS   **Peer-Reviewed & Archival Publications** (*Marks first-author contribution)

18. **Sarah H. Cen*,** Andrew Ilyas, Hedi Driss, Charlotte Park, Aspen Hopkins, Chara Podimata, Aleksander Madry. "Large-Scale, Longitudinal Study of Large Language Models During the 2024 US Election Season." arXiv. Link.

17. **Sarah H. Cen*,** Salil Goyal, Zaynah Javed, Ananya Karthik, Percy Liang, Daniel E. Ho. "Audits Under Resource, Data, and Access Constraints: Scaling Laws For Less Discriminatory Alternatives." AAAI /ACM Conference on Artificial Intelligence, Ethics, & Society (AIES'25). Link.

16. Rishi Bommasani, Scott R Singer, Ruth E Appel, **Sarah H. Cen**, A Feder Cooper, Lindsey A Gailmard, Ian Klaus, Meredith M Lee, Inioluwa Deborah Raji, Anka Reuel, Drew Spence, Alexander Wan, Angelina Wang, Daniel Zhang, Daniel E Ho, Percy Liang, Dawn Song, Joseph E Gonzalez, Jonathan Zittrain, Jennifer Tour Chayes, Mariano-Florentino Cuellar, Li Fei-Fei. "The California Report on Frontier AI Policy." arXiv. Link.

15. Aspen K. Hopkins*, **Sarah H. Cen***, Andrew Ilyas, Isabella Struckman, Luis Videgaray, Aleksander Mądry. "AI Supply Chains: An Emerging Ecosystem of AI Actors, Products, and Services." Eighth AAAI /ACM Conference on Artificial Intelligence, Ethics, & Society (AIES'25). Link.

14. **Sarah H. Cen** and Rohan Alur. "From Transparency to Accountability and Back: A Discussion of Access and Evidence in AI Auditing." ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO'24). Link.

13. **Sarah H. Cen***, Andrew Ilyas*, and Aleksander Mądry. "User Strategization and Trustworthy Algorithms." 2025 ACM Conference on Economics and Computation (EC'24). Link.

   Also appeared in Responsible Decision Making in Dynamic Environments Workshop at ICML 2022 as an **Oral**.

12. **Sarah H. Cen***, Andrew Ilyas*, Jennifer Allen, Hannah Li, and and Aleksander Mądry. "Measuring Strategization in Recommendation: Users Adapt Their Behavior to Shape Future Content." 2025 ACM Conference on Economics and Computation (EC'24). R&R to *Management Science*. Link.

11. Anish Agarwal, **Sarah H. Cen***, Devavrat Shah, and Christina Lee Yu. "Network Synthetic Interventions: A Framework for Panel Data with Network Interference." R&R to *Operations Research*. Link.

10. Cindy Zhang*, **Sarah H. Cen***, and Devavrat Shah. "Matrix Estimation for Individual Fairness." International Conference on Machine Learning (ICML), 2023. Link.

9. **Sarah H. Cen***, Aleksander Mądry, and Devavrat Shah. "A User-Driven Framework for Regulating and Auditing Social Media." arXiv, 2023. In preparation for submission to ICML 2024. Link.

8. **Sarah H. Cen*** and Manish Raghavan. "The Right to Be an Exception to a Data-Driven Rule." MIT Case Studies in Social and Ethical Responsibilities of Computing (SERC), 2023. Invited paper. In preparation for submission the Conference on Fairness, Accountability, and Transparency (FAccT) 2024. Link.

7. Julien Perolat*, Bart de Vylder*, **et al.** "Mastering the game of Stratego with model-free multiagent reinforcement learning." *Science*, 2022. Link.

6. **Sarah H. Cen*** and Devavrat Shah. "Regulating algorithmic filtering on social media." Neural Information Processing Systems (NeurIPS), 2021. **Spotlight**. Link

5. **Sarah H. Cen*** and Devavrat Shah. "Regret, stability & fairness in matching markets with bandit learners." International Conference on Artificial Intelligence and Statistics (AISTATS), 2022. Link.

4. **Sarah H. Cen*** and Paul Newman. "Radar-only ego-motion estimation in difficult settings via graph matching." IEEE International Conference on Robotics and Automation (ICRA), 2019. Link.

3. Rob Weston*, **Sarah H. Cen**, and Ingmar Posner. "Probably unknown: Deep inverse sensor modelling radar." ICRA 2019. Link.

2. **Sarah H. Cen*** and Paul Newman. "Precise ego-motion estimation with milli-meter-wave radar under diverse and challenging conditions." ICRA 2018. Link.

1. **Sarah H. Cen***, Vaibhav Srivastava, and Naomi Ehrich Leonard. "On robustness and leadership in markov switching consensus networks." IEEE Conference on Decision and Control (CDC), 2017. Link.

Public Policy

3. The California Report on Frontier AI Policy. Directly contributed to the writing of California's SB 53, which was passed in 2025. Written with Rishi Bommasani, Scott R Singer, Ruth E Appel, A Feder Cooper, Lindsey A Gailmard, Ian Klaus, Meredith M Lee, Inioluwa Deborah Raji, Anka Reuel, Drew Spence, Alexander Wan, Angelina Wang, Daniel Zhang, Daniel E Ho, Percy Liang, Dawn Song, Joseph E Gonzalez, Jonathan Zittrain, Jennifer Tour Chayes, Mariano-Florentino Cuellar, Li Fei-Fei.

2. Aleksander Mądry's Testimony to the Senate Committee on Judiciary (link). Written by **Sarah H. Cen**, Andrew Ilyas, Aleksander Mądry, and Luis Videgaray. "Artificial Intelligence and Human Rights." June 13, 2023.

1. Aleksander Mądry's Testimony to the House Committee on Oversight and Accountability (link). Written by **Sarah H. Cen**, Andrew Ilyas, Aleksander Mądry, and Luis Videgaray. "Advances in AI: Are We Ready For a Tech Revolution?" March 8, 2023.

Blogs

3. *Blog series on AI Accountability & Transparency* (September 2023)

   Written by **Sarah H. Cen**, Cosimo L. Fabrizio, Dan Huttenlocher, Aleksander Mądry, Martha Minow, and James Siderius. (*Marks first-author contribution)

   1. *Introduction: AI Accountability & Transparency (link)
   2. *AI Risk Scores: What Do They Mean? (link)
   3. *How Much Access Is Needed to Audit an AI System? (link)
   4. Ensuring an Adequate Mechanism of Relief for AI-Induced Harms
   5. A Public Health Approach to AI Regulation

2. *Blog Series on AI Deployment* ([April-August 2023](link)) – full report [here](link)

   Written by **Sarah H. Cen**, Aspen Hopkins, Andrew Ilyas, Aleksander Mądry, Isabella Struckman, and Luis Videgaray. (*Marks first-author contribution)

   1. We need to talk about AI deployment (link)
   2. *AI supply chains and why they matter (link)
   3. +Who will provide AI to the world? (link)
   4. *How Upstream AI impacts Downstream AI (link)
   5. *Three Proposals for Regulating AI (link)

1. *Blog Series on Social Media* ([December 2022](link))

   Written by **Sarah H. Cen**, Andrew Ilyas, and Aleksander Mądry. (First authorship shared with Andrew Ilyas on all posts)

   1. Social Media Platforms & the Information Exchange They Run (link)
   2. What's To Know About Regulating Social Media in the US? (link)
   3. Why Can't We Regulate Social Media Like Previous Media? (link)
   4. Regulating Social Media: Lessons from Existing Regulation (link)

MEDIA &
PUBLICITY

Articles on Longitudinal Study of LLMs during the 2024 Election   2024

Science (link), by Celina Zhao
"How AIs responded to the 2024 U.S. presidential election—in real time"

Forbes (link), by Ron Schmelzer
"AI Models Acted Like Voters In 2024 And The Findings Are Startling"

MIT CSAIL News (link), by Rachel Gordon
"Peering inside political AI: How LLMs responded to the 2024 election"

Fortune (link), by Patrick Kulp
"MIT researchers studied 16 million election-related AI responses"

Forbes Article (link), by Lance Eliot   2022
"AI Ethics Fighting Passionately For Your Legal Right To Be An Exception"

Profile in MIT News (link), by Grace Chua   2022
"Living better with algorithms"

MIT EECS Article (link), by Rachel Yang   2022
"Thriving Stars of AI"

+ Was not an author on this blog post only. It is included for completeness of the series.

SELECTED TALKS	**Invited Talks**	
	Explorations in Statistics, Probability, Learning & Optimization (exSPLOre)	2026
	Simons Theory of Algorithmic Fairness	2026
	INFORMS Session on Networks	2025
	Stanford RAIN Seminar	2025
	ICML 2nd Workshop on Reliable and Responsible Foundation Models	2025
	FAR.AI Technical Innovations for AI Policy Conference	2025
	Karlsruhe Institute of Technology (KIT) Guest Lecture	2025
	INRIA Theory of Auditing Group	2025
	Columbia University's Frontiers of AI Symposium	2025
	Northeastern University Network Science Institute	2025
	FAccT Tutorial on Algorithmic Auditing and Generative AI	2025
	CS & Law Research Workshop	2024
	INFORMS Session on Online Marketplaces & Incentives	2024
	The National Academies of Sciences, Engineering, and Medicine's Committee on Science, Technology, and Law Disinformation Workshop	2024
	Cornell Tech Digital Life Initiative (DLI)	2023
	INFORMS Session on Fairness in Platforms & Recommendation	2023
	INFORMS Session on Decisions Under Not-So-Perfect Data	2023
	Cornell Young Researchers Workshop (YRW)	2023
	Computing & Society, Stanford CS	2023
	Thriving Stars of AI, MIT EECS	2022
	AI & Humanity Program, Berkeley Simons Institute	2022
	AI Ethics Reading Group, MIT	2022
	Session on Online Platforms, INFORMS	2022

Conference Talks

Machine Learning in Economics Summer Conference (MLESC25)	2025
ACM CS & Law Lightning Talks	2025
ACM Economics & Computation (EC)	2024
ACM EAAMO	2024
Conference on Digital Experimentation (CODE)	2023
Marketplace Innovations Workshop	2023
Responsible Decision-Making Workshop at ICML	2022
Harvard Ethical Issues in Computing & AI	2022
Privacy Law Scholars Conference	2022
Information Credibility Workshop at ICWSM	2021
Stanford GSB Rising Scholars	2020
IEEE Conference on Decision & Control	2017

Institutional Talks

Stanford ML Seminar	2024
LIDS Student Conference	2022, 2023
Econometrics Seminar	2022
LIDS Stats and Tea	2021, 2022, 2023
Machine Learning at MIT	2020
MLxMIT Retreat	2020
Graduate Women in EECS Research Summit	2019

Posters

NeurIPS	2025
ACM CS & Law	2025
American Causal Inference Conference (ACIC)	2023
Marketplace Innovations Workshop (MIW)	2023
MIT Social & Ethical Responsibilities of Computing Symposium	2023
Causality for Real-World Impact Workshop at NeurIPS	2022
Responsible Decision-Making Workshop at ICML (**Oral**)	2022
Berkeley Simons Institute	2022
NeurIPS (**Spotlight**)	2021
Women in Data Science	2020, 2021
IEEE Int'l Conference on Robotics & Automation (ICRA)	2018, 2019

TEACHING

Teaching Assistant, Science & Technology Policy (MIT, 17.925)	2021
Instructor: Bill Bonvillian	
Teaching Assistant, Inference and Information (MIT, 6.437)	2020
Instructors: Professor Gregory Wornell and Professor Lizhong Zheng	
Instructor & Organizer, MOSTEC workshop for high school students	2019
Instructor, Oxford Centre for Doctoral Training	2018

Panelist, MIT EECS panel for incoming students	2022, 2023
Research Mentor for Undergraduates, seven students	2020-2023

SERVICE &
OUTREACH

Conference, Workshop or Session Organizer

TTIC Workshop on Incentives for Data Sharing & Collaborative Learning	2025
Session Chair, INFORMS	2023
Founding Organizer, MIT Women in EECS Research Summit	2019
Program Chair, MIT Global Startup Workshop (~1000 people)	2019
Lead Organizer, Oxford Conference for Females in Science & Technology	2018
Associate Editor & Session Organizer, IEEE ITS Conference	2018

Department Service & Outreach

PhD Admissions reviewer, MIT EECS	2019
Mentor, MIT EECS Graduate Application Assistance Program (GAAP)	2022
Organizer, MIT LIDS & IDSS DEI Book Club	2021
Mentor, MIT IDSS Graduate Application Assistance Program (GAAP)	2020
Organizer, MIT IDSS Graduate Application Assistance Program (GAAP)	2020
Instructor & Organizer, MIT workshop for high school students	2019

SELECTED
ACTIVITIES

Co-president, MIT AI Ethics Reading Group	2019-2021
DEI Chair, MIT IDSS Student Committee	2020-2021
Executive board, MIT LIDS Social Committee	2019-2020
Co-president, MIT Graduate Women in EECS	2019
Content Lead, MIT Global Startup Workshop	2018-2019

Tennis

Varsity player, Oxford University Tennis Team	2016-2018
Founder & captain, Princeton University Club Tennis Team	2014-2016

Miscellany

MIT Science Policy Congressional Visit Days	2023
Team lead, Princeton Robotics Club	2015
Web editor, The Daily Princetonian	2013-2014
News writer, The Daily Princetonian	2012-2013