SOCIAL MEDIA GOVERNANCE INITIATIVE
SPRING 2023 CONVENING

BEYOND MODERATION

MARCH 30 + MARCH 31

THE JUSTICE COLLABORATORY
Yale Law School

thank you

This convening was made possible through the generosity of an Oscar M. Ruebhausen grant by the Yale Law School.

Thank you for organizing support from

USC Marshall
School of Business
Neely Center for Ethical Leadership and Decision Making

We would like to thank the group of individuals who helped to contribute and shape this event

Molly Aunger
Lindsay Blackwell
Ravi Iyer
Matt Katsaros
Caroline Nobo
Sarita Schoenebeck
Joseph Seering
Sudhir Venkatesh

agenda

Day One

Thursday, March 30

Yale Divinity School, Old Refectory • 409 Prospect Street

9–9:15am

Opening Remarks

Matt Katsaros, Tracey Meares, Caroline Nobo, Tom Tyler

9:15–10:15am

Taking Stock of the Trust & Safety Landscape

SPEAKERS Sarah T. Roberts and Sudhir Venkatesh

The landscape of online trust and safety is quickly shifting. There are efforts to coalesce an industry of workers through organizations like the TSPA. An entire economy has emerged over the last few years providing T&S resources to platforms of all sizes. These companies have collectively raised hundreds of millions in venture capital to provide algorithms, intelligence, moderator tools, etc for platforms both big and small. Breakthroughs in AI have dramatically shifted the way companies approach safety on their platforms. In this opening conversation, Sudhir Venkatesh and Sarah T. Roberts will take stock of the current moment for online safety with a look towards what these shifts mean for the future and how the path we've taken to get here can inform where we go next.

10:15–11:45am

Emerging Research in Online Governance

SPEAKERS Jina Yoon, Monika Yadav, and Adina Gitomer
MODERATOR Amanda Menking

We have invited three young scholars on the cutting edge of research to share their ideas on digital governance. Some directly study digital patterns while examining social, political, and economic issues that are relevant to the online world.

12:00–12:45pm

Lunch

12:45–2:15pm

Design and Architecture of Healthy Online Spaces

SPEAKERS Delia Mocanu, Julia Kamin, and Jen Weedon
MODERATOR Ravi Iyer

Content moderation based systems have strict limitations in terms of both outcomes and legitimacy that suggest the need for alternatives. In this panel, we will hear about alternatives to content moderation based systems for encouraging a safer and improved information environment. We will hear from the Prosocial Design Network about evidence-based design patterns and from practitioners on solutions like Birdwatch, that provide let users to crowdsource misinformation mitigation.

agenda

Day One

Thursday, March 30

Yale Divinity School, Old Refectory • 409 Prospect Street

2:15–2:30pm

Break

2:30–3:30pm

Q&A with Community Moderators

SPEAKERS Sery, Binchlord, and Vanessa B
MODERATOR Sarah Gilbert

Many platforms like Reddit, Wikipedia and Discord rely extensively on many individuals who are passionate about their online communities. These volunteers do much of the work to build and maintain healthy communities across the internet. In this session, we hold a Q&A with three individuals who have been doing this work for many years to hear where they see future opportunities for building better online spaces.

3:30–3:45pm

Break

3:45–5:15pm

New Work on Community Driven Governance

SPEAKERS Amy Zhang, Eshwar Chandrasekharan, and Joseph Seering
MODERATOR Sanjay Kairam

The panel will discuss approaches to online governance which put communities at the center of governance structures, providing community members tools, training, and systems to build thriving and self-governing online communities.

5:30–7:30pm

Cocktails at The Study

We will make our way over to The Study - 1157 Chapel Street, New Haven - where we will have time to continue discussions, socialize, and enjoy some drinks.

agenda

Day Two

Friday, March 31

Sterling Law Building, Room 120 • 127 Wall Street

9–10:30am

Radical Futures for Social Media

SPEAKERS Jane Im, Ishita Chordia, and Shamika Klassen
MODERATOR Lindsay Blackwell

This panel seeks to look past the current moderation apparatuses constructed by major online platforms as the dominant approach towards online safety. Instead, these researchers are reimagining what it can look like to build social spaces online that promote safety and build trust.

10:30–10:45am

Break

10:45am–12:15pm

The Utility of Academic Research in Industry: A conversation on bridging the gap from research to practice

SPEAKERS Harsha Bhatlapenumarthy, Beth Goldberg and Alex Leavitt
MODERATOR Sarita Schoenebeck

Bridging the gap between industry and academia is of constant consideration for many scholars and practitioners alike. How can the insightful and interesting work emerging from the academy lead to more productive changes implemented at the platforms being studied and beyond? In this panel, a range of individuals from Product to Policy will explore with moderator Sarita Schoenebeck opportunities to improve the path for research to be put into practice.

12:15–1:15pm

Lunch

agenda

Day Two

Friday, March 31

Sterling Law Building, Room 120 • 127 Wall Street

1:15–3pm

Breakouts

Throughout the two days, we will be asking attendees to propose topics for closing breakout sessions. We will spend time breaking off into smaller groups where we can continue conversations on topics that arose over the two days focused on opportunities for collaboration and future work.

3–3:30pm

Breakouts Readouts

3:30–3:45pm

Closing Remarks

speakers

Taking Stock of the Trust & Safety Landscape

Thursday / March 30 / 9:15–10:15am

Sarah T. Roberts

Associate Professor, Gender Studies
UCLA

Sarah T. Roberts is an associate professor at UCLA (Gender Studies, Information Studies, Labor Studies), specializing in Internet and social media policy and culture, and the intersection of media, technology and society. She is the faculty director and co-founder of the UCLA Center for Critical Internet Inquiry (C2i2), co-director of the Minderoo Initiative on Technology & Power, and a research associate of the Oxford Internet Institute.

Roberts researches information work and workers, and is a leading global authority on "commercial content moderation," the term she coined to describe the work of those responsible for making sure media content posted to major commercial social platforms fit within legal, ethical, and the site's own guidelines and standards. She is frequently consulted on matters of policy, worker welfare, and governance related to content moderation issues and the broader social media landscape. Additionally, she has experience in industry as a consultant and researcher, most recently working as a Staff Researcher for Twitter in 2022.

She is a 2018 Carnegie Fellow and winner of the 2018 Electronic Frontier Foundation (EFF) Barlow Pioneer Award in recognition of her work on commercial content moderation. She is a 2023 honoree of the 100 Brilliant Women in AI Ethics.

Her book, *Behind the Screen: Content Moderation in the Shadows of Social Media*, was released in June 2019 and in paperback with a new preface in early 2022 (Yale University Press) and in a French edition, *Derrière les écrans* (Éditions La Découverte) in October 2020. A Mandarin edition will appear in early 2023.

Sudhir Venkatesh

William B Ransford Professor of Sociology
Columbia University

Sudhir Venkatesh has been on the faculty at Columbia since 1999. Periodically, he leaves the University. Recent excursions include a tenure as a Senior Advisor to the FBI Director, a stint managing teams at Facebook and Twitter, and two years running a Global MBA program in Berlin. Before becoming a parent, he studied the underground economy. Since then, he has turned to the tech sector. He is completing an ethnography of the world of online trust and safety. He is the recipient of a Junior Fellowship from Harvard's Society of Fellows. He is the narrator/creator of Sudhir Breaks the Internet, appearing on the Freakonomics Radio Network.

speakers

Emerging Research in Online Governance

Thursday / March 30 / 10:15–11:45am

Adina Gitomer

PhD Student
Network Science Institute, Northeastern University

I am a third-year PhD student in the Communication Media and Marginalization (CoMM) Lab at Northeastern University's Network Science Institute. My research focuses on the various ways that people — youth in particular — make use of social media for political participation and social change. While holding space for the harms caused by social media, I strive to shine a light on how it may be leveraged to further progressive goals and radical inclusivity. I tend to take a mixed methods approach, blending computational network and statistical techniques with qualitative analyses.

Monika Yadav

PhD Candidate, Sociology
Columbia University

Monika Yadav is a PhD candidate and Paul F. Lazarsfeld Fellow in the Department of Sociology at Columbia University. Her research work centers around studying the causes and consequences of political polarization, with a specific focus on the effects of online misinformation.

Jina Yoon

PhD Student, Paul G. Allen School of Computer Science & Engineering
University of Washington

Jina Yoon is a PhD student at the University of Washington researching ways to make people nicer to each other on the internet. Her work explores how to design social computing systems with empathy in mind to bridge misunderstandings and cultural differences online. Currently, her research focuses on supporting teen Discord moderators, information credibility, neurodivergent technology use, and diverse gaming communities.

Prior to starting her doctoral degree, Jina graduated from Brown University in 2018 where she studied Computer Science and Modern Culture & Media. She also has several years of industry experience from working at Microsoft and Riot Games. Jina's work has received support from sources including the NSF CSGrad4US Fellowship, ARCS Foundation, LEAP Fellowship, and NCWIT.

MODERATOR

Amanda Menking

Director of Programs
Trust & Safety Foundation (TSF)

Amanda Menking is the Director of Programs at the Trust & Safety Foundation (TSF) and at the Trust & Safety Professional Association (TSPA). She joined TSF and TSPA after spending a decade in academia as a PhD student, post-doc, and professor, where she researched bias, knowledge production, and safety in online communities and taught courses about design, research, and human-computer interaction. She's also a veteran of the Seattle tech scene, having worked for two start-ups and as a vendor at Microsoft prior to pursuing her Ph.D. in Information Science. Amanda is most interested in doing meaningful work with thoughtful, kind people with the aim of building more just and equitable worlds. At TSF, she's currently working on supporting the launch of the Trust & Safety Research Coalition and expanding programming to bring together the wide range of stakeholders working in and doing research about the trust and safety ecosystem.

speakers

Design and Architecture of Healthy Online Spaces

Thursday / March 30 / 12:45–2:15pm

Julia Kamin

Library Team Lead, Science Board Member
Prosocial Design Network

Julia Kamin is a researcher based in New York City. She works with organizations that leverage social science to improve civic discourse and mitigate toxic political polarization both on and off line. She is the Director of Research at Civic Health Project and the Library Team Lead at Prosocial Design Network. Julia received her PhD in Political Science from the University of Michigan, where her research focused on political psychology, social media and polarization.

Delia Mocanu

Data Scientist

Delia Mocanu is a Data Scientist based in San Francisco. Her work spanned quality, integrity, and adversarial behavior efforts at Twitter (Birdwatch), Facebook (News Feed), and Capital One (First Party Fraud). Her PhD research in Network Science was centered on understanding the dynamics of online engagement and information dissemination. She is an advocate for content agnostic solutions in mitigating challenges arising from inauthenticity and undue amplification.

Jen Weedon

Adversarial Planning
Niantic Labs

Jen is an expert in global information security threats and the use of intelligence to help keep users safe and secure online. She has spent her career as an analyst and leader, building teams to mitigate online harms across sectors. Most recently, she held leadership roles at Facebook establishing and supporting intelligence and investigative teams to illuminate adversarial and emerging threats, and helping design solutions with cross-functional teams in the product integrity space.

MODERATOR

Ravi Iyer

Managing Director
Psychology of Technology Institute at USC's Neely Center

Ravi is currently the Managing Director of the Psychology of Technology Institute, which is a project of USC's Neely Center. Before that, he led data science, research, and product teams across Facebook toward improving the societal impact of social media. He began that work focused on policy-based content moderation, but learned the painful lesson that a focus on content moderation was a dead end and often a distraction from the product design and algorithmic changes that would truly make a difference. Before working at Facebook, he helped co-found and build the initial algorithms for Ranker.com. He also has a PhD in Social Psychology from the University of Southern California and has published numerous articles on moral psychology.

speakers

Q&A with Community Moderators

Thursday / March 30 / 2:30–3:30pm

Sery

Bot Developer

Sery (he/him) has been both a streamer and a moderator on Twitch for 5 years. He's also the developer of Sery_Bot, a bot designed to moderate hateful activity on the platform.

Binchlord

Volunteer Moderator

binchlord (he/they) is a volunteer moderator who has been working primarily within LGBTQ+ spaces on Discord for the past 5 years. He has also been involved in some Reddit moderation and many of Discords moderation initiatives including launching new servers for brand partners, providing moderation for events, and generating guides on various aspects of moderation.

Vanessa B

Community Coordinator

Vanessa Brasfield has been a moderator of online communities and spaces since 2009. Their expertise and experiences cover Twitch, Twitter, and Discord for online gaming communities and moderation of activism education spaces on various forums and Facebook.

MODERATOR

Sarah Gilbert

Research Director
Citizens and Technology Lab

Dr. Sarah Gilbert (she/her/hers) is a postdoctoral associate at Cornell University and Research Director of the Citizens and Technology Lab. Her work focuses on understanding and designing healthy online communities, studying topics like what influences participation, how people learn in online communities, how volunteer moderators' labor impacts community governance, and how research in those areas can be done ethically. She also moderates the Reddit community, r/AskHistorians.

speakers

New Work on Community Driven Governance

Thursday / March 30 / 3:45–5:15pm

Eshwar Chandrasekharan

Assistant Professor, Department of Computer Science
University of Illinois at Urbana-Champaign

Eshwar Chandrasekharan is an Assistant Professor of Computer Science at the University of Illinois at Urbana-Champaign. His research builds a foundation for evaluating and improving approaches to online moderation, and developing new AI-backed sociotechnical systems. Dr. Chandrasekharan's work has appeared at high-impact conferences venues, and received considerable press coverage. He recently won a Facebook research award for his project on measuring healthy online behavior. Professor Chandrasekharan has worked with large-scale Internet platforms including Twitter, Reddit and Facebook, and his research has impacted their efforts to improve online governance. For example, he developed Crossmod, a new AI-backed moderation system that is currently deployed in an online community with over 14 million subscribers. Professor Chandrasekharan holds a B.Tech and M.Tech in Computer Science from the Indian Institute of Technology Madras, and a Ph.D. in Computer Science from the Georgia Institute of Technology.

Joseph Seering

Postdoctoral Scholar
Stanford University Computer Science Department

Joseph Seering is a postdoctoral scholar in Computer Science at Stanford University and an affiliated fellow at the Yale Law School SMGI, and will be joining the faculty at the KAIST School of Computing in Fall 2023. His work focuses on understanding the social and organizational dynamics of moderation systems on online social platforms in order to drive the development of new forms of social tools. His work begins with empirical analyses to understand community behaviors and challenges, and he translates the resulting findings into the design and development of new systems.

Amy Zhang

Assistant Professor, Allen School of Computer Science & Engineering
University of Washington

Amy X. Zhang is an assistant professor at University of Washington's Allen School of Computer Science and Engineering. Previously, she was a 2019-20 postdoctoral researcher at Stanford University's Computer Science Department after completing her Ph.D. at MIT CSAIL in 2019, where she received the George Sprowls Best Ph.D. Thesis Award at MIT in computer science. During her Ph.D., she was an affiliate and 2018-19 Fellow at the Berkman Klein Center at Harvard University, a Google Ph.D. Fellow, and an NSF Graduate Research Fellow. Her work has received a best paper award at ACM CSCW and a best paper honorable mention award at ACM CHI. She received an M.Phil. in Computer Science at the University of Cambridge on a Gates Fellowship and a B.S. in Computer Science at Rutgers University, where she was captain of the Division I Women's tennis team.

MODERATOR

Sanjay Kairam

Staff Scientist
Reddit

Sanjay Kairam is a computational social scientist studying online community interactions, community moderation, and governance processes. He currently works with the Community Team at Reddit, developing models, identifying strategies, and designing interventions to support successful outcomes for online communities. He previously led the Community Health Science team at Twitch. Sanjay received a PhD in Computer Science (2016), MA in Philosophy (2006), and BS in Mathematics (2006), all from Stanford University.

speakers

Radical Futures for Social Media

Friday, March 31, 9–10:30am

Ishita Chordia

Ph.D. Candidate, Information School
University of Washington

Bio: Ishita Chordia is a 5th year doctoral student at the University of Washington Information School. Her dissertation investigates how safety applications like Nextdoor and Citizen influence our sense of safety, community, and wellbeing. Her orientation as a researcher is deeply influenced by her experience working with radical groups like Berkeley Mutual Aid and the Policing Alternatives and Diversion Initiative. She was raised in the Jain tradition, and her passion for social justice is fueled by nonviolence, the central commitment of Jainism. Ishita has a Master's degree in Computer Science from the Georgia Institute of Technology and a Bachelor's degree in Economics from Duke University. She currently lives in East Atlanta with her fiance.

Jane Im

Ph.D. Candidate, School of Information
University of Michigan

Jane Im is a fifth-year Ph.D. candidate at the University of Michigan School of Information and the Division of Computer Science & Engineering. Jane's research aims to design safe social computing systems and tackles problems that range from interpersonal harm to institutional privacy issues. She focuses on the insight that these problems arise because software is designed without considering users' consent and power dynamics, in both user-to-user and user-to-company contexts. She combines systems-building and empirical studies, to design and evaluate systems' privacy controls and governance tools. Jane's research has won an Honorable Mention Award from ACM SIGCHI Conference on Human Factors in Computing and influenced designs of social media platforms. She completed her undergraduate studies in Business Management and Computer Science at Korea University.

Shamika Klassen

PhD Candidate, Information Science
University of Colorado, Boulder

Shamika is a person who is passionate about people and technology! After graduating from Stanford University with a degree in African and African-American studies in 2011, she served a year with AmeriCorps in NYC. She went on to study technology and ethics by developing technowomanism at Union Theological Seminary in the city of New York. There, she also created and developed the concept of a Tech Chaplain. She received her Master of Divinity from Union in 2017. She currently attends CU Boulder as a doctoral candidate in their Information Science department and is studying technology, ethics, and social justice issues.

MODERATOR

Lindsay Blackwell

Head of Trust & Safety
Sidechat

Lindsay Blackwell is a recognized expert in content moderation and social media governance. She is currently Head of Trust & Safety at Sidechat, a startup social media platform. In her previous roles, Lindsay led teams of policy experts, operations specialists, and machine learning engineers in efforts to better define, detect, and ultimately prevent harmful content and behavior. She served as lead researcher on Facebook's Hate Speech and Violence & Incitement teams, where she investigated issues of bias, fairness, and severity in automated enforcement systems, and pursued similar work as a senior researcher on Twitter's Content Health team.

speakers

The Utility of Academic Research in Industry: A conversation on bridging the gap from research to practice

Friday, March 31, 10:45am–12:15pm

Harsha Bhatlapenumarthy

Governance Manager / Lead, T&S Curriculum Working Group
Meta / TSPA

Harsha is an experienced Trust and Safety specialist having worked in the industry for the last 10 years based in India, Ireland and now in the US. As a volunteer at TSPA, she is leading the roadmap to develop and democratize the trust and safety curriculum. At Meta, she is leading programs focusing on the growth and efficiency of the Oversight Board. In her previous roles, she led multiple global initiatives to operationalize policies and scale enforcement, developed strategies and influenced product and policy direction to combat online harms such as sexual exploitation, fraud, scams, and misinformation.

Beth Goldberg

Head of Research & Development
Jigsaw (Google)

Beth Goldberg is the Head of Research and Development at Jigsaw, a unit within Google that studies online threats to open societies. She leads an interdisciplinary team of data scientists, ethnographers, and social scientists who work closely with academics and civil society. Her team's research focuses on the incentives driving online disinformation, harassment, and violent extremism, and the psychological interventions that can build resilience to these harms. Beth has led and published ethnographies of white supremacists and conspiracy theory purveyors, subsequently advising tech products and policies based on her findings. In 2022, she applied inoculation theory to build resilience to misinformation techniques across social media platforms for tens of millions of people. Beth has graduate degrees from Yale and was a fellow at Yale Law School Information Society Project, and received a BS from Georgetown School of Foreign Service.

Alex Leavitt

Senior Principal Researcher, Safety
Roblox

Alex Leavitt (they/them) is a social scientist who focuses on digital harm & safety issues in the technology industry. Most recently, they were a senior UX Researcher at Meta, working on global issues of trust & safety, such as misinformation, polarization, and information needs during crisis and for marginalized populations. While at Meta, Alex also led multiple academic collaborations, including a multi-million dollar grant program for academic social science research focused on "integrity" harms on social media platforms. In their next role, Alex will lead trust & safety research at Roblox, focusing on distributed moderation, digital civility, child safety, and VR/metaverse harms.

Moderator: Sarita Schoenebeck

Associate Professor
University of Michigan

Sarita Schoenebeck is an Associate Professor in the School of Information at the University of Michigan. Her research examines social and technical approaches to creating more safe and equitable experiences online. She is the recipient of the NSF CAREER award, Best Paper and Honorable Mention awards at CHI and CSCW, and ACM CSCW and UMSI Service awards. She is a Fellow at the Center for Democracy and Technology, the Yale Information Society Project, and the Yale Justice Collaboratory. Sarita received her PhD in Human-Centered Computing from Georgia Tech.