TRUST & SAFETY RESEARCH CONFERENCE

AGENDA

Thursday, September 26, 2024

8:00 – 8:30 **Registration and Continental Breakfast**

8:30 – 8:45 **Opening Remarks** (McCaw Hall)
Jeffrey T. Hancock, *Harry and Norman Chandler Professor of Communication, Stanford University*
Faculty Director, Stanford Internet Observatory and Social Media Lab

8:45 – 9:15 **Keynote** (McCaw Hall)
Camille François, *Associate Professor of Practice of International and Public Affairs,*
School of International and Public Affairs, Columbia University

9:15 – 9:45 **Break**

9:45 – 11:00 *Select one of the below, simultaneously held, sessions to attend.*

Research Presentation: 2024 Conference Proceedings of the Journal of Online Trust & Safety
McCaw Hall Mainstage

Lightning Talks:
Future of Search
Fisher Conference Center
Lane/Lyons/Lodato

2024 Conference Proceedings of the Journal of Online Trust & Safety (McCaw Hall Mainstage)
Research Presentations feature ten-minute presentations with time for questions.
Moderated by Victoria Cosgrove, Stanford University

- ○ A Survey of Scam Exposure, Victimization, Types, Vectors, and Reporting in 12 Countries
  Mo Houtti, Google
- ○ Factors Associated with Help-Seeking Among Online Child Sexual Abuse Material Offenders: Results of an Anonymous Survey on the Dark Web
  Tegan Insoll, Protect Children
- ○ Characteristics and Prevalence of Fake Social Media Profiles with AI-generated faces
  Kai-Cheng Yang, Northeastern University
- ○ Assessing Algorithmic Impact Assessments: Practitioners' Challenges and Needs
  Amar Ashar, Spotify & Berkman Klein Center and Henriette Cramer, Papermoon.AI

- Nuances and Challenges of Moderating a Code Collaboration Platform
  Margaret Tucker, GitHub
- Optimizing Data Portability: A Multi-Stakeholder Approach to Advancing Research on the Digital Information Environment
  Lama Mohammed, NYU

Lightning Talks: Future of Search (Fisher Conference Center Lane/Lyons/Lodato)
A special lightning talk session that looks at the Trust & Safety issues unique to search products and novel ways their harms can be reduced.
Co-Moderated by Ronald Robertson, Stanford Internet Observatory and Daniel Griffin, Trieve.ai

- LLMs and Web Search: Questioning the Impact on User Subjectivities and the Findability of Knowledge
  Nora Freya Lindemann, University of Osnabrück, Germany
- Examining The Influence of AI-Generated Search Results on User Behavior and Trust in Search Outputs
  Aleksandra Urman, University of Zurich
- Building Responsible Meta AI Search Systems
  Yvonne Lee, Meta
- New Contexts, Old Heuristics: How Young People in India and the US Trust Online Content in the Age of Generative AI
  Rachel Xu, Google Jigsaw
- Integrating External Expertise in Privacy into Product Roadmaps
  Mary Ioannidis, Google
- Good AI Legal Help, Bad AI Legal Help
  Margaret Darin Hagan, Stanford Legal Design Lab
- Searching for a New Search Algorithm
  Will Bryk, Exa
- AI Manipulation and Human Autonomy
  Inyoung Cheong, University of Washington
- The Future Of Trust In LLMs — Lessons From You.com
  Bryan McCann, You.com

11:00 – 11:30 **Break**

11:30 – 12:30 *Select one of the below, simultaneously held, sessions to attend.*

Lightning Talks: AI for Content Moderation
McCaw Hall Mainstage

Research Session: AI, Security, and Online Behavior
Fisher Conference Center Lane/Lyons/Lodato

Workshop: Academic Research Programs from YouTube and Google
Fisher Conference Center Barnes/McDowell/Cranston

AI for Content Moderation (McCaw Hall Mainstage)
Lightning talks feature five-minute, rapid fire presentations with time for questions.
Moderated by Samidh Chakrabarti, Stanford University

- Using LLMs for Labeling Task: Progress and Potential Risks
  Dave Willner, Stanford University
- GenAI/LLMs tech is Swiss Army Knife for Guardians of the Internet
  Shiwani Gupta, Google
- Navigating the Landscape of Automated Content Moderation: Insights from Ofcom's Research
  Pedro Freire, Ofcom - UK Office of Communications
- Utility of Generative AI vs Discriminative AI for Content Moderation
  Tom Siegel, TrustLab, Inc
- Identifying Best Practices for the Use of AI and Automation to Detect, Enforce, and Review Abusive Content and Behavior
  David Sullivan, Digital Trust & Safety Partnership
- Harmful YouTube Video Detection: A Taxonomy of Online Harm and MLLMs (GPT) as Alternative Annotators
  Claire Wonjeong Jo, University of California Davis
- Contested Pathways to Trusted and Safe AI through Third-Party Audits
  Chris Tenove, University of British Columbia
- Lessons Learned: Prepping for AI Automation in Trust & Safety Operations
  Jimin Lee, Change.org

AI, Security, and Online Behavior: Innovations and Impacts(Fisher Conference Center Lane/Lyons/Lodato)
Research Presentations feature ten-minute presentations with time for questions.
Moderated by Justin Hendrix, Tech Policy Press

- Emergent Cognitive Capabilities in Large Language Models
  Michal Kosinski, Stanford University
- Reach of Deceptive Online Networks is Limited but Highly Concentrated
  Ruth Elisabeth Appel, Stanford University
- Online Moderation in Games: How Intervention Affects Player Behavior
  Michael Alvarez, Caltech
- A Cross-Country Exploration of Security and Privacy Concerns and Advice in Ten Majority World Countries
  Collins W. Munyendo, The George Washington University
- Introduction to National Internet Observatory
  Jeffrey Gleason, Northeastern University

Academic Research Programs from YouTube and Google

(Fisher Conference Center Barnes/McDowell/Cranston)

Outside researchers bring valuable insights and expertise when conducting studies related to Google products, including Maps, Play, Search, Shopping and YouTube particularly when it comes to Trust & Safety challenges. This workshop has three main purposes:

1. We'll aim to demonstrate the current ways in which academics and think tanks can leverage existing Google and YouTube data;
2. We'll shed light on the topics and products that researchers have already requested data to explore;
3. We'll facilitate a conversation to better understand the needs of researchers in academia, and discuss how we can address those while simultaneously adhering to regulatory requirements and protecting user data privacy and security.

The workshop will also feature a discussion centered on end-to-end data needs (e.g., what are the types of data researchers are most interested in, what formats do they need that data in, what additional information do academics need in the form of documentation to make sense of raw data).

Facilitated by Brent Besson and Angela McKay, Google

12:30 – 1:45 **Lunch** (Ford Gardens)

1:00 – 1:30 **Lunch Plenary: Fireside Chat** (McCaw Hall)

Frank H. McCourt, Jr., *Executive Chairman of McCourt Global and Founder of Project Liberty*

Nathaniel Persily, *James B. McClatchy Professor of Law; Senior Fellow, Freeman Spogli Institute for International Studies; Co-Director, Cyber Policy Center*

1:45 – 2:45 *Select one of the below, simultaneously held, sessions to attend.*

Panel: Decentralized Platforms and Trust & Safety	**Lightning Talks: Regulation**	**Workshop: The Case for Open Source T&S Software**
McCaw Hall Mainstage	Fisher Conference Center Lane/Lyons/Lodato	Fisher Conference Center Barnes/McDowell/Cranston

Decentralized Platforms and Trust & Safety (McCaw Hall Mainstage)

Moderated by Mike Masnick, Copia Institute/Techdirt

- *Evan Prodromou, W3C*
- *Aaron Rodericks, Bluesky*
- *Samantha Lai, Carnegie Endowment for International Peace*
- *Liz Arcamona, Threads*

Regulation (Fisher Conference Center Lane/Lyons/Lodato)

Lightning talks feature five-minute, rapid fire presentations with time for questions.

Moderated by Daphne Keller, Stanford University

- Burden of Proof: Lessons Learned for Regulators from The Oversight Board's Implementation Work
  Manuel Parra Yagnam, Oversight Board
- A Risk-Based Approach to Age Assurance
  Cami Goray, University of Michigan
- Navigating New Frontiers: Article 21 of the Digital Services Act and the Future of Content Moderation
  Raphael Kneer, User Rights GmbH
- Regulating 'Trust and Safety' Under the Digital Services Act
  Linda Weigl, University of Amsterdam
- The EU Digital Services Act: Takeaways from One Year of Compliance
  Gerard de Graaf, European Commission, EU Office in San Francisco
- Latest Developments on Children's Rights Online
  James R. Marsh, Marsh Law
- Brussels' Effect Limited? Perspectives from Japan and Canada on Online Harm Legislation
  Toru Maruhashi, Meiji University
- Global Legislative Guidance for Online Spaces Free from Violence
  Katherine Townsend, Open Data Collaborative
- The Role of International Standards in Aligning Age Verification
  Alex Zeig, The Age Verification Providers Association
- Whose Free Speech?
  Belen Bricchi, Duke University

Stronger Together: The Case for Open Source T&S Software
(Fisher Conference Center Barnes/McDowell/Cranston)
Over the past few years, Trust & Safety teams worldwide have learned hard lessons that can now inform new efforts to develop a proper, purpose-built technology stack for Trust and Safety. The tools available today are designed and built for the internet of the 2010s but the online landscape has changed dramatically, bringing novel user behaviors and harm patterns. This space is now ripe for innovation and investment, with a clear opportunity to build and provide essential open "bricks" of this tooling stack. Join the T&S Tooling Consortium to learn how academics, researchers, and professionals can benefit from open source software and contribute to the movement.
Facilitated by Camille Francois, Eli Sugarman, Juliet Shen, Dave Willner, Rama Elluru

2:45 – 3:00 **Break**

3:00 – 4:00 *Select one of the below, simultaneously held, sessions to attend.*

Lightning Talks: Media Literacy	**Lightning Talks:** **Session 1: Mental Health & Well-being** **Session 2: Data Access**	**Lightning: Understanding Algorithms & Online Environments**
McCaw Hall Mainstage	Fisher Conference Center Lane/Lyons/Lodato	Fisher Conference Center Barnes/McDowell/Cranston

Media Literacy (McCaw Hall Mainstage)

Lightning talks feature five-minute, rapid fire presentations with time for questions.

Moderated by Angela Lee, Stanford University

- ○ Fact-checking Information Generated by a Large Language Model can decrease Headline Discernment
  Matthew DeVerna, Indiana University
- ○ Thoroughly Tracking the Takes and Trajectories of News Narratives from Trustworthy and Worrisome Websites
  Hans Hanley, Stanford University
- ○ Navigating Online Information Spaces: Strategies to Counteract Online Misinformation and Enhance Trust
  Sunny Liu, Stanford Social Media Lab
- ○ Reducing Misinformation Sharing at Scale using Digital Accuracy Prompt Ads
  Hause Lin, Massachusetts Institute of Technology
- ○ Correcting Misinformation with a Large Language Model
  Xinyi Zhou, University of Washington
- ○ The Effect of AI Labeling on Perceptions of Images
  Zeve Sanderson, NYU Center for Social Media & Politics
- ○ Collaboratively adding Context to Social Media posts reduces the Sharing of False News
  Thomas Renault, Université Paris 1 Panthéon - Sorbonne
- ○ Building Resilience to Misinformation in Communities of Color: Results from Two Studies of Tailored Digital Media Literacy Interventions
  Ryan Moore, Stanford University
- ○ How Scientific Retractions Enable Further Misinformation (and What to Do About it)
  Rod Abhari, Northwestern University
- ○ Labeling AI-Generated Content: Promises, Perils, and Future Directions
  Zivvy Epstein, MIT

Mental Health & Well-being (Fisher Conference Center Lane/Lyons/Lodato)

Moderated by Shubhi Mathur, Stanford Internet Observatory

- ○ 988 Suicide Crisis Services: How Online Discussions of Service Experiences can Improve Service Efficacy and Dissemination
  Nora Kelsall, Columbia Mailman School of Public Health, Department of Epidemiology
- ○ Building Bonds: Harnessing AI for Mental Health and Connection
  Yulia Sullivan, Baylor University
- ○ Exploring Interpretable Crisis Moderation Using LLMs and Diagnostic Inventories
  Karen Mosoyan, BlueFever

- Social Contagion and #Sadtok: The Risks and Benefits of Teens Self-diagnosing Mental Health Disorders from Social Media
  Ian Dull, ReD Associates
- Exploring the use of Virtual Reality for Content Moderators to Enhance Rapid Decompression from Occupational Stress during Short Wellness Breaks
  Natalie Campbell, TikTok

Data Access (Fisher Conference Center Lane/Lyons/Lodato)

Moderated by Zakir Durumeric, Stanford University

- Analyzing DSA Research Access
  Cameron Hickey, National Conference on Citizenship
- Data Sharing in K-12 EdTech Mobile Apps: Looking Under the Hood
  Lisa LeVasseur, Internet Safety Labs
- Making Social Media Safer Requires Meaningful Transparency
  Jeff Allen, Integrity Institute
- An Incentive-Compatible Framework for Online Surveys with Sensitive Questions
  John Ternovski, US Air Force Academy
- Behind the Curtain: Understanding the Datasets that Platforms Have and What You Can Learn with them
  Matt Motyl, Integrity Institute

Understanding Algorithms and Online Environments (Fisher Conference Center Barnes/McDowell/Cranston)

Moderated by Tracy Navichoque, Stanford University

- The Benefits of Optimizing for Quality Instead of Engagement
  Ravi Iyer, University of Southern California Neely Center
- Understanding Platform Users' Algorithmic Knowledge
  John Wihbey, Northeastern University
- The Cursed Equilibrium of Algorithmic Traumatization
  Cristiana Firullo, Cornell University
- AI Imaginaries Shape Identity Infusion and Digital Futures
  Bu Zhong, Hong Kong Baptist University
- User or Algorithm? Investigating what drives Congenial and Problematic Consumption on YouTube
  Muhammad Haroon, University of California, Davis
- Homogenizing Harm Across Realities: A Comparative Study of Web 2.0 and XR Community Guidelines.
  Kyooeun Jang, University of Southern California

4:00 – 5:30 **Happy Hour and Poster Session** (Ford Gardens)

During happy hour, poster presenters will be on hand to discuss their research.

- ○ A Self-Eating Snake: The Challenges for Constituent Processes in the Social Media Era and What to Learn from the Chilean failure. *José Acevedo, Rutgers University*
- ○ Incentivizing News Consumption on Social Media Platforms Using LLMs and Realistic Bot Accounts. *Hadi Askari, University of California, Davis*
- ○ More Than Meets the Eye: Exploring the Efficacy of Media Provenance for Synthetic Content Analysis. *Wilson Chen, University of Washington*
- ○ The Role of Narrative in Misinformation Games. *Nisha Devasia, University of Washington*
- ○ Looking Back to Move Forward: How 20 years of Empirical T&S Research Unveils a Better Path Forward. *Michael Bochkur Dratver, The Justice Collaboratory at Yale Law School*
- ○ Assessing US Military Information Operations: An Exploration into USCENTCOM J39. *Divya Ganesan, Stanford University*
- ○ Steps Toward Reliably Measuring Secondary Trauma Among Content Moderators. *Alexandra Gonzalez, Cornell University*
- ○ The Digitalized Space and Social Inequality in China. *Niko Han, Peking University and University of Oxford*
- ○ How to See 1000 Images: Innovative Image Analysis Methods for Problematic Information Studies. *Nina Lutz, University of Washington*
- ○ Mapping the Digital Divide of VPNs: How VPN providers Fail to Protect and Reach the MENA Region. *Mina Mohammadi, Oxford Internet Institute, University of Oxford*
- ○ Digital Footprint or a Personal Right: Understanding the Opinions and Attitudes Toward Data Privacy Among Internet Users in the United States. *Lukasz Niparko, University of Nebraska at Lincoln*
- ○ Beyond the Regular Benchmarks: Evaluate Large Foundation Models' Potential Usage in Adversarial Activities. *Tu Ouyang, Case Western Reserve University*
- ○ Crafting Synthetic Reality: Examining Visual Realism and Misinformation Potential of Photorealistic AI-Generated Images. *Qiyao Peng, University of California, Santa Barbara*
- ○ Meta "Meta Papers": Analyzing Framings and Coverage of the US 2020 Election Project. *Joseph Schafer, University of Washington*
- ○ Bridging Nodes and Narrative Flows: A Graph-Theoretic Analysis of Telegram's Disinformation Ecosystem. *Devang Shah, SimPPL*
- ○ The Double-Edged Sword of User Agency: Empowerment and Risk in Decentralized Social Media Platforms. *Aneesh Shamraj, SimPPL*
- ○ A Legal and Ethical Analysis of the Use of AI in Journalism: A Case Study on the Financial Times. *Zoey Soh, University College London*
- ○ Yellowstone Is Not Erupting: Rumor Correction and How TikTok Users Made Sense of a Small-Scale Hoax. *Julie Vera, University of Washington*
- ○ Clearing the Haze: Examining the Impact of the EU Digital Services Act on Content Moderation Transparency. *Alessia Zornetta, UCLA School of Law*

4:30 – 5:30 **Moderated Content Live!** (Fisher Conference Center Lane/Lyons/Lodato)
A live recording of the Stanford Law School and Cyber Policy Center's *Moderated Content*, a podcast about content moderation, with Evelyn Douek and Alex Stamos. The recording will feature special guests engaging in wonky conversation about the regulation of what you see, hear and do online.

Friday, September 27, 2024

8:00 – 8:30 **Registration and Continental Breakfast**

8:30 – 8:45 **Opening Remarks** (McCaw Hall)

8:45 – 9:30 **Keynote** (McCaw Hall)
Arvind Narayanan, Professor of Computer Science and Director of the Center for Information Technology Policy, Princeton University

9:30 - 10:00 **Break**

10:00 – 11:00 *Select one of the below, simultaneously held, sessions to attend.*

Panel: Abuse Mitigation in Interoperable End-to-End Encrypted Messaging	**Lightning Talks: Building Trust & Safety**	**Workshop: Research with the Meta Content Library & API**
McCaw Hall Mainstage	Fisher Conference Center Lane/Lyons/Lodato	Fisher Conference Center Barnes/McDowell/Cranston

Abuse Mitigation in Interoperable End-to-End Encrypted Messaging
(McCaw Hall Mainstage)
Moderated by Alissa Cooper, Knight-Georgetown Institute

- *Julia Len, MIT*
- *Jenny Blessing, University of Cambridge*
- *Dick Brouwer, WhatsApp*
- *Stephan Somogyi, Google*

Building Trust & Safety (Fisher Conference Center Lane/Lyons/Lodato)
Lightning talks feature five-minute, rapid fire presentations with time for questions.
Moderated by Dave Willner, Stanford University

- Building Trust and Safety on Facebook
  Lluis Garcia Pueyo, Meta
- "There's an Information war going on": Understanding Motivations of Content Abusers
  Chelsea Johnson, LinkedIn
- Safety Operations: Preventing Illegal and Harmful Behavior at Scale
  Tom Thorley, GitHub
- Striking the Right Balance between Access to Information and User Safety: A Case Study of SafeSearch BLUR Design, Launch and Measurement from Trust & Safety Perspective
  Elzbieta Brzoz, Google
- Digital Responses to Crises: An Action Plan for Platforms and CSOs Confronting Online Threats
  Rachelle Faust, National Democratic Institute

- ○ The 7 Data Points That Increase Content Moderation Efficiency by 3x
  Hill Stark, ActiveFence
- ○ Leveraging User Surveys to Track Online Experiences: Lessons from 5 Waves of Neely Index Data Collection
  Juliana Schroeder, University of California, Berkeley
- ○ XR is not Social Media. And that's a problem.
  Michael Karanicolas, UCLA Institute for Technology, Law & Policy
- ○ A Strategic Approach to Navigating Integrity in Immersive Technologies
  Kelly Lundy, Meta

Research with the Meta Content Library & API
(Fisher Conference Center Barnes/McDowell/Cranston)
This session will introduce Trust and Safety researchers to the Meta Content Library and demonstrate new data fields and functionalities available within the latest version of Meta Content Library's user interface (UI) and API. The hands-on demonstration will deploy research use cases to demonstrate how the data from the API and the UI can be used to shed light on questions relevant to trust and safety researchers. We will also provide an overview for how individuals and research teams can apply for access to these tools, as well as provide an opportunity for attendees to share feedback with us about our products and services.

Meta Content Library gives researchers comprehensive access to posts, videos, photos, and reels posted to public Pages, Groups, and Events on Facebook. For Instagram, the library includes content from public posts, albums, videos, and photos from personal, creator and business accounts. Robust metadata about each of these data types (e.g. view count, reshares, reactions, etc.) enables in-depth quantitative and qualitative analysis.

Facilitated by Yair Rubenstein and Phil Edwards, Meta

11:00 – 11:30 **Break**

11:30 – 12:30 *Select one of the below, simultaneously held, sessions to attend.*

Lightning Talks: Polarization and Elections	**Lightning Talks: Digital Threats**	**Lightning Talks: Using AI to generate harmful content**
McCaw Hall Mainstage	Fisher Conference Center Lane/Lyons/Lodato	Fisher Conference Center Barnes/McDowell/Cranston

Polarization and Elections (McCaw Hall Mainstage)
Lightning talks feature five-minute, rapid fire presentations with time for questions.
Moderated by Izzy Gainsburg, Stanford Polarization and Social Change Lab

- ○ Exploring the Interaction of Trust in Science and Vaccine Hesitancy
  Pranav Goel, Northeastern University
- ○ Otherization via Disinformation: Text, Context and the Baha'is in Iran
  Fares Hedayati, Baha'i International Community

- Foreign Information Manipulation and Interference: Lessons from the EU Elections
  Rachele Gilman, Global Disinformation Index
- Understanding Online Hate Speech in Context
  Thomas Davidson, Rutgers University
- The Musk Effect: Changes in Twitter's Misinformation and Partisan Composition
  Burak Ozturan
- Measuring the Effects of Harmful Social Media Narratives in Conflict Settings
  Bailey Ulbricht, Stanford Law School
- Where Do Election Deniers Get their News?
  Hong Qu, Northeastern University
- Revolutionary Rhetoric: Moderating the fine line between Patriotism and Dangerous Speech
  Cathy Buerger, Dangerous Speech Project
- Election Misinformation: A Case Study from Shasta County California
  Paul Spencer, Disability Rights California
- Hate Speech and Misinformation on WhatsApp: Insights from a Large Data Donation Program in India & Brazil
  Kiran Garimella, Rutgers University

Digital Threats(Fisher Conference Center Lane/Lyons/Lodato)

Moderated by Olivia Natan, UC Berkeley

- Beyond Borders: Lessons from Ghana's Fight Against Child Online Exploitation and CSAM - Implications for Africa: Challenges & Successes
  Emmanuel Adinkrah, Ghana Internet Safety Foundation
- Assessing the Gendered Dimension of the Nexus between Data-Exploiting Cyberattacks and the Proliferation of Harmful Content Online
  Pavlina Pavlova, New America
- Mean Megaphones: Criminal Advantages in Social Media Amplification
  Lucas Almeida, Northeastern University
- Sextortion: Prevalence and Correlates in 10 Countries
  Rebecca Umbach, Google
- AI and Your Amygdala: Partners in Cyber-Crime
  Scott Hellman, FBI
- Keeping Online Marketplaces Safe in the World of AI
  Sarika Oak, Udemy
- Uncovering a DPRK Hiring Scheme Targeting Remote IT Jobs
  Benjamin Racenberg, Nisos
- Improving the Governance of Online Platforms with Truth Warrants
  Swapneel Mehta, BostonU and MIT
- Mapping the Network Maze: Identifying and Tracking Coordinated Spam and Scam Campaigns on Social Media
  Fabio Giglietto, University of Urbino

- Authentic or Artificial? AI's Impact on Verification
  Steven Chua, Google

Using AI to generate harmful content (Fisher Conference Center Barnes/McDowell/Cranston)
Lightning talks feature five-minute, rapid fire presentations with time for questions.
Moderated by Elena Cryst, Stanford University

- From Open-Source to Primetime: The Making of an AI News Anchor
  Maty Bohacek, Stanford University
- Artificial Deception: How Bad Actors Leverage AI to Spread Disinformation
  Sarah Brandt, NewsGuard
- Generative AI Misuse: A Taxonomy of Tactics and Insights from Media Reports
  Rachel Xu, Google Jigsaw
- Generative Propaganda: Evidence of AI's Impact from a State-Backed Disinformation Campaign
  Patrick Warren, Clemson University Media Forensics Hub
- Generative AI and the Changing Business of Propaganda
  Madeleine Daepp, Microsoft Research
- The Future of Trust & Safety: Gen AI and Alleged Reality
  Amie White, Hinge

12:30 – 1:30 **Lunch**
Attendees are welcome to enjoy their lunch in the gardens or bring their lunch to the breakout.

1:30 – 2:30 *Select one of the below, simultaneously held, sessions to attend.*

Panel: Conceptualizing Trust & Safety: Interactions between Core Functions & Research	**Research Presentations: Child Safety**	**Workshop: Teaching Trust & Safety**
McCaw Hall Mainstage	Fisher Conference Center Lane/Lyons/Lodato	Fisher Conference Center Barnes/McDowell/Cranston

Conceptualizing Trust & Safety: Interactions between Core Functions & Research
(McCaw Hall Mainstage)
Moderated by Rebecca Umbach, Google

- *Katie Harbath, Duco Experts*
- *Angela McKay, Google*
- *Savannah Badalich, Discord*
- *Brian Fishman, Cinder*

Child Safety (Fisher Conference Center Lane/Lyons/Lodato)
Research Presentations feature ten-minute presentations with time for questions.
Moderated by David Thiel, Stanford Internet Observatory

- Trust, Safety, Trafficking: How to Stop CSAM at the Source
  Erin Kilbride, Human Rights Watch
- Beyond the West: Exploring the Impact and Regulation of Non-Consensual Image Disclosure Abuse (NCIDA) in Non-Western Contexts
  Amna Batool, School of Information, University of Michigan, Ann Arbor
- Why Johnny Spent Money on Roblox: Manipulative Design in the Roblox Monetization Ecosystem
  Yael Eiger, University of Washington
- Practical Solutions to Combat Financial Sextortion: Insights from Infiltrating the Secret Marketplace of Scammers and Collaborating with Leading Tech Platforms
  Avi Jager, ActiveFence
- A Digital Pandemic: Uncovering the Role of Yahoo Boys in the Rise of Financial Sextortion Targeting Minors
  Paul Raffile, Network Contagion Research Institute
- The Financial Sextortion of Minors: An overview of a Study Exploring Three Years of CyberTipline Reports and the Rise of Financial Sextortion
  Tim O'Gorman, Thorn

Teaching Trust & Safety
(Fisher Conference Center Barnes/McDowell/Cranston)
This session will allow participants to share information about courses they teach related to trust and safety and hear from instructors around the country about innovations in teaching in this field.

- *Alexios Mantzarlis, Cornell Tech*
- *Laura McLester, University of Alabama at Birmingham*
- *Juliet Shen, Columbia University*
- *Olga Belogolova, Johns Hopkins University*
- *Anthony Mensah, Stanford Engineering*
- *Étienne Brown, San Jose State University*
- *Jeff Lazarus, Trust & Safety Professional Association*

Facilitated by Elena Cryst, Stanford University

2:30 – 2:45 **Break**

2:45 – 3:15 **Closing Remarks**

3:15 – 4:00 **Happy Hour** (Ford Gardens and McCaw Hall)