



“What is Safety?”

Building Bridges Across Approaches to Digital Risks and Harms

Ashley Marie Walker
amwalker@google.com
Google
New York City, USA

Michael Ann DeVito
m.devito@northeastern.edu
Northeastern
Boston, USA

Karla Badillo-Urquiola
kbadill3@nd.edu
University of Notre Dame
Notre Dame, USA

Rosanna Bellini
rfb242@cornell.edu
Cornell Tech
New York City, USA

Stevie Chancellor
steviec@umn.edu
University of Minnesota
Minneapolis, USA

Jessica L. Feuston
jesfeuston@meta.com
Meta
New Haven, USA

Kathryn Henne
kathryn.henne@anu.edu.au
Australian National Univ.
Canberra, Australia

Patrick Gage Kelley
patrickgage@acm.org
Google
New York City, USA

Shalaleh Rismani
shalaleh.rismani@mail.mcgill.ca
McGill University
Montreal, Canada

Renee Shelby
reneeshelby@google.com
Google Research
San Francisco, USA

Renwen Zhang
r.zhang@nus.edu.sg
National Univ. of Singapore
Singapore

CCS CONCEPTS

• Security and privacy → Human and societal aspects of security and privacy; • Human-centered computing → HCI theory, concepts and models.

KEYWORDS

Safety, digital safety, risks, harms, what is safety

ACM Reference Format:

Ashley Marie Walker, Michael Ann DeVito, Karla Badillo-Urquiola, Rosanna Bellini, Stevie Chancellor, Jessica L. Feuston, Kathryn Henne, Patrick Gage Kelley, Shalaleh Rismani, Renee Shelby, and Renwen Zhang. 2024. “What is Safety?” Building Bridges Across Approaches to Digital Risks and Harms. In *Companion of the 2024 Computer-Supported Cooperative Work and Social Computing (CSCW Companion '24)*, November 9–13, 2024, San Jose, Costa Rica. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3678884.3681824>

1 MOTIVATION

Safety is a frequently discussed and increasingly important topic in Human-Computer Interaction (HCI) and one historically well-represented at CSCW [30]. However, CSCW is a diverse community that engages multiple fields and perspectives; therefore, the community employs multiple, sometimes contradictory, definitions of technologically-mediated “safety” both at the conference and in the field overall. While different concepts of safety may have clear meaning within a computing subfield, the overall importance of safety as a topic to HCI and the need to communicate clearly across subfields and to other disciplines underscores the importance of exploring how differing approaches to defining and operationalizing safety compare and potentially build on/communicate with each other. Moreover,

by comparing, contrasting, and synthesizing CSCW’s multiple approaches to safety, we can better understand and account for safety as a whole. A core goal of this workshop is to make progress towards a collective understanding of safety. Without clarity on such an important concept, there will be wasted efforts in intervention design for mitigating harms for impacted users and communities, and we squander the potential for effective, academic collaboration. In this workshop, we take the first step towards initiating a conversation between different HCI perspectives on safety and mapping out the multitude of different goals, aspirations, and working definitions of what academic research on and about safety looks like.

Multiple approaches to safety exist at CSCW, focused on various dimensions of physical, psychological, and sociotechnical aspects of safety as it intersects with computing technologies and approaches to enact it. For example, HCI work on “digital safety” examines online behaviors that may result in a wide range of interpersonal harms, including forms tech-facilitated violence (e.g., cyberbullying [2], gender violence [1, 3], racism and misogyny [33], exposure to dangerous content (e.g., harmful social media challenges, content encouraging eating disorders), and how the design of platforms and the logics baked in create conditions of unsafety that disproportionately affect historically marginalized communities, including transgender and disabled communities [22, 38]. This literature also focuses on ways to foster “safer” digital experiences through community-centered practices [20, 25, 35] and design interventions [9, 23] focused on recognition of interdependence and care (e.g., [40, 41]). Some of this work aligns with how tech companies think of “Trust & Safety” organizations, including topics like child safety, misinformation, illegal content, violence and extremism, and more. In parallel, there is ample work in HCI that deals with topics related to safety but themselves may not self-describe as safety, such as mental health crisis and self-injurious behaviors (e.g. [8]).

Another key thread of HCI research focuses on the perceptual side of safety, such as the sensemaking and folk theorization that actually drives user decision making [13]. This emerging body of literature examines safety from an affective point-of-view [14, 27, 42] offering important insights into the user side of safety questions.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CSCW Companion '24, November 9–13, 2024, San Jose, Costa Rica

© 2024 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1114-5/24/11.

<https://doi.org/10.1145/3678884.3681824>

Nevertheless, this approach to safety often fails to engage with technical realities of how digital technologies are developed, which are shaped by the complex interplay between users, platforms, and governments/regulatory bodies. For example, while folk theorization and community-based approaches provide crucial context on user experiences and needs, they sometimes capture broad user anger, irritation, and frustration as opposed to contextual experiences of harm when interacting with novel and/or unsafe-seeming systems [13, 15]. This may look like users and non-users calling for platforms to “do better” to implement safer experiences, which may be out of touch with the practical limits of how platforms function (e.g., [17]), or how optimizing for one safety concern may lead to other problems in disciplining technologies to developer prescribed safety and content moderation policies (e.g., [24]).

Finally, an entirely different set of definitions and approaches exist related to machine learning, responsible AI (artificial intelligence) development, and human-AI interaction within HCI. ML systems are increasingly embedded in our digital ecosystems (e.g., social media platforms, recommendation systems, apps, stand alone text and image generation tools) necessitating novel research on what safety means when these systems mediate interpersonal and social interactions [19, 28]. Against the backdrop of regulatory calls for “safe AI” (e.g., [11, 26]), HCI research in this space focuses on identifying and reducing potential harmful social impacts of AI technologies for direct users and broader communities [39], including outlining unsafe design patterns [10] and differential experiences of harm and safety in AI technologies [12].

Much HCI work in this space also focuses on studying developer and practitioner culture and improving practices to responsibly develop “safe” AI systems (e.g., [31]). This includes research that draws on the longstanding field of safety engineering [5, 29]) exploring how it provides an avenue for systematizing the identification and mitigation of social and ethical risks [36, 37]. While this body of work enriches understanding of sociotechnical approaches to establishing a safety culture and examining harms from AI system, research on human-AI interaction would benefit from deeper engagement with equity-oriented HCI approaches, including engaging historically marginalized communities to understand how technology mediates experiences of safety (e.g., [21, 38]). This deeper engagement is critical in light of the growing conversation on AI safety as it emphasizes how unsafe experiences/events emerge from complex interactions between a technology and corresponding social structures shaped by axes of inequality.

Additionally, HCI contains a wealth of knowledge in areas that relate to and impact digital safety. For example, the HRI subfield and work on assistive and medical technology provide insight into physical safety issues, while ergonomics-related and human factors research provides insight into organizational and procedural safety. While the primary focus of this workshop is digital safety, we welcome participation from other areas of expertise within CSCW.

1.1 The Need to Bridge Different HCI Approaches to Safety

The different and distinct approaches to safety in HCI carry important meaning in their own contexts. However, strengthening shared understandings of the multiple meanings of safety in HCI has important theoretical and methodological benefits. Larger projects, such

as building frameworks and engaging in policy and regulatory work, require us to look across these definitions more closely. Doing so enables us to work more effectively towards a shared understanding of digital safety which matches user perception to technical and regulatory reality. This pragmatic approach can also account for increasingly-important safety questions in AI technologies that may be very different from what concerns users about AI.

To accelerate our progress towards such a shared understanding, we propose a CSCW workshop to gather key safety researchers in the CSCW community to compare, contrast, and synthesize their experiences. This is a crucial first step towards a more robust framework for discussing and doing work on online safety across multiple HCI-related subfields. We do so with the following objectives:

- Create a collaborative space for scholars, industry practitioners, and policy makers interested in studying safety to exchange interdisciplinary ideas and practices.
- Identify challenges and opportunities within the digital safety field and begin to synthesize an approach for the future.
- Brainstorm how HCI methodologies and perspectives contribute to the shifting and growing conversation about safety of emerging technologies.

2 ORGANIZERS

Our organizing committee is comprised of researchers with extensive experience researching digital safety related topics from multiple areas noted above. Multiple organizers have prior experience running successful CSCW and HCI workshops, including hybrid formats on safety and trauma-informed design [6], justice-oriented perspectives [4], the needs of vulnerable populations [16, 32], conducting sensitive research [7], and designing AI [34].

2.1 Organizer Bios

- **Ashley Marie Walker, Google:** Walker is a researcher on the Trust & Safety team within Google. Their work focuses on understanding the systems of marginalization and subsequent risks in sociotechnical systems and how empirically informed policy design can help with mitigation.
- **Michael Ann DeVito, Northeastern University (she/her):** Dr. DeVito is an expert in user perception and adaptation in the context of visibility, marginalization, and online abuse/harassment. Her work is aimed at making sociotechnical systems safer and more understandable for end users.
- **Karla Badillo-Urquiola, University of Notre Dame (she/her/ella):** Dr. KB is a Clare Boothe Luce Assistant Professor of Computer Science & Engineering. She directs the EPOCH Research laboratory, designing technology-driven solutions that empower people and protect the well-being of youth in marginalized communities (e.g., foster care, juvenile justice, Global South).
- **Rosanna Bellini, Cornell Tech (she/her):** Dr. Bellini is a Post-doctoral Scholar. She specializes in mitigating technology-enabled abuse in intimate partner violence contexts by working directly with abusive partners and around instances of financial abuse.
- **Stevie Chancellor, University of Minnesota (she/her):** Chancellor is Assistant Professor of Computer Science & Engineering. She researches human-centered AI for predicting and evaluating

physically dangerous mental health behaviors in social media (e.g. eating and substance use disorders, suicidality).

- **Jes Feuston, Meta:** Dr. Feuston is an integrity and well-being researcher who specializes in qualitative methods including ethnography, participatory design, and usability evaluations, to develop products that support safe, responsible, and human-centered experiences on social media.
- **Kathryn Henne, Australian National University (she/her):** Dr. Henne is Professor and Director of RegNet, the ANU School of Regulation and Global Governance, where she leads the Justice and Technoscience Lab. Her research focuses on how to regulate technology to promote health and public safety. Her recent work examines anti-violence apps, automated decision-making in welfare delivery systems, and predictive policing technologies.
- **Patrick Gage Kelley, Google:** Kelley has worked in Privacy, Counter-Abuse, and Trust & Safety within Google. He has worked with at-risk users (content creators, people on political campaigns, youth) and is exploring how AI can be responsibly designed and understood to help people make safer AI decisions.
- **Shalaleh Rismani, McGill University (she/her):** Shalaleh Rismani is a Ph.D. candidate at McGill University and Mila in Montreal, Canada. Her research leverages perspectives at the intersections of safety engineering, responsible AI development, and human-computer interaction to develop systematic approaches to evaluate and mitigate harms from algorithmic systems.
- **Renee Shelby, Google Research (she/her):** Dr. Shelby studies the social impacts of technology, particularly the relationships between social inequality, technology design, and governance. Her work identifies potential harms and modes of harm reduction from algorithmic systems by centering the experiences of people in different historically marginalized communities.
- **Renwen Zhang, National University of Singapore (she/her):** Dr. Zhang is Assistant Professor of Communications and New Media. Her research focuses on technology-mediated communication and human-machine communication, with emphasis on social support, well-being, and mental health.

3 PUBLISHING WORKSHOP PROCEEDINGS

Given the potentially sensitive nature of the conversations of this workshop we will invoke Chatham House Rules and not host proceedings of participant submissions on our public website. We will allow for private sharing within the group to facilitate virtual and asynchronous connections, and publish summarized notes from the discussion to the website after CSCW.

4 HYBRID WORKSHOP STRUCTURE

For this one-day workshop, we are planning a hybrid event so as best to reach the widest range of participants who may be interested in this topic. To limit the logistical complications that arise with running synchronous in-person and virtual sessions, we will structure the bulk of the virtual engagement to take place asynchronously, with optional synchronous engagement opportunities for those who are unable to attend CSCW in-person.

4.1 Pre-Workshop Plan

In the two weeks leading up to the workshop, we will invite participants to an event Padlet, to allow participants to volunteer questions

as seed for the workshop discussions. One week before CSCW, we will open an event-specific Slack, populating Slack channels from common themes in Padlet responses. Members of the organizing committee will be present in all Slack channels to seed discussions and foster community dynamics and discussion. This model has been successfully in multiple iterations of the Queer in HCI Special Interest Group at CHI [18]. To ensure workshop's activities are accessible for all participants, we will ask invited participants about their accessibility needs and accommodate them.

4.2 Post Workshop Plans

After CSCW, we will continue to stand up the event-specific Slack for an additional 4 weeks. The organizing committee will provide topic specific report outs from the small group discussions sessions in-person at CSCW no more than 2 weeks after the event. For our report out (and indeed for this entire event), we will be invoking Chatham House Rules. We will be reporting out the major discussion points from the different in-person and virtual channels in the workshop, but not providing direct attribution of any commentary, reflecting the potentially sensitive nature of the conversations.

REFERENCES

- [1] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Joshua Gracie, Munmun De Choudhury, Pamela J. Wisniewski, and Gianluca Stringhini. 2022. Understanding the Digital Lives of Youth: Analyzing Media Shared within Safe Versus Unsafe Private Conversations on Instagram. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 148, 14 pages. <https://doi.org/10.1145/3491102.3501969>
- [2] Zahra Ashktorab and Jessica Vitak. 2016. Designing Cyberbullying Mitigation and Prevention Solutions through Participatory Design With Teenagers. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). Association for Computing Machinery, New York, NY, USA, 3895–3905. <https://doi.org/10.1145/2858036.2858548>
- [3] Rosanna Bellini. 2023. Paying the Price: When Intimate Partners Use Technology for Financial Harm. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 872, 17 pages. <https://doi.org/10.1145/3544548.3581101>
- [4] Rosanna Bellini, Debora de Castro Leal, Hazel Anneke Dixon, Sarah E Fox, and Angelika Strohmayer. 2022. "There is No Justice, Just Us": Making Mosaics of Justice in Social Justice Human-Computer Interaction. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI EA '22). Association for Computing Machinery, New York, NY, USA, Article 97, 6 pages. <https://doi.org/10.1145/3491101.3503698>
- [5] Carl S Carlson. 2012. *Effective FMEAs: Achieving Safe, Reliable, and Economical Products and Processes Using Failure Mode and Effects Analysis*. Vol. 1. John Wiley & Sons.
- [6] Rosanna Bellini Tawfiq Ammari Michael Ann DeVito Bryan Semaan and Nazanin Andalibi Casey Randazzo, Carol F. Scott. 2023. Trauma-Informed Design: A Collaborative Approach to Building Safer Online Spaces. <https://sites.google.com/corp/view/traumainformedcscw>
- [7] Stevie Chancellor, Nazanin Andalibi, Lindsay Blackwell, David Nemer, and Wendy Moncur. 2019. Sensitive Research, Practice and Design in HCI. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI EA '19). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3290607.3299003>
- [8] Stevie Chancellor, Zhiyuan Lin, Erica L. Goodman, Stephanie Zerwas, and Munmun De Choudhury. 2016. Quantifying and Predicting Mental Illness Severity in Online Pro-Eating Disorder Communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (San Francisco, California, USA) (CSCW '16). Association for Computing Machinery, New York, NY, USA, 1171–1184. <https://doi.org/10.1145/2818048.2819973>
- [9] Janet X. Chen, Allison McDonald, Yixin Zou, Emily Tseng, Kevin A Roundy, Acar Tamersoy, Florian Schaub, Thomas Ristenpart, and Nicola Dell. 2022. Trauma-Informed Computing: Towards Safer Technology Experiences for All. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 544, 20 pages. <https://doi.org/10.1145/3491102.3517475>
- [10] Ishita Chordia, Lena-Phuong Tran, Tala June Tayebi, Emily Parrish, Sheena Erete, Jason Yip, and Alexis Hiniker. 2023. Deceptive Design Patterns in Safety Technologies: A Case Study of the Citizen App. In *Proceedings of the 2023 CHI Conference*

- on *Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 193, 18 pages. <https://doi.org/10.1145/3544548.3581258>
- [11] European Commission. 2021. The Artificial Intelligence Act. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
- [12] Isha Datey, Hanan Khalid Aljasim, and Douglas Zytka. 2022. Repurposing AI in Dating Apps to Augment Women's Strategies for Assessing Risk of Harm. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, Taiwan) (CSCW'22 Companion). Association for Computing Machinery, New York, NY, USA, 150–154. <https://doi.org/10.1145/3500868.3559472>
- [13] Michael Ann DeVito. 2021. Adaptive Folk Theorization as a Path to Algorithmic Literacy on Changing Platforms. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 339 (oct 2021), 38 pages. <https://doi.org/10.1145/3476080>
- [14] Michael Ann DeVito. 2022. How Transfeminine TikTok Creators Navigate the Algorithmic Trap of Visibility Via Folk Theorization. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 380 (nov 2022), 31 pages. <https://doi.org/10.1145/3555105>
- [15] Michael A. DeVito, Darren Gergle, and Jeremy Birnholtz. 2017. "Algorithms Ruin Everything": #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 3163–3174. <https://doi.org/10.1145/3025453.3025659>
- [16] Michael A. DeVito, Ashley Marie Walker, Jeremy Birnholtz, Kathryn Ringland, Kathryn Macapagal, Ashley Kraus, Sean Munson, Calvin Liang, and Herman Saksono. 2019. Social Technologies for Digital Wellbeing Among Marginalized Communities. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) (CSCW '19). Association for Computing Machinery, New York, NY, USA, 449–454. <https://doi.org/10.1145/3311957.3359442>
- [17] Michael Ann DeVito, Ashley Marie Walker, and Julia R. Fernandez. 2021. Values (Mis)Alignment: Exploring Tensions Between Platform and LGBTQ+ Community Design Values. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 88 (apr 2021), 27 pages. <https://doi.org/10.1145/3449162>
- [18] Michael A. DeVito, Ashley Marie Walker, Caitlin Lustig, Amy J. Ko, Katta Spiel, Alex A. Ahmed, Kimberley Allison, Morgan Scheuerman, Briana Dym, Jed R. Brubaker, Ellen Simpson, Naveen Bagalkot, Noopur Raval, Michael Muller, Jennifer Rode, and Mary L. Gray. 2020. Queer in HCI: Supporting LGBTQIA+ Researchers and Research Across Domains. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/3334480.3381058>
- [19] Roel I. J. Dobbe. 2022. System Safety and Artificial Intelligence. *arXiv:2202.09292 [eess.SY]*
- [20] Diana Freed, Natalie N. Bazarova, Sunny Consolvo, Eunice J Han, Patrick Gage Kelley, Kurt Thomas, and Dan Cosley. 2023. Understanding Digital-Safety Experiences of Youth in the U.S.. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 191, 15 pages. <https://doi.org/10.1145/3544548.3581128>
- [21] Shubhangi Gupta, Sylvia Janicki, Pooja Casula, and Nassim Parvin. 2023. Rethinking Safe Mobility: The Case of Safetipin in India. In *Proceedings of the 2022 International Conference on Information and Communication Technologies and Development* (Seattle, WA, USA) (ICTD '22). Association for Computing Machinery, New York, NY, USA, Article 20, 6 pages. <https://doi.org/10.1145/3572334.3572392>
- [22] Oliver L. Haimson, Daniel Delmonaco, Peipei Nie, and Andrea Wegner. 2021. Disproportionate Removals and Differing Content Moderation Experiences for Conservative, Transgender, and Black Social Media Users: Marginalization and Moderation Gray Areas. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 466 (oct 2021), 35 pages. <https://doi.org/10.1145/3479610>
- [23] Oliver L. Haimson, Dyke Gorrell, Denny L. Starks, and Zu Weinger. 2020. Designing Trans Technology: Defining Challenges and Envisioning Community-Centered Solutions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3313831.3376669>
- [24] Susan Hao, Piyush Kumar, Sarah Laszlo, Shivani Poddar, Bhaktipriya Radharapu, and Renee Shelby. 2023. Safety and Fairness for Content Moderation in Generative Models. *arXiv:2306.06135 [cs.LG]*
- [25] Heidi Hartikainen, Afsaneh Razi, and Pamela Wisniewski. 2021. Safe Sexting: The Advice and Support Adolescents Receive from Peers Regarding Online Sexual Risks. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 42 (apr 2021), 31 pages. <https://doi.org/10.1145/3449116>
- [26] The White House. 2022. Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People. <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>
- [27] Naveena Karusala and Neha Kumar. 2017. Women's Safety in Public Spaces: Examining the Efficacy of Panic Buttons in New Delhi. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 3340–3351. <https://doi.org/10.1145/3025453.3025532>
- [28] Heidy Khlaaf. 2023. Toward Comprehensive Risk Assessments and Assurance of AI-Based Systems.
- [29] Nancy G Leveson. 2016. *Engineering a Safer World: Systems Thinking Applied to Safety*. The MIT Press.
- [30] Wayne G. Lutters and Mark S. Ackerman. 2002. Achieving safety: a field study of boundary objects in aircraft technical support. In *Proceedings of the 2002 ACM Conference on Computer Supported Cooperative Work* (New Orleans, Louisiana, USA) (CSCW '02). Association for Computing Machinery, New York, NY, USA, 266–275. <https://doi.org/10.1145/587078.587116>
- [31] Michael A. Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach. 2020. Co-Designing Checklists to Understand Organizational Challenges and Opportunities around Fairness in AI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376445>
- [32] Nora McDonald, Karla Badillo-Urquiola, Morgan G. Ames, Nicola Dell, Elizabeth Keneski, Manya Sleeper, and Pamela J. Wisniewski. 2020. Privacy and Power: Acknowledging the Importance of Privacy Research and Design for Vulnerable Populations. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI EA '20). Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3334480.3375174>
- [33] Tyler Musgrave, Alia Cummings, and Sarita Schoenebeck. 2022. Experiences of Harm, Healing, and Joy among Black Women and Femmes on Social Media. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 240, 17 pages. <https://doi.org/10.1145/3491102.3517608>
- [34] Karla Badillo-Urquiola John S. Seberger Denise E. Agosto Nora McDonald, Afsaneh Razi and Pamela Wisniewski. 2023. AI through the Eyes of Gen Z: Setting a Research Agenda for Emerging Technologies that Empower Our Future Generations. <https://aiyouthcscw2023.wordpress.com/organizers/>
- [35] Kathryn E. Ringland, Christine T. Wolf, Lynn Dombrowski, and Gillian R. Hayes. 2015. Making "Safe": Community-Centered Practices in a Virtual World Dedicated to Children with Autism. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) (CSCW '15). Association for Computing Machinery, New York, NY, USA, 1788–1800. <https://doi.org/10.1145/2675133.2675216>
- [36] Shalaleh Rismani, Renee Shelby, Andrew Smart, Renelito Delos Santos, Ajung Moon, and Negar Rostamzadeh. 2023. Beyond the ML Model: Applying Safety Engineering Frameworks to Text-to-Image Development. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) (AI/ES '23). Association for Computing Machinery, New York, NY, USA, 70–83. <https://doi.org/10.1145/3600211.3604685>
- [37] Shalaleh Rismani, Renee Shelby, Andrew Smart, Edgar Jatho, Joshua Kroll, Ajung Moon, and Negar Rostamzadeh. 2023. From Plane Crashes to Algorithmic Harm: Applicability of Safety Engineering Frameworks for Responsible ML. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 2, 18 pages. <https://doi.org/10.1145/3544548.3581407>
- [38] Morgan Klaus Scheuerman, Stacy M. Branham, and Foad Hamidi. 2018. Safe Spaces and Safe Places: Unpacking Technology-Mediated Experiences of Safety and Harm with Transgender People. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 155 (nov 2018), 27 pages. <https://doi.org/10.1145/3274424>
- [39] Renee Shelby, Shalaleh Rismani, Kathryn Henne, Ajung Moon, Negar Rostamzadeh, Paul Nicholas, N'Mah Yilla-Akbari, Jess Gallegos, Andrew Smart, Emilio Garcia, and Gurleen Virk. 2023. Sociotechnical Harms of Algorithmic Systems: Scoping a Taxonomy for Harm Reduction. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (Montréal, QC, Canada) (AI/ES '23). Association for Computing Machinery, New York, NY, USA, 723–741. <https://doi.org/10.1145/3600211.3604673>
- [40] Emily Tseng, Mehrnaz Sabet, Rosanna Bellini, Harkiran Kaur Sodhi, Thomas Ristenpart, and Nicola Dell. 2022. Care Infrastructures for Digital Security in Intimate Partner Violence. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 123, 20 pages. <https://doi.org/10.1145/3491102.3502038>
- [41] Lauren Wilcox, Renee Shelby, Rajesh Veeraraghavan, Oliver L. Haimson, Gabriela Cruz Erickson, Michael Turken, and Rebecca Gulotta. 2023. Infrastructuring Care: How Trans and Non-Binary People Meet Health and Well-Being Needs through Technology. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 489, 17 pages. <https://doi.org/10.1145/3544548.3581040>
- [42] Darcia Wilkinson and Bart Knijnenburg. 2022. Many Islands, Many Problems: An Empirical Examination of Online Safety Behaviors in the Caribbean. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 102, 25 pages. <https://doi.org/10.1145/3491102.3517643>