Developing Sociotechnical Solutions to Mitigate New Harms in Immersive and Embodied Virtual Spaces: A Workshop at CHI 2025

Guo Freeman
School of Computing
Clemson University
Clemson, South Carolina, USA
guof@clemson.edu

Julian Frommel
Utrecht University
Utrecht, Netherlands
j.frommel@uu.nl

Regan L. Mandryk
Computer Science
University of Victoria
Victoria, British Columbia, Canada
reganmandryk@uvic.ca

Jan Gugenheimer
TU-Darmstadt
Darmstadt, Germany
Telecom Paris
Paris, France
jan.gugenheimer@tu-darmstadt.de

Lingyuan Li
School of Information
The University of Texas at Austin
Austin, Texas, USA
lingyuan.li@ischool.utexas.edu

Douglas Zytko
University of Michigan-Flint
Flint, Michigan, USA
dzytko@umich.edu

Afsaneh Razi
Department of Information Science,
College of Computing & Informatics
Drexel University
Philadelphia, Pennsylvania, USA
afsaneh.razi@drexel.edu

Cliff Lampe
School of Information
University of Michigan
Ann Arbor, Michigan, USA
cacl@umich.edu

Abstract

Mitigating new harms in immersive and embodied virtual spaces (e.g., embodied harassment in social VR, new AI-powered online attacks, and harmful virtual world design to manipulate users) is a critically needed HCI research agenda for achieving safer online environments in the future, which requires cross-disciplinary, community-wide discussion, and collective reflections. Building upon our CHI 2024 workshop on identifying and understanding these new harms, this workshop aims to gather researchers and practitioners from various domains to collectively design and develop concrete and actionable sociotechnical solutions that specifically target new harms in immersive and embodied virtual worlds. This includes but is not limited to the four themes identified in our CHI 2024 workshop: monetizing embodied harms, blurring reality with the online world, platforming perpetrators by investigating their motivations and emotions, and embodied harms specifically targeting children. Through this workshop, we will not only synthesize and map our existing interdisciplinary efforts and challenges in this space but also collaboratively create a roadmap detailing our developed sociotechnical solutions to address these new harms in immersive and embodied virtual spaces as a community.

Permission to make digital or hard copies of all or part 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).
CHI EA '25, April 26–May 01, 2025, Yokohama, Japan
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1395-8/25/04
https://doi.org/10.1145/3706599.3706708

CCS Concepts

• **Human-centered computing** → **Collaborative and social computing**; **Human computer interaction (HCI)**.

Keywords

toxicity, online harm, online safety, harassment, embodiment, harm mitigation, artificial intelligence, immersive virtual worlds

ACM Reference Format:
Guo Freeman, Julian Frommel, Regan L. Mandryk, Jan Gugenheimer, Lingyuan Li, Douglas Zytko, Afsaneh Razi, and Cliff Lampe. 2025. Developing Sociotechnical Solutions to Mitigate New Harms in Immersive and Embodied Virtual Spaces: A Workshop at CHI 2025. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3706599.3706708

1 Background and Motivation

How diverse online users encounter, experience, and manage various forms of toxic, problematic, and harmful interactions (e.g., harassment, verbal abuse, hate speech, and flaming on social media platforms and online forums) remains a severe and pervasive issue in today's online social spaces, which seriously damages victims' mental, emotional, and physical well-being [5, 17, 33]. Moving towards gaming and virtual world contexts paints an even more direct picture. Indeed, as early as 1993, Dibbell's "A Rape In Cyberspace" detailed sexual assaults of women users in one of the first virtual worlds called LambdaMOO [8]. Now toxicity is largely accepted and normalized in online gaming [4, 14, 34] with statistics suggesting that 83% of adult gamers experience harassment in online multiplayer games with severe effects like causing distress [3]. With the rise of new immersive technologies such as Augmented and Virtual

Reality that heavily rely on immersion, presence, and embodiment, these toxic behaviors have the potential to be amplified or even take completely new shapes. Typical examples of these new and understudied forms of harm in immersive and embodied virtual spaces range from *embodied harassment* in social Virtual Reality (VR) that is especially harmful to marginalized online users, including women, LGBTQ+ individuals, and disabled persons [13, 27, 28] to new forms of online racism such as *racist Zoombombing* [23] and *Ugandan Knuckles* in VRChat [1], new AI-powered online attacks such as *hate raids* on Twitch [7], user- or developer-generated harmful design to manipulate people's actions [20], and new types of perceptual manipulations leveraging the core abilities of extended reality (XR) technology to create illusions to the user [6, 18, 32].

To further identify and deepen our understanding of these new harms, we have successfully held our CHI 2024 workshop [11]. By leveraging a speculative dystopian design fiction approach, at that workshop, 26 experts on this topic (14 women and 12 men with diverse ethnicities), including computer scientists, HCI scholars, VR and games researchers, social scientists, and policy makers, engaged in a structured design fiction exercise [10] to collectively create six design fictions. These design fictions highlight four main themes regarding our shared concerns in this problem space, including: **monetizing embodied harms** (e.g., how harm in embodied online spaces can be monetized to escalate its damage on the victims), **blurring reality with the online world** (e.g., the continuous collision of our offline identity and digital representations), **platforming perpetrators by investigating their motivations and emotions** (e.g., understanding why toxicity is always part of the online experience), and **embodied harms specifically targeting children** (e.g., how would placing children in such technical environments affect social dynamics and our future generation?). These insights have been published as a poster at CSCW 2024 [12]. It is also through our successful CHI 2024 workshop that we believe that developing novel sociotechnical solutions to mitigate these new harms is a critically needed HCI research agenda for achieving safer online environments in the future, which requires more cross-disciplinary, community-wide discussion, and collective reflections. In particular, such community-wide efforts should go beyond just *identifying and understanding* these new harms but to collectively design and develop more concrete and actionable *sociotechnical solutions* to mitigate such harms in immersive and embodied virtual spaces. This thus becomes the main goal of this follow-up workshop at CHI 2025.

2 Goals of the Workshop

In response to these new and understudied harms in immersive and embodied online spaces, a growing body of HCI research has begun to not only unpack the unique nature of such harms but also envision new solutions and technical interventions to mitigate them in various ways. For example, focusing on mitigating embodied harassment in social VR, various platform-embedded new safety features such as voice modulators for users to conceal their gender, sexual, and racial identity [13], better reporting and documentation mechanisms [13], consent mechanics to address interpersonal harm [30, 35], and AI-powered moderation systems [9, 29] have been proposed. Research has investigated how people cope with exposure to harms, such as toxicity [15, 21], and how tools integrated into games could provide support to those targeted by toxic behavior [26]. Additionally, researchers have proposed a shift from content moderation to design moderation to help better regulate user-generated virtual worlds/experiences, especially for youth [20]. Likewise, there have been various research efforts to utilize AI to detect and mitigate online risks for adolescents [24, 25]. For instance, researchers have utilized multimodal approaches to detect unsafe interactions [2] and found that meta-data from youth's private conversations provides sufficient evidence for machine learning models to detect unsafe interactions. This means that with minimum information, privacy-invasive tools such as parental monitoring apps [16] or prioritizing security over children's safety by implementing end-to-end encryption[1] could be addressed. Moreover, Slater et al. [31] opened the discussion among VR researchers about how increasing realism in VR might lead to even more threats to the user and make them susceptible to perceptual manipulations (e.g., memory source confusion in VR dominantly relies on the ability of the user to leverage visual cues that are associated uniquely to one source [19]). Right now, there are virtually no safeguards against such manipulations, which highlights a significant gap for ensuring safety in XR platforms.

This small body of existing work has demonstrated our community's collective efforts to develop solutions to mitigate new harms in various immersive and embodied online spaces. However, two concerns emerge. First, while these existing efforts have proposed various potential solutions, it is still unclear whether they aim to address all harms for all populations or if we still need more **targeted** efforts to develop nuanced and specific solutions for various types of harms targeting different user groups. Second, these existing efforts mainly focus on developing technical interventions. Given how most of these new harms are social in nature (e.g., interpersonal harms), there seems to be an inherent dilemma regarding using technical entities (e.g., computer programs) to address highly complex and context-sensitive social problems. As such, we believe that effective solutions for mitigating these new harms in immersive and embodied virtual spaces should not merely be technical but **sociotechnical.**

Therefore, the main goal of this workshop, as a follow-up to our successful CHI 2024 workshop [11], is to gather researchers and practitioners from various domains, including online safety, policy-making, social sciences, AI development, cybersecurity, games and virtual worlds, and XR, to **collectively design and develop concrete and actionable sociotechnical solutions that specifically target new harms in immersive and embodied virtual worlds.** This includes but is not limited to the four themes identified in our CHI 2024 workshop: *monetizing embodied harms, blurring reality with the online world, platforming perpetrators by investigating their motivations and emotions, and embodied harms specifically targeting children* [12]. Focusing on these themes (and other additional themes workshop attendees are interested in exploring), this workshop aims to:

- Synthesize existing strategies and recommendations to deal with each theme of new harms

[1] https://www.itpro.com/security/encryption/359943/what-is-end-to-end-encryption-and-why-is-everyone-fighting-over-it

- Articulate coping approaches for managing exposure to each theme of new harms that can be integrated into system designs
- Evaluate how, if at all, existing harm mitigation strategies (e.g., AI or moderation) can or cannot be used to effectively mitigate each theme of new harms
- Brainstorm, design, and prototype both social/legal (e.g., policy-making and community norms) and technical solutions to address each theme of new harms
- Discuss potential risks and ethical implications of the developed sociotechnical solutions to mitigate each theme of new harms
- Summarize concrete and generalizable takeaways and actionable approaches to tackle new harms in immersive and embodied virtual spaces that may go beyond these specific themes

3 Organizers

Our team of organizers represents experts in the area of games, live streaming, XR, online communities, youth online safety, and social VR with diverse research approaches (e.g., qualitative, experimental, design, and machine learning). Each brings unique expertise and vision to the topic.

Guo Freeman (main contact) is an Associate Professor of Human-Centered Computing at Clemson University. Her work focuses on how interactive technologies such as multiplayer online games, esports, live streaming, and social VR shape interpersonal relationships and group behavior; and how to design safe, inclusive, and supportive social VR spaces to combat emergent harassment risks especially for marginalized users.

Julian Frommel is an Assistant Professor in Interaction/Multimedia at Utrecht University. He is interested in the design and implementation of interactive digital systems that provide enjoyable, meaningful, safe, and healthy experiences for users, including research on how to mitigate negative effects of toxicity and harassment in online games and other online spaces.

Regan Mandryk is a Professor of Computer Science at the University of Victoria, Canada. Her work focuses on how people of all ages use playful technologies for social, cognitive, and emotional wellbeing, how toxicity, discrimination, and harassment thwart the connection and recovery benefits provided by multiplayer games, and how we can design playful collaborative systems that benefit—not harm—player wellbeing.

Jan Gugenheimer is an Assistant Professor at TU-Darmstadt and Telecom-Paris, working on HCI-related topics in the field of Extended Reality. His focus is on understanding how new unique properties of immersive technologies can be used to deceive and manipulate a user's actions and beliefs and how we have to design the technology to prevent such potential misuse.

Lingyuan Li is a Postdoctoral Fellow in the School of Information at the University of Texas at Austin. She delves into the intricacies of mediated experiences shaped by emerging technologies such as social VR, digital peer-to-peer payments, human-centered AI, and live streaming, and how we can leverage AI to mitigate emergent personal harm and risks and design more inclusive, safer, and supportive online spaces.

Douglas Zytko is an Associate Professor in the College of Innovation & Technology at the University of Michigan-Flint. His research uses consent as a lens to understand and design mitigative solutions for sexual harm. Most relevant to this workshop, his work explores data donation of online dating sexual experiences as a human-centered approach to improving sexual risk detection AI.

Afsaneh Razi is an Assistant Professor at Drexel University's Information Science Department. Her research area is positioned at the intersection of HCI and AI to address socio-technical issues. Specifically, her work aims to address the critical and timely problem of online safety by leveraging a multi-disciplinary approach of human-centered AI to accurately detect risks vulnerable users encounter online and develop online safety interventions.

Cliff Lampe is a Professor and Associate Dean for Academic Affairs at the University of Michigan School of Information. He also serves as Chair of the CHI Steering Committee and is a member of the CHI Academy. His work has highlighted moderation in online spaces, the effects of harassment and disinformation in online spaces, and recently the motivations behind harassers in these spaces.

4 Website and Pre-Workshop Plans

Similar to our CHI 2024 workshop[2], we have created a website (https://sites.google.com/view/chi-2025-workshop-solutions) to advertise the CFP, provide details about the workshop and submissions, and introduce the organizers. Leading up to the workshop, we will add the list of accepted workshop attendees, a schedule of workshop activities, and information related to accessibility accommodation. Following the workshop, we anticipate continuing to use the website to promote the outcomes of the workshop.

Prior to the workshop, we will make this workshop broadly visible in our community and further increase the awareness of the critical need for developing concrete and actionable sociotechnical solutions to mitigate new online harms in immersive and embodied virtual spaces. We will invite participants of our CHI 2024 workshop to attend this follow-up workshop to continue our collective efforts last year. We will also widely advertise this workshop to attract new participants with fresh perspectives. We will especially focus on including participants who represent diverse backgrounds and reach out to our connections in communities that focus on working with specific populations of underrepresented individuals. We will additionally reach out via mailing lists and other communities on Slack and Discord, not only associated with communities and universities in North America and Europe, but also via global networks. Moreover, we will point potential attendees to SIGCHI resources on financial support for attendee participation (e.g., Gary Marsden Travel Awards) in the CHI conference.

5 Hybrid Plans and Asynchronous Engagement

We are planning a hybrid workshop. Because of the interactive nature of the planned activities, we feel that attendees would benefit from in-person attendance over online attendance. However, due to the various considerations of reduced travel (e.g., sustainability, financial or visa restrictions, and ongoing travel restrictions due

[2]https://sites.google.com/view/chi-2024-workshop-newharms/home

to the pandemic and global mobility issues), we will support participation from those unable or unwilling to travel upon request. Plenary sessions and breakout sessions will be supported using videoconferencing tools (e.g., Zoom). Depending on the ratios of in-person and remote attendees, we will form remote breakout groups or integrate participants into hybrid groups during the interactive workshop activities, a format that we successfully implemented in our prior workshop at CHI 2023 [22] and CHI 2024 [11]. Our effective hybrid solution is no-cost to the conference, assuming a standard AV setup with a room projector and speaker. We will ask about accessibility requirements long before the conference and arrange auto-captioning (which benefits everyone).

6 Workshop Activities

This full day workshop will contain four main sessions of activities, demarcated by coffee breaks and lunch.

Session One: Building Community and Explaining Themes of New Harms We Target. At the beginning of the workshop, the organizers will communicate the goals for the day. Participants and organizers will also introduce themselves to situate the knowledge that people are offering and expect to gain throughout this workshop. In doing so, we aim to foster a sense of community, allowing participants and organizers to get to know each other, understand why we are interested in this problem space, and comprehend our shared group knowledge, concerns, and approaches.

Next, we will explain the four themes of new harms in immersive and embodied virtual worlds identified in our CHI 2024 workshop [11, 12] to both new and returning workshop attendees. We will also encourage all attendees to propose any other type of new harms than these four themes that they would like to focus on (e.g., harms concerning accessibility issues in these spaces). *The ultimate goal of this session is to (1) build our mutual understandings and targeted efforts to focus on a list of specific themes of new harms in immersive and embodied virtual spaces, rather than all harms; and (2) establish small groups who will work together to collectively tackle one theme of new harms respectively, based on the team's shared interests.*

Section Two: Collective Mapping: Existing Solutions and Limitations. After participants and organizers start to have a sense of community and form small groups based on which theme of new harms they would like to work on, participants will then work in breakout groups to collectively map existing strategies and recommendations to deal with each theme of new harms (if any) and limitations of these existing solutions. Participants will be especially encouraged to reflect upon why these existing solutions do or do not work from a sociotechnical perspective.

At the end of this session, we expect to collectively generate a visualization (e.g., a map or a chart) of existing sociotechnical solutions to mitigate each theme of new harms as well as specific limitations, concerns, and trade-offs of these solutions. *The ultimate goal of this session is to synthesize our existing interdisciplinary efforts and map out the challenges to develop concrete solutions to address specific forms of new harms in immersive and embodied virtual spaces, which will orient Session 3 workshop activities after the lunch break.*

Session Three: Design and Prototype New Sociotechnical Solutions. After lunch, we will then work in our breakout groups to design and prototype new concrete and actionable approaches, directions, and principles to better mitigate each theme of new harms in immersive and embodied virtual spaces. These proposed new solutions can either focus on addressing existing solutions' limitations or provide alternative approaches to mitigate these new harms. Participants will also be encouraged to not only emphasize technical interventions (e.g., how to design and implement a consent feature in social VR) but consider necessary and structural changes at the social and legal level (e.g., how to appropriately revise community norms). Each group will report back to the plenary session on their discussion. *The ultimate goal of this session is to collaboratively develop and elaborate concrete and generalizable takeaways and actionable approaches to tackle new harms in immersive and embodied virtual spaces, which may go beyond the specific themes we focus in this workshop.*

Session Four: Creating a Roadmap. The last session of this workshop will focus on collaboratively creating a roadmap detailing our developed sociotechnical solutions to mitigate new harms in immersive and embodied virtual spaces as a group. We envision that this roadmap will include at least three sections based on our workshop activities throughout all 4 sessions: (1) detailing and visualizing existing solutions to address specific themes of new harms in immersive and embodied virtual spaces and limitations of each solution; (2) the list of guidelines, recommendations, sketches, and prototypes that each breakout group has developed to either address these existing solutions' limitations or provide alternative approaches to mitigate these new harms; (3) our collective stance, recommendations, and envisions of this problem space.

We will then wrap up by summarizing the day and discussing post-workshop plans for continued community building, progress, and opportunities to work together. Similar to how we published the outcomes of our CHI 2024 workshop as a CSCW 2024 poster [12], *the ultimate goal of this session is to keep refining this roadmap after the workshop and submit it for publication (e.g., submitting to CSCW 2025 as a poster).* Our hope is that this potential publication can serve as a starting point for the broader HCI community to continue such conversations and exploration.

7 Accessibility

See Section 5 for a description of how we plan to accommodate accessibility needs in the workshop itself.

8 Post-Workshop Plans and Plans to Publish Workshop Outcomes

This year, to allow for focused work toward a tangible outcome, we want to support the participation of topic experts and participants who are interested in the topic. Therefore, for this workshop, we plan to invite interested researchers and practitioners to submit a brief bio and position statement outlining their interest in the topic (each up to 200 words). In the position statement, participants should describe why they are interested in the area, which work they have already done, and if they are interested in specific topics related to the workshop. All the submitted materials will be hosted on our workshop website, providing a repository for the discussion. Instead of publishing those position statements and workshop proceedings via arXiv, we will focus on submitting the outcomes of our collective efforts at this follow-up workshop for publication

(e.g., as a poster for CSCW 2025), similar to how we published the outcomes of our CHI 2024 workshop as a CSCW 2024 poster [12].

In addition to research outputs, we also aim for the workshop to motivate future events (e.g., panels/workshops at other SIGCHI venues), research collaborations, and grant applications.

9 Expected Size of Attendance

We aim to be inclusive in the selection and accept all submissions that are in scope of our workshop. Last year, more than 30 in-person attendees and online participants participated in our CHI 2024 workshop. We expect a similar size of attendance for this follow up CHI 2025 workshop.

10 Call for Participation

Immersive and embodied virtual spaces are increasingly prevalent. While generally beneficial, the paradigm shifts introduced by such environments (e.g., immersion, embodiment, or technological advancements) can also lead to new harms for users, such as embodied harassment in social VR, new AI-powered online attacks, user- or developer-generated harmful design to manipulate people's actions, and new types of perceptual manipulations in XR to create illusions to the user. In our CHI 2024 workshop, we brought together experts to identify and understand these new harms. Building on this, we want to bring together a set of interdisciplinary researchers and practitioners from HCI and adjacent fields to **collectively design and develop concrete and actionable sociotechnical solutions that specifically target new harms in immersive and embodied virtual worlds.** This includes (but is not limited to) the four themes identified in our CHI 2024 workshop: *monetizing embodied harms, blurring reality with the online world, platforming perpetrators by investigating their motivations and emotions, and embodied harms specifically targeting children.* We invite interested researchers and practitioners to submit a position statement and bio (each up to 200 words). In the position statement, participants should describe their experience and interest in the topic and workshop. Submissions will be accepted via the workshop website: https://sites.google.com/view/chi-2025-workshop-solutions. The workshop will be organized as a hybrid event at CHI 2025, but we encourage in-person attendance. One participant from each submission must register for the workshop and at least one day of the CHI conference.

Acknowledgments

This work was partially supported by the National Science Foundation under award 2112878.

References

[1] Julia Alexander. 2018. 'Ugandan Knuckles' is overtaking VRChat. https://www.polygon.com/2018/1/8/16863932/ugandan-knuckles-meme-vrchat

[2] Shiza Ali, Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Chen Ling, Munmun De Choudhury, Pamela J Wisniewski, and Gianluca Stringhini. 2023. Getting Meta: A Multimodal Approach for Detecting Unsafe Conversations within Instagram Direct Messages of Youth. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–30.

[3] Anti-Defamation League. 2021. Hate is no game: Harassment and positive social experiences in online games 2021. https://www.adl.org/hateisnogame#executive-summary

[4] Nicole A Beres, Julian Frommel, Elizabeth Reid, Regan L Mandryk, and Madison Klarkowski. 2021. Don't You Know That You're Toxic: Normalization of Toxicity in Online Gaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, Yokohama, Japan, 1–15. https://doi.org/10.1145/3411764.3445157

[5] Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2017. Classification and its consequences for online harassment: Design insights from heartmob. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–19.

[6] Elise Bonnail, Wen-Jie Tseng, Mark Mcgill, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2023. Memory Manipulations in Extended Reality. 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 875, 20 pages. https://doi.org/10.1145/3544548.3580988

[7] Jie Cai, Sagnik Chowdhury, Hongyang Zhou, and Donghee Yvette Wohn. 2023. Hate Raids on Twitch: Understanding Real-Time Human-Bot Coordinated Attacks in Live Streaming Communities. *arXiv preprint arXiv:2305.16248* (2023).

[8] Julian Dibbell. 1993. A Rape in Cyberspace: or, How an Evil Clown, a Haitian Trickster Spirit, Two Wizards, and a Cast of Dozens Turned a Database Into a Society. http://www.juliandibbell.com/texts/bungle_vv.html

[9] Cristina Fiani, Robin Bretin, Mark McGill, and Mohamed Khamis. 2023. Big Buddy: Exploring Child Reactions and Parental Perceptions towards a Simulated Embodied Moderating System for Social Virtual Reality. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*. 1–13.

[10] Casey Fiesler. 2018. Black Mirror, Light Mirror: Teaching Technology Ethics Through Speculation. https://cfiesler.medium.com/the-black-mirror-writers-room-teaching-technology-ethics-through-speculation-f1a9e2deccf4.

[11] Guo Freeman, Julian Frommel, Regan L Mandryk, Jan Gugenheimer, Lingyuan Li, and Daniel Johnson. 2024. Novel Approaches for Understanding and Mitigating Emerging New Harms in Immersive and Embodied Virtual Spaces: A Workshop at CHI 2024. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[12] Guo Freeman, Julian Frommel, Regan L Mandryk, Jan Gugenheimer, Lingyuan Li, Daniel Johnson, Cecilia Aragon, Syed A Asif, Jakki Bailey, Barkallah Meryem, et al. 2024. Understanding and Mitigating New Harms in Immersive and Embodied Virtual Spaces: A Speculative Dystopian Design Fiction Approach. *Companion of the 2024 ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW'24)* (2024).

[13] Guo Freeman, Samaneh Zamanifard, Divine Maloney, and Dane Acena. 2022. Disturbing the Peace: Experiencing and Mitigating Emerging Harassment in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–30.

[14] Julian Frommel, Daniel Johnson, and Regan L Mandryk. 2023. How perceived toxicity of gaming communities is associated with social capital, satisfaction of relatedness, and loneliness. *Computers in Human Behavior Reports* 10 (2023), 100302.

[15] Julian Frommel and Regan L Mandryk. 2024. Toxicity in Online Games: The Prevalence and Efficacy of Coping Strategies. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 959, 12 pages. https://doi.org/10.1145/3613904.3642523

[16] Arup Kumar Ghosh, Karla Badillo-Urquiola, Shion Guha, Joseph J LaViola Jr, and Pamela J Wisniewski. 2018. Safety vs. surveillance: what children have to say about mobile apps for parental control. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

[17] Nitesh Goyal, Leslie Park, and Lucy Vasserman. 2022. " You have to prove the threat is real": Understanding the needs of Female Journalists and Activists to Document and Report Online Harassment. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[18] Jan Gugenheimer, Wen-Jie Tseng, Abraham Hani Mhaidli, Jan Ole Rixen, Mark McGill, Michael Nebeling, Mohamed Khamis, Florian Schaub, and Sanchari Das. 2022. Novel Challenges of Safety, Security and Privacy in Extended Reality. 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 108, 5 pages. https://doi.org/10.1145/3491101.3503741

[19] Marcia K Johnson, Shahin Hashtroudi, and D Stephen Lindsay. 1993. Source monitoring. *Psychological bulletin* 114, 1 (1993), 3.

[20] Yubo Kou and Xinning Gui. 2023. Harmful Design in the Metaverse and How to Mitigate It: A Case Study of User-Generated Virtual Worlds on Roblox. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 175–188. https://doi.org/10.1145/3563657.3595960

[21] Samuli Laato, Bastian Kordyaka, Velvet Spors, and Juho Hamari. 2024. How StarCraft II Players Cope with Toxicity: Insights from Player Interviews. In *HCI in Games*, Xiaowen Fang (Ed.). Springer Nature Switzerland, Cham, 203–219. https://doi.org/10.1007/978-3-031-60695-3_14

[22] Regan L Mandryk, Julian Frommel, Nitesh Goyal, Guo Freeman, Cliff Lampe, Sarah Vieweg, and Donghee Yvette Wohn. 2023. Combating Toxicity, Harassment, and Abuse in Online Social Spaces: A Workshop at CHI 2023. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[23] Lisa Nakamura, Hanah Stiverson, and Kyle Lindsey. 2021. *Racist Zoombombing*. Routledge.

[24] Afsaneh Razi, Ashwaq AlSoubai, Seunghyun Kim, Shiza Ali, Gianluca Stringhini, Munmun De Choudhury, and Pamela J Wisniewski. 2023. Sliding into my DMs: Detecting uncomfortable or unsafe sexual risk experiences within Instagram direct messages grounded in the perspective of youth. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–29.

[25] Afsaneh Razi, Seunghyun Kim, Ashwaq Alsoubai, Gianluca Stringhini, Thamar Solorio, Munmun De Choudhury, and Pamela J. Wisniewski. 2021. A Human-Centered Systematic Literature Review of the Computational Approaches for Online Sexual Risk Detection. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 465 (oct 2021), 38 pages. https://doi.org/10.1145/3479609

[26] Elizabeth Reid, Regan L Mandryk, Nicole A Beres, Madison Klarkowski, and Julian Frommel. 2022. Feeling Good and In Control: In-game Tools to Support Targets of Toxicity. *Proceedings of the ACM on Human-Computer Interaction* 6, CHI PLAY (2022), 1–27.

[27] Kelsea Schulenberg, Guo Freeman, Lingyuan Li, and Catherine Barwulor. 2023. Creepy Towards My Avatar Body, Creepy Towards My Body: How Women Experience and Manage Harassment Risks in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* CSCW (2023). https://guof.people.clemson.edu/papers/cscw23women.pdf

[28] Kelsea Schulenberg, Guo Freeman, Lingyuan Li, and Bart Knijnenburg. 2024. Does Who You Are or Appear to Be Matter?: Understanding Identity-Based Harassment in Social VR Through the Lens of (Mis) Perceived Identity Revelation. *Proceedings of the ACM on humancomputer interaction* (2024).

[29] Kelsea Schulenberg, Lingyuan Li, Guo Freeman, Samaneh Zamanifard, and Nathan J. McNeese. 2023. Towards Leveraging AI-based Moderation to Address Emergent Harassment in Social Virtual Reality. In *Proceedings of the 2023 CHI Confrence on Human Factors in Computing Systems*.

[30] Kelsea Schulenberg, Lingyuan Li, Caitlin Lancaster, Doug Zytko, and Guo Freeman. 2023. "We Don't Want a Bird Cage, We Want Guardrails"': Understanding & Designing for Preventing Interpersonal Harm in Social VR through the Lens of Consent. *Proceedings of the ACM on Human-Computer Interaction* CSCW (2023). https://guof.people.clemson.edu/papers/cscw23consent.pdf

[31] Mel Slater, Cristina Gonzalez-Liencres, Patrick Haggard, Charlotte Vinkers, Rebecca Gregory-Clarke, Steve Jelley, Zillah Watson, Graham Breen, Raz Schwarz, William Steptoe, et al. 2020. The ethics of realism in virtual and augmented reality. *Frontiers in Virtual Reality* 1 (2020), 1.

[32] Wen-Jie Tseng, Elise Bonnail, Mark McGill, Mohamed Khamis, Eric Lecolinet, Samuel Huron, and Jan Gugenheimer. 2022. The Dark Side of Perceptual Manipulations in Virtual Reality. 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 612, 15 pages. https://doi.org/10.1145/3491102.3517728

[33] Jessica Vitak, Kalyani Chadha, Linda Steiner, and Zahra Ashktorab. 2017. Identifying women's experiences with and strategies for mitigating negative effects of online harassment. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 1231–1245.

[34] Michel Wijkstra, Katja Rogers, Regan L Mandryk, Remco C Veltkamp, and Julian Frommel. 2023. Help, My Game Is Toxic! First Insights from a Systematic Literature Review on Intervention Systems for Toxic Behaviors in Online Video Games. In *Companion Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 3–9.

[35] Douglas Zytko and Jonathan Chan. 2023. The Dating Metaverse: Why We Need to Design for Consent in Social VR. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2489–2498.