Facing the Illusion and Reality of Safety in Social VR

Qingxiao Zheng
School of Information Sciences
University of Illinois at Urbana-Champaign, USA
qzheng14@illinois.edu

Tue Ngoc Do
Department of Computer Science
University of Illinois at Urbana-Champaign, USA
tuedo2@illinois.edu

Lingqing Wang
Department of Industrial Engineering
Tsinghua University, China
wanglq19@mails.tsinghua.edu.cn

Yun Huang
School of Information Sciences
University of Illinois at Urbana-Champaign, USA
yunhuang@illinois.edu

Abstract

The ethical design of social Virtual Reality (VR) is not a new topic, but "safety" concerns of using social VR are escalated to a different level given the heat of the Metaverse. For example, it was reported that nearly half of the female-identifying VR participants have had at least one instance of virtual sexual harassment. Feeling safe is a basic human right — in any place, regardless in real or virtual spaces. In this paper, we are seeking to understand the discrepancy between user concerns and designs in protecting users' safety in social VR applications. In this paper, we study safety concerns on social VR experience by analyzing Twitter posts and synthesizing practices on safety protection adopted by mainstream social VR platforms. We argue that future research and platforms should explore the design of social VR with boundary-awareness.

CCS Concepts: • **Human-centered computing → Virtual reality**; **Interaction paradigms**; **Interaction paradigms**; **Virtual reality**; • **Software and its engineering → Software safety**.

Keywords: Metaverse, virtual reality, safety, social VR, boundary

1 Introduction

The 3D virtual world, or metaverse [7], is broadly defined to be the recent flush of technologies including and related to Virtual Reality (VR) and Augmented Reality (AR), consisting of permanent virtual worlds that persist when we log off. This new dimension of user interaction is likely to spark new social norms and cultural behaviors.

However, attention has been drawn towards the safety, or lack of, features currently existing in commercial VR spaces. For example, a recent news report [2] asserted that virtual worlds lack adequate safety precautions; reporting on a group of researchers who recorded a social transgression (e.g. racism or sexual harassment) on average once per seven minutes, and often with minors present. Heavy criticism has been raised on some VR front runners, for focusing on growth at the expense of safety and privacy [5, 6, 12].

Furthermore, considerations of the immersive facet of virtual reality have led some to explore new avenues of online attacks. In comparison to traditional social medias, in which a user has the physical barrier of a screen held at arms length, psychological attacks within these spaces are potentially more traumatic [2, 3, 16]. There seems to be less boundaries in VR that can rule and determine what are reasonable, psychologically safe and permissible ways for other people to behave around self and how self will respond when someone steps outside those limits. To this end, we are trying to explore the following questions:

RQ1: *What safety concerns are emerging in social VR platforms?*

RQ2: *What mechanisms do mainstream social VR products use to protect user safety?*

2 Related Social VR Research

Social VR has been defined as "a growing set of multi-user applications that enable people to interact with one another in virtual space through VR headmounted displays" [11]. Today's social VR applications are similar to "collaborative virtual environments" studied in the late 1990s and early 2000s, which applied virtual reality to support multiple user interactions [4].

CHI EA '22, Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality, April 29-May 5, 2022, New Orleans, LA, USA

Zheng, et al.

There are a set of works mapping social VR landscape and design research by proposing a taxonomy of social VR application design ecology [15, 21]. Some work specially investigated codes of ethics in the virtual worlds [1, 8, 13]. Some studied social VR from human factor perspective. For example, one study investigated how violations of interpersonal space can cause discomfort in real-world situations [19], and another work learned how users shape pro-social behavior in VR [14].

New questions have arisen from the embodied and spatialized experiences that immersive presence in VR affords in recent years. Harassment in social VR, for instance, has proven particularly fraught, with the effects of abuse magnified by embodied presence [15]. However, to the best of our knowledge, research related to safety mechanisms applied in social VR is still very limited.

3 Method

We conducted an exploratory study. The seed findings will be expanded with a more comprehensive investigation in the future.

3.1 Scraping Social Media Posts

To answer RQ1, we used the Tweepy and PRAW Python packages to scrape recent three months' content from Twitter and Reddit, respectively. These social media platforms were chosen due to their more public natures and author familiarity. We used search queries of social VR synonyms (e.g., social VR, social virtual reality, virtual reality, virtual worlds) and names of commercial social VR applications [21]. We ended up with 18,274 data entries and proceeded by manually reviewing all tweets and comments after filtering by keywords. Using a thematic analysis approach [18], we synthesized major concerns and discussion topics of social media posts related to social VR.

3.2 Investigating Mainstream Social VR Platforms

To answer RQ2, we investigated 14 existing social VR commercial applications previously summarized [10, 21]. Based on number of downloads in across VR app stores of Oculus Store[1], Steam VR[2], and Sidequest [3], we chose four social VR apps. They are Rec Room[4], which allows users to manipulate a variety of objects and engage with mini-games with other users; VRChat[5], which is known for non-normative social interactions and performative memes; Horizon Worlds[6], which allows users to explore virtual worlds and engage in content consumption and creation; and AltspaceVR[7], which

---

[1]https://www.oculus.com/experiences/quest

[2]https://store.steampowered.com/vr

[3]https://sidequestvr.com/all-apps

[4]https://recroom.com

[5]https://hello.vrchat.com/

[6]https://www.oculus.com/horizon-worlds

[7]https://altvr.com

is used for live, virtual events, empowering artists, brands, and businesses. One researcher experienced these social VR apps using an Oculus Quest 2 headset and triggered the safety mechanisms built within to examine these apps' safety protection protocols. Then, one co-author summarized the observations of the safety mechanisms applied by the apps. Additionally, co-authors analyzed the relationship between safety-related design features and the kinds of social interactions they may support or constrain, by reviewing their experiences of onboarding, exploration, and play.

4 Preliminary Takeaways

4.1 Safety Concerns about Social VR (RQ1)

4.1.1 Safety of Minors. A key concern mentioned by multiple users across platforms were the existence of **minors** in these VR environments. Although users seemed to be split on whether to blame parents or developers for negligence, users unanimously expressed that minors did not belong on the platform, as their safety could not be guaranteed. The underlying reasoning was the perceived accessibility of dangerous/unfitting material. For example, one Twitter user made an analogy to the physical world "*There's a benefit to the physical distance between the average house and the Vegas strip. Harder for your kids to just wander over there.*"

Virtual reality has attracted many young users with chat rooms and popular games, such as Roblox. But whether it can adequately protect children is a major concern. Many users of Horizon Worlds have reported that Facebook, now Meta, does not sufficiently enforce the age restriction, which leads to worries for minors' safety and a negative adult user experience. These problems have been widely discussed and addressed on social media. However, some platforms in virtual reality spaces still lag behind common concerns, lacking preexisting methods such as parental controls and guardrails for young users, to the frustrations of some community members.

4.1.2 Distrust of Corporations. Another concern is distrust. Specifically, regarding the track record of Meta, there seemed to be some pushback with Reddit users feeling uncomfortable plugging in at all. One Reddit user remarked "*I can assure you I will not be using it* [Oculus]...*Zuckerberg has revealed over and over he doesn't [care (replaced word for polite language)] about safety or privacy.*"

There also seems to be a disdain for Meta's current moderation systems. Regarding **sexual harassment** and **abusive behavior**, one commentor expresses in regards to a blocking feature (at the time unimplemented), "*I would not be surprised if Facebook empowers users even in this minor way. You know they aren't gonna bother banning creeps.*"

In December 2021, a beta tester of Horizon Worlds wrote about being virtually groped by another user, creating quite a stir. After that, while VR platforms are claiming to develop different tools and improve their accessibility to protect users,

CHI EA '22, Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality, April 29-May 5, 2022, New Orleans,
Facing the Illusion and Reality of Safety in Social VR
LA, USA

more harassment cases are reported. Sexual harassment exists in all kinds of digital spaces. However, the immersive, multi-sensory and 'realistic' nature of VR triggers stronger emotional responses, thus even worse repercussions for users in such circumstances.

4.1.3 Humiliating Behaviors. Currently, the public seems to largely rehash many known concerns with other social media without considering the additional dimension and complications a virtual reality can bring. For example, regarding a now-famous incident of a woman being groped while in VR, some commentors were dismissive of her experiences, suggesting that woman just "turn it off", and falsely equalizing the experience with milder internet norms such as teabagging (the act of repeatedly crouching). As more incidents were logged with these technologies and social space, and new norms and cultures are yet established, it remains to be seen where public opinions start to lean in terms of user safety in the virtual worlds.

4.1.4 Integration with the *Reality*. Many users have been discussing that Social VR platforms might blur the boundary between reality and the virtual world. Despite users' worries that the VR world may lead to addiction easily, causing more serious **isolation** from real-world and cognitive problems. It also raises concerns when integrating into the real world. For example, there are several news reports on the physical injury in the real-world caused by body movement in the virtual world, when users immerse themselves in the playing.

Some other emerging issues have also caught public attention. One is the **sleeping phenomenon**. Widely discussed on YouTube channels, there is a unique phenomenon that users regularly sleep on virtual platforms despite all the inconveniences in real world. For example, the headset might hinder head movement while sleeping, and the cable might tie around users' neck. Reasons for sleeping in a social VR platform vary from social satisfaction when sleeping with friends to avoiding the real-world noisy environment. Though users make the trade-off to do so, they are generally not clear about the potential consequence.

Another emerging issue is phantom touch, or more broadly speaking, **phantom sense**, which means users feel the sensation as if they are their avatars, without any haptic technology. One Reddit user wrote "*when someone headpats me (well, my avatar, but to me, they're the same thing...) it gives me a tingly feeling on my head ...*" There are many explanations and guides to help users develop phantom touch in Reddit and other discussion forums. Some users claim they naturally gain phantom sense, and other provides various methods to trick the mind into that. For example, one user in the Steam community recommended asking a friend to pet users' face in VR, suggesting that "*area most sensitive for most people is the small part between the upper lips and the bottom of the nose - especially to the left and right of the center.*" While this phenomenon is not clear, it adds exciting layers to user experience in virtual world. Many users are trying to practice and keep the feelings of phantom touches in VR. However, many users also express their worries about whether they can still differentiate between reality and the virtual environment. After all, having a bad experience for the avatars, like being shoved or stabbed, might not be fun.

In summary, we found that safety concerns on social media include "intensified old concerns in the new world." Harassment and abusive behaviors, distrust of user privacy, content appropriateness for children etc., have long been discussed in previous social platforms. And now, they exacerbate users' concerns because of the double-edged, more authentic experience that social VR platforms promise to provide, which would also bring more real lousy experience at the same time.

Also, some "new concerns in the new world" have recently become salient, such as problems caused by the separation of body and mind in real and virtual worlds, and emerging issues like sleeping phenomenon and phantom sense. How work and lifestyles would be reshaped, and what that might impact user psychological and physical safety are not yet addressed extensively.

4.2 Designs Features for User Safety (RQ2)

In a prior work, we advocate raising "boundary-awareness" when designing AI-based interactive systems to support human-human interaction [20]. When reviewing safety mechanisms adopted by main stream social VR platforms, we identified six major social boundary settings, four safety protection mechanisms, and community protocols.

4.2.1 Setting Social Boundaries. We found six boundary related design practices to keep users comfortable with social distance when hanging out in social VR.

(1) Intimacy Proxemic. Hall's personal space theory describes that there are four zones in personal space [9]. Recently, Horizon Worlds released the personal boundary feature with a default setting making users feel like an almost 4-foot distance between their avatar and others. Similarly, Rec Room has an adjustable "personal space bubble," an invisible space that allows users to control how close other players interact with their avatars from close, medium, to large. We also found a similar function in AltspaceVR.

(2) Intimacy Rank. When playing VRchat, users' trust feeds into a "trust rank" with several levels, i.e., trusted user, known user, user, new user, visitor, and nuisance.

(3) Social Spaces. Rec Room has a dorm room, user-created rooms, and clubs. The dorm room is a completely private space. Room owners of user-created rooms and clubs are fully responsible for ensuring their rooms do not devolve into a toxic mess, even if they are going on vacations in the real world. We found that local expectations are set for

CHI EA '22, Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality, April 29-May 5, 2022, New Orleans, LA, USA

Zheng, et al.

conduct in public spaces. For example, it must be private if a club involves sexual themes.

(4) Image Control. Users can adjust both their own avatars' identities, such as gender, race, and voice, and select to hide or show specific users' avatars. For example, in VRchat, when a user encounters a user that, despite their higher Trust Rank, is wearing an avatar they feel uncomfortable, they can choose "hide avatar" in their social panel.

(5) Trust System. Users can adjust control settings to maintain desirable social boundaries with others. For example, VRChat trust and safety system can keep users safe from nuisance by adjusting microphones, screen-space shaders, loud sounds, or visually noisy or malicious particle effects.

(6) User Demographics. Age limits are set in social VR. For example, Rec Room enables a junior account managed by an account owned by the parent or guardian for players under 13 years of age, which come with safety features to prevent children from disclosing their identifiable information in the virtual world. It also allows age-based matchmaking to recommend players close to their age range.

4.2.2 Quick Reactions When Boundaries are Crossed.
Different safety self-protection mechanisms were designed to allow users to react accordingly when their social boundaries are violated in the virtual world, ranging from reactions and safe zones to safety reports.

Safety Reactions. Simple communication gestures and shortcuts allowed quick-action remediation in challenging situations. For example, Rec Room's 'talk to the hand' and pointing gestures can instantly trigger users' comfort and moderation menu, allowing users to mute, block, vote kick, and report the player. These simple mechanisms enable users to report a hurting experience instantly without interrupting or degrading their experience.

Safe Mode. VRchat has a shortcut called "safe mode," which can immediately disable all features on all users around. Horizon worlds offer a one-touch button that can quickly remove users from a situation. Similarly, AltspaceVR has a "title screen" feature, which immediately drops users back into the command center, removing their avatars from the previous space and scene.

Safety Reports. Users can report users, rooms or event places that violate codes of conduct or ask for extra support.

4.2.3 Consent and Codes of Conduct.
All applications have informed Consent before using entering the application. The informed consent contents regarding health and safety warnings range from the immersive effects on users' psychological status, a notice of risky content during social interactions, and user privacy. Some applications, for example, Rec Room, also require users to finish a tutorial of codes of conduct in the virtual environment; these codes were introduced to newcomers and repeated displayed to members of the communities.

5 Opportunities for Future Research
Initial results reveal many old safety concerns in the new virtual world (e.g., user trust, abusive language and behavior, threats to minors) which current design features have addressed. However, some new issues, such as the "phantom sense," are not considered by any reviewed social VR apps. There are many novel safety facets of this emerging design space of social VR. Below, we propose some topics to explore.

5.1 Re-defining User Safety in Social VR
Do we need to redefine safety in social VR? VR has many shared characteristics with our experiences in real life. Because players can disguise themselves into another human, animal, or object and use transportation behaviors, they may have different social boundaries and appropriateness definitions than as usual. Due to its nascent stage and creative formats, it is yet to investigate the social norms being applied in the new world. One approach is distinguishing universal rules and consequences for real-world and virtual-world violations and harassment.

5.2 Boundary-Awareness for Safer Social Systems
Boundary-aware design is an emerging and under-explored research area[20]. Our findings on social VR safety mechanisms demonstrate the dynamics of boundary consideration in social system design. Studying social systems, considering the importance of various boundaries and the interactions occurring at different levels of boundaries, may add value to the human-human interactions through these social technologies. For example, building granular controls on different boundary levels (e.g., on information sharing, emotional boundaries, differentiating reality and virtual, persona building, to name a few) may help foster safety in social VR spaces. Researchers can also develop a boundary-aware framework to support the design process and evaluations with such awareness.

5.3 Values of Platforms, Creators, and Users
Building safe and responsible VR ecology will have a competitive edge in the virtual world. For platforms, toolkits providers that have the autonomy to limit content creators' production capabilities may easily restrict creators' development capabilities by restraining risky developer support. Creators, such as those in the Decentraland, a first-ever platform to create, explore, and trade land in the virtual world owned by its users, have high autonomy in the "fully decentralized world." Platforms can hardly control the content created. For users, we found community consensuses and discussions regarding safety in social VR are not fully developed, perhaps due to its novelty. One possible approach perhaps can be community-based safety reports, where bystanders can intervene and stop a violent attack by one person on another [17].

CHI EA '22, Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality, April 29-May 5, 2022, New Orleans,
Facing the Illusion and Reality of Safety in Social VR
LA, USA

A shared understanding is needed and we posit that more future studies could uncover more fruitful understandings.

References

[1] Devon Adams, Alseny Bah, Catherine Barwulor, Nureli Musaby, Kadeem Pitkin, and Elissa M Redmiles. 2018. Ethics emerging: the story of privacy and security perceptions in virtual reality. In *Fourteenth Symposium on Usable Privacy and Security (SOUPS 2018)*. 427–442.

[2] Maura Barrett and Forte Douglas. 2019. *Metaverse virtual worlds lack adequate safety precautions.* https://www.nbcnews.com/tech/internet/metaverse-virtual-worlds-lack-adequate-safety-precautions-critics-say-rcna15418

[3] Tanya Basuarchive. 2021. *The metaverse has a groping problem already.* https://www.technologyreview.com/2021/12/16/1042516/the-metaverse-has-a-groping-problem

[4] Barbara Becker and Gloria Mark. 1998. Social conventions in collaborative virtual environments. In *Proceedings of Collaborative Virtual Environments*. 17–19.

[5] Lindsay Blackwell, Nicole Ellison, Natasha Elliott-Deflo, and Raz Schwartz. 2019. Harassment in social VR: Implications for design. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 854–855.

[6] Ellysse Dick. 2021. *Balancing User Privacy and Innovation in Augmented and Virtual Reality*. Technical Report. Information Technology and Innovation Foundation.

[7] John David N Dionisio, William G Burns III, and Richard Gilbert. 2013. 3D virtual worlds and the metaverse: Current status and future possibilities. *ACM Computing Surveys (CSUR)* 45, 3 (2013), 1–38.

[8] Colin M Gray and Shruthi Sai Chivukula. 2019. Ethical mediation in UX practice. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[9] Edward T Hall, Ray L Birdwhistell, Bernhard Bock, Paul Bohannan, A Richard Diebold Jr, Marshall Durbin, Munro S Edmonson, JL Fischer, Dell Hymes, Solon T Kimball, et al. 1968. Proxemics [and comments and replies]. *Current anthropology* 9, 2/3 (1968), 83–108.

[10] Marcel Jonas, Steven Said, Daniel Yu, Chris Aiello, Nicholas Furlo, and Douglas Zytko. 2019. Towards a taxonomy of social vr application design. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. 437–444.

[11] Anya Kolesnichenko, Joshua McVeigh-Schultz, and Katherine Isbister. 2019. Understanding emerging design practices for avatar systems in the commercial social vr ecology. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 241–252.

[12] Fade Lorne. 2021. *What Is Virtual Reality, And How Can It Be Used In The Workplace?* https://www.forbes.com/sites/forbesbusinesscouncil/2021/04/05/what-is-virtual-reality-and-how-can-it-be-used-in-the-workplace/?sh=1905b7b27a1e

[13] Michael Madary and Thomas K Metzinger. 2016. Real virtuality: a code of ethical conduct. Recommendations for good scientific practice and the consumers of VR-technology. *Frontiers in Robotics and AI* 3 (2016), 3.

[14] Joshua McVeigh-Schultz, Anya Kolesnichenko, and Katherine Isbister. 2019. Shaping pro-social interaction in VR: an emerging design framework. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[15] Joshua McVeigh-Schultz, Elena Márquez Segura, Nick Merrill, and Katherine Isbister. 2018. What's It Mean to" Be Social" in VR? Mapping the Social VR Design Ecology. In *Proceedings of the 2018 ACM Conference Companion Publication on Designing Interactive Systems*. 289–294.

[16] Frenkel Sheera and Browning Kellen. 2021. *The Metaverse's Dark Side: Here Come Harassment and Assaults.* https://www.nytimes.com/2021/12/30/technology/metaverse-harassment-assaults.html

[17] Mel Slater, Aitor Rovira, Richard Southern, David Swapp, Jian J Zhang, Claire Campbell, and Mark Levine. 2013. Bystander responses to a violent incident in an immersive virtual environment. *PloS one* 8, 1 (2013), e52766.

[18] Gareth Terry, Nikki Hayfield, Victoria Clarke, and Virginia Braun. 2017. Thematic analysis. *The SAGE handbook of qualitative research in psychology* 2 (2017), 17–37.

[19] Laurie M Wilcox, Robert S Allison, Samuel Elfassy, and Cynthia Grelik. 2006. Personal space in virtual reality. *ACM Transactions on Applied Perception (TAP)* 3, 4 (2006), 412–428.

[20] Qingxiao Zheng, Yiliu Tang, Yiren Liu, Weizi Liu, and Yun Huang. 2022. UX Research on Conversational Human-AI Interaction: A Literature Review of the ACM Digital Library. *arXiv preprint arXiv:2202.09895* (2022).

[21] Douglas Zytko, Ryan Handley, Bert Guerra, and Rukkmini Goli. 2022. A Taxonomy of Social VR Design. *arXiv preprint arXiv:2201.02253* (2022).