Identity and Safety in Social VR: Findings from an Experimental Avatar-Based Interview Study

Matthias Quent & Sara Lisa Vogl[1]

Introduction

Virtual Reality (VR) has evolved in recent years from a niche technology to a widely adopted medium with numerous applications. The modern use of the term VR was coined by the computer scientist Jaron Lanier in the 1980s (LaValle, 2020, p. 5). According to Steven M. LaValle, "The most important idea of VR is that the user's perception of reality has been altered through engineering, rather than whether the environment they believe they are in seems more 'real ' or 'virtual'. A perceptual illusion has been engineered." (LaValle, 2020, p. 6) This immersive technology enables users to enter worlds vastly different from their physical surroundings, opening new dimensions of experience and interaction. Applications of VR systems range from video games and immersive cinema to telepresence (e.g. Google's Street View) and educational settings (LaValle, 2020, pp. 9–22).

Among the most significant and increasingly relevant applications of VR systems are Social VR platforms. Social VR refers to virtual reality platforms that emphasize interpersonal interactions, enabling users to meet, communicate, and engage in shared activities within virtual environments. One of the most popular and pioneering platforms in this domain is VRChat. Gaining broader popularity since its public release in 2017, VRChat offers a diverse array of user-generated worlds and avatars, allowing participants to interact in an almost limitless range of settings. Beyond real-time voice and text-based conversations, users can partake in activities such as games, karaoke, movie nights, and educational sessions. As of 2023, VRChat has over four million registered users, exemplifying the growing appeal of virtual spaces for social connection and collaborative experiences. Users can access the platform via PC, VR headsets and a mobile app on smartphones. High-quality VR experiences require VR headsets connected to powerful computers**.** The technology is particularly useful for displaying high-quali-

1 With assistance of Lilly Walter.

ty, high-resolution avatars. The visual level of detail can be reduced through software settings. Fully immersive experiences require expensive hardware.

Arne Vogelgesang highlights, that „Social VR as a practice can be understood as embodied social role-playing in a system of networked and bounded virtual 3D spaces inhabited by avatars connected to human users." (Vogelgesang, 2024, pp. 1–2) Krell and Wettmann (2023) explore corporeal interactions within the Social VR platform VRChat, emphasizing that users experience intense bodily closeness and intimacy despite spatial separation. Central to their analysis is the concept of "body synchronization", which describes the alignment of physical movements with virtual avatars, leading to the simultaneous perception of physical and digital embodiment in a "coextensive space." This hybrid embodiment results in phenomena such as "phantom sense," where virtual touches trigger physical sensations. Routine interactions such as hugging, head-patting, and dancing, as well as intimate sexual practices (Erotic Roleplay, ERP), illustrate how mediated corporeality facilitates strong affective and social bonds. The authors argue for a theoretical expansion of interaction theory, stressing that mediatised situations should be understood as situated on a continuum of interactional intensity and corporeal involvement, approaching physical co-presence. While VRChat offers vast opportunities for creativity and social interaction, it also presents significant challenges. Sabri et al. (2023) discuss these issues in their study 'Challenges of Moderating Social Virtual Reality' highlighting the difficulties associated with moderating such platforms. The anonymity and deep immersion provided by VR can foster both positive and negative behaviors, posing a challenge for moderators who must ensure a safe and enjoyable environment without stifling user freedom and creativity. These challenges range from combating harassment and abuse to enforcing age restrictions and ensuring privacy. These challenges are not unique to VRChat but reflect broader societal concerns about immersive technologies. Matthias Quent (2023) highlights in his analysis of democratic cultures in the 'next Internet' that while VR offers significant opportunities at an individual level, it also confronts major social challenges especially at a social level (Quent, 2023, p. 30). He sees immersive technologies like VR as offering significant opportunities, such as supporting psychotherapy for anxiety and depression, fostering empathy and inclusion through perspective-taking, and enabling innovative educational experiences. However, they also pose risks, as seen in platforms like Roblox, where user-generated content can promote discrimination, far-right ideologies, and group-focused enmity, violating community standards (Quent, 2023, p. 38). A study

by Hinduja and Patchin (2024) investigates risks and harms experienced by adolescents within metaverse and social VR environments. Based on a representative survey of 5.005 US adolescents aged 13–17, the authors examine the prevalence of various negative experiences, including bullying, hate speech, sexual harassment, and grooming. Their findings reveal notable gender differences, with girls being more frequently targeted for sexual harassment and predatory behaviors, primarily due to their gender. Despite the availability of safety mechanisms within VR platforms, these are underutilized, although girls employ them slightly more than boys.

Reference to the Metaverse

The "Metaverse" refers to a collective virtual shared space. It is a vision of a fully immersive digital universe that people can inhabit, work, play, and socialize in, akin to a massive, interconnected digital society. Platforms like VRChat and other Social VR experiences serve as early indicators or prototypes of what the metaverse might look like in the future. They allow users to transcend physical limitations, enabling them to interact with others in a virtual universe with diverse worlds and activities. In VRChat, for instance, the idea of user-generated content—wherein participants can create their own avatars, design their worlds, and curate their experiences—aligns closely with the broader metaverse vision. Current social VR platforms pave the way for a comprehensive and integrated digital ecosystem, where physical and virtual worlds become increasingly interconnected. As technologies advance and more platforms emerge, the vision of a unified metaverse becomes progressively tangible, with current social VR platforms serving as its foundational building blocks and environments for learning and experimentation.

Research Design and Data Collection Methods

As part of our experimental research, we aimed to find out why people use Social VR, what experiences they have had there, particularly regarding aspects of identity, safety, and hate speech, and how they protect themselves from harassment. Questions asked in the interviews conducted inside the social VR platform VRChat included aspects of gender, harassment, hate speech, and safety.

To approach these questions, we conducted four qualitative guided interviews as avatars in VRChat in autumn 2023. The anonymized participants come from different countries and were recruited as a self-selected sample from the virtual network of VR designer Sara Lisa Vogl. The interviews were transcribed, coded according to the key questions and analyzed. The results of this explorative pilot research are limited and cannot be generalized. However, they provide first-hand insights into user behavior and experiences in Social VR and thus fill an empirical gap in the primarily theoretical and technological research literature. Further research with more interviewees in virtual worlds can build on this. Through the unique ability of VR to completely alter appearances and virtual embodiment, the interviewees can show up in various identities that are unrelated to their real-life bodies and identities regarding age, gender and abilities.

Key Findings from the Interviews

Social VR Usage

To gain insights into the patterns and diversity of VR usage among our participants, and how they relate to their identity and safety issues inside and outside of VR, we asked interviewees when and for what purpose they have been using Social VR. Duration of usage between participants is 2–5 years in Social VR, mainly VRChat. For some of the participants, the experience of Social VR is an alternative to real-life social interaction, while still performing shared activities and feeling connected to friends. Especially for self-described introverts, meetings in Social VR have the benefit of added layers of control over the environment. People can adjust the number of users visible to them as well as colors and audio of virtual worlds and avatars. They can also completely blend out other users they do not want to see or interact with. They describe how they can adjust their VR experience to their preferences and avoid having to be exposed to people or situations they have difficulties coping with: “I do not need to interact with people that I don't want to, and having that extra bit of control over the situation is just amazing for me, especially because I'm an introvert.”

Additionally, they value the control VRChat offers over their social environment. They can adjust the volume of individual users, hide overly bright avatars, or block individuals they do not wish to interact with. They further highlight that in VRChat, people get to know each other mainly based on

their personality instead of their physical appearance. This aspect helps overcome insecurities related to physical appearance, boosting confidence by focusing interactions on the personality projected through the virtual avatar.

For example, Interviewee 2 discusses using VRChat as a means of coping with social anxiety. They explain that the social interaction within VRChat functions as a form of training in a controlled environment where positive feedback encourages them to engage in social exchanges. Interviewee 3 uses Virtual Reality primarily to detach from the stresses of reality and to relax. They express satisfaction with this approach as it allows them to meet new people, establish friendships, and practice their English skills.

Gender-Identity and Embodiment of Diverse Genders

Another aim of our research was exploring how the interviewees experienced and expressed their gender identity in Social VR. We asked them about their gender identity in real life and how it matched or differed from their virtual avatar appearance and gender identification. We also inquired about their feelings of embodiment and agency in Social VR, and how they were perceived and interacted with by other users. Through this set of questions, we expected to gain insights into the challenges and opportunities of fundamentally free and fluid gender expression and embodiment in Social VR, as well as if and how the according experience relates to their real-life identity and gender expression.

Interviewees describe voice characteristics as the main element of gender identification by other users in VR. They also report that, similarly to reality, some people do not respect their chosen pronouns: "Of course you've run into some people that don't respect that, but you'll run into that anywhere, unfortunately." One Interviewee states that their pronouns and gender identification in VR change with the Avatar they wear and the way they present in VR: "In VR I usually go with the avatar, so when I am a female, I prefer people use she. When I am a male, I prefer he, and when I am an object so like a lamp or stuff like that, it doesn't matter."

This differs from their real life, where they represent their birth gender and use corresponding pronouns only: "No, it's something I just used to do in VRChat – in real life I am not into this stuff and people call me for what is it."

Experienced Harassment or Hate Speech

Harassment – especially sexual harassment – as well as anti-Semitic, racist, transphobic hate comments, or extremist content not only have harmful individual and societal consequences but can also lead people to avoid certain virtual platforms and spaces. Protecting users from harassment and creating an inclusive, human dignity-oriented culture is also particularly important in immersive virtual environments, especially since it must be assumed that the intense experiences through embodiment and virtual reality are experienced more intensively by users than in text-based media. Few studies or figures are yet available on the spread of hate messages in immersive environments, partly because the platform providers (so far) – unlike the particularly large social media platforms – are not obliged to provide public and transparent information on measures against hate. Due to real-time presence, real-time verbal communication, and the absence of recordings, the documentation of hate messages in Social VR is particularly difficult. In the interviews, respondents reported varying degrees of experience with harassment or hate speech in VR.

One interviewee observes “definitely a lot of racism in this game, like in a lot of unfiltered language that is very racist or just offensive to people and genders and all that stuff. It's very offensive." This can be encountered by users in very normal situations. Influenced by political discourse in the world, the interviewee observes mostly hate-filled messages against the LGTBIQ+ community: "Hateful comments and stuff towards the entirety of LGBTQ [community]. [...] Because it's become pretty prominent outside of VR Chat as well [...].”

In Social VR, harassment and hatred can be found primarily in certain groups or spaces. The manner or characteristics to which the experienced hate messages refer are closely related to the virtual appearance as an avatar, so that the user is attacked for their avatar. However, the user relates this to personal identity and does not say “[...] my avatar was attacked for being a furry.” but "I've been hated on for being a furry.". This speaks to a great deal of identification with the avatar and that attacks are perceived personally, rather than as avatar related. In addition, there is gender-based discrimination: "I've been hated on for being a female." and the association of both characteristics with weight-based discrimination: "I've been hated on for being a female furry because apparently that immediately makes you 500 pounds." In such situations, the user makes short work of it: "It's just the block button and then you don't need to talk to those people anymore."

This raises the question of whether avatars and sub-communities in virtual realities give rise to new forms of digital hate communities that are directed against certain classes of avatar identities.

Experiences of harassment are reported by one user mainly in public spaces, which they therefore avoid and their own circle of friends in VR: "Harassment typically comes in whenever I go to public worlds [...] that's one of the main reasons why I don't go to public worlds anymore."

Most notably, the interviewee experiences harassment in verbal form because the user has hidden other avatars by default in the settings: "It's mostly speech, especially because I personally have everyone's avatars turned off by default, and then I have to manually go in and show people's avatars." However, there are also particular glitches related to the specific implementation of the embodiment in the metaverse. For example, offensive images can be used as avatars or high-resolution avatars can cause the hardware to crash due to overload. The user protects themselves against this by disabling avatars by default.

The statements reveal the high technical level of the early VR experts, who developed various tricks to deal with unintended, sometimes malicious interventions in the environments. This addresses key technological challenges to the goal of photorealistic real-time presences or highly authentic avatars: The consumer hardware available today is predominantly not yet capable of achieving this reliably. The varying quality of technological equipment can lead to this being used to the disadvantage of people – consciously or unconsciously – with less powerful hardware – a new form of digital divide in immersive environments. For the development of innovative technologies on the one hand and the establishment of interoperability on the other, this represents a major challenge that is directly relevant to users' sense of security and stability in immersive virtual environments.

Overall, the interviewees agree that while social media and VRChat share some certain similarities regarding harassment, the immersive nature of VRChat intensifies negative interactions, making them feel more personal and intrusive, requiring more proactive measures to manage and mitigate. The small survey indicates that harassment in VRChat is particularly prevalent in public spaces, prompting users to seek refuge in private groups. Common forms of misconduct include verbal harassment and technical exploits, with users frequently needing advanced technical skills to protect themselves. The digital divide exacerbates this issue, with less powerful hardware leaving users vulnerable.

Although VRChat provides various protective mechanisms, their effectiveness remains limited. Users adopt personal strategies, such as avoiding public spaces and employing advanced blocking techniques, to enhance their safety and cultivate a more positive experience in virtual environments.

Blocking as a Personal and Structural Security Strategy

We specifically examined how the interviewees used and perceived the blocking system in VRChat. We asked them about the number and characteristics of people they had blocked, the reasons, outcomes, and the challenges and drawbacks they experienced. Through further investigation of this popular, but profoundly different to real-life protection mechanism, we aim to gain insights into the role and impact of blocking as a personal and structural security strategy in Social VR.

One interviewee states, "I don't think there's really any counter speech, it's just a straighter blocking." The person observes varying levels of sensitivity in dealing with hate speech and in utilizing the blocking function. The interviewee makes a direct comparison to traditional social media, noting that in VR chat there is "[…] definitely not as much censorship in chat as there is on social media [...] But via chat, it's probably the worst. The difference is that you have to be unlucky enough to end up in the same room as those people, whereas with social media, everyone has access to it." Subsequently, the technical barrier to accessing Social VR still provides some protection from being flooded with hate messages compared to traditional social media. At the same time, the interviewee points out that such negative content has a more intense effect in Social VR and requires more active countermeasures: "It's just like I feel like via chat is worse in person [...] But I'd say via chat is way more in your face about and then social media is." Regarding their view of potentially getting blocked by other users, they stated: "I mean, I don't know if I've ever been blocked. It doesn't tell you like when you do. Honestly, it doesn't make a difference to me."

Another person reports about their experiences with blocking as something that can be confusing and even hurtful:

> *"So, I have gotten blocked before [..] for absolutely no reason at all. I know I didn't do anything or say anything hurtful. […] And when I get blocked, if it's from a random person […] sometimes I'll just laugh it off and I'll be like, OK, whatever. They have their reason. Now, if it's someone that's*

a friend who has blocked me before, I'll feel kind of hurt and I'll be like, well, what did I do? What did I do to hurt you? What can I do to make this better?"

One interviewee describes how the blocking feature affects the visibility of avatars of users:

"The biggest thing is that if you block someone, it blocks it for you, but not for everyone else in the instance. So, I've run into things where I blocked someone and then like the poor person standing next to me didn't want to block them for whatever reason. And then they started harassing the person next to me trying to get them to talk to me […]."

Overall, the Interviewees describe blocking as a commonly used tool for managing harassment in VRChat, however, its effectiveness is limited to the user who initiates the block. But the feature can lead to social complications and emotional reactions, depending on the context and the nature of the relationship between users.

Potential Legal Consequences of Actions in Social VR

A central focus of our study was to examine the legal implications of user actions in Social VR, specifically within VRChat. This is a challenging topic because anonymity remains still highly valued by most users, and currently, no Social VR platform implements a mandatory legal identity verification system. We asked users about their awareness and opinions regarding existing and emerging laws applicable to Social VR platforms to understand the challenges and opportunities of legal governance in these environments. We also explored their perspectives on ethical and moral concerns arising from Social VR usage, including issues related to identity, impersonation, consent, and harassment.

One interviewee suggests that there should be certain crimes that are relevant to be prosecuted in the metaverse and ruled out at the user's physical location. Specifically, they highlight that avatar theft and unauthorized appropriation of 3D files as major yet unprosecuted issues, emphasizing: "Now when it comes to the stealing avatars [...] that is a breach of copyright. You are stealing somebody's intellectual property."

Another interviewee sees the safety of minors and the presence of pedophiles on VRChat as the most urgent issues to be tackled on both platform and policy side:

> *"Yes, there are so many pedophiles in this game it is actually ridiculous. […] And they will kind of prowl upon anyone who has some sort of need or wanting for attention, regardless of age. And there are a lot of really social, angsty teenagers in this game aged like 13 to 17, […] some of these guys or girls are really like just sinking their teeth into because they can give them attention that they need in exchange for [...] certain types of like child pornography and like abuse towards minors happening on the platform."*

The interviewees highlighting critical issues such as harassment, avatar theft, and the presence of pedophiles within VRChat. They emphasize the necessity for legal accountability concerning criminal activities in VR, stressing the importance of protecting minors and managing intellectual property rights within the platform.

What Should Policy Makers, Regulators or Designers Know about VR Experiences Like VRChat?

The final question of our study aimed to gather recommendations and insights for policymakers, regulators or designers who are interested in or responsible for Social VR experiences such as VRChat. We asked participants about their expectations and suggestions for enhancing the quality, safety and accessibility of these platforms. Interviewees highlighted several critical issues related to user safety, policy implications, and social inclusion. One participant expressed significant concerns about minors' vulnerability to sexual exploitation by pedophilic users, as well as potential legal repercussions for adults inadvertently interacting with minors. They underline the practical challenges of prosecuting online crimes due to the anonymity and inadequate identification mechanisms prevalent on these platforms.

One interviewee advocated the introduction of mandatory identity verification measures, such as digital ID checks and security recordings, to effectively hold users accountable and facilitate the legal prosecution for online offenses. At the same time, they emphasized the necessity of ensuring cost-free verification processes to prevent the exclusion of economically disadvantaged individuals or users from low-income countries, thus fostering equitable access and social inclusion. Another interviewee stressed the necessity for policymakers and regulatory authorities to gain direct, immersive experiences with Social VR platforms to create informed, user-centric regulatory frameworks. Additionally, this participant pointed out that the heightened sense of immersion characteristic of virtual environments in-

tensifies the urgency for comprehensive safety regulations and policy interventions, acknowledging that virtual interactions carry tangible emotional and psychological impacts for users behind virtual avatars. Interviewees recommended that policymakers and platform designers enhance safety features within Social VR, such as implementing effective vote-kick systems, strengthening protection for minors, and considering mandatory ID verification for improved user accountability. VRChat has now implemented an age verification system through IDs, at least for VRC+ users. The interviewees emphasized the necessity for policymakers to gain first-hand experience to better understand the platform's dynamics and stressed the importance of making security measures accessible to all users, regardless of their economic situation. Additionally, they emphasized the importance of parental supervision to safeguard children from inappropriate content and interactions. They also highlight the unique benefits of VR for improving social interaction skills, particularly for introverts, due to the controlled environment and customizable features available. Furthermore, they advocated higher age ratings for Social VR spaces given the elevated immersion and potential risks, suggesting that regulatory framework should be adjusted to the advanced nature of VR technology.

Discussion

The experimental study is limited by the relatively small sample size of only four participants. Additionally, the sample is restricted in terms of age and gender diversity, although it includes male, female, and diverse gender identities. Geographically, participants represent predominantly the United States and Europe, further limiting the generalizability of the findings. Despite these limitations, the study provides first-hand information and experiences from users and demonstrates that it is relatively straightforward to conduct avatar-based interviews as a researcher in the metaverse. This is especially true when following a community-based approach involving gatekeepers in virtual communities.

Experiencing and expressing personalities independently of physical bodies gives users the opportunity to experiment with social interactions and learn more about them. Friendships, relationships and Smalltalk with strangers are embraced in VR for various reasons, such as users' location or social anxiety. While anonymity in online spaces has significantly impact users' behavior positively and negatively, protective mechanisms are still

underdeveloped, buggy and not the main priorities for platforms. Ensuring minors 'safety through robust identity verification emerges as the most pressing area requiring improvement.

Key concerns include limited understanding among policy-, law- and decision makers regarding the immersive dimension of Social VR. Users emphasize the uniquely immersive nature of VR and its profound social implications, such as fostering friendships, relationships and personal development through prolonged interaction in digital environments.

The surveys clearly indicate that Social VR allows users to experience and express personalities independently of their physical bodies, facilitating experimentation and learning about social interactions. Friendships, relationships, and casual interactions with strangers are highly valued, especially by those geographically isolated or socially anxious. Social VR offers opportunities such as global social interactions with like-minded people, geographical logistical and environmental independence, accelerated learning and collaboration through high levels of immersion, and creative self-expression through interactive tools and accessible audiences. Nonetheless, the medium faces challenges, such as fraud and harassment enabled by anonymity, the particularly dangerous nature of harassment due to VR's immersive quality, and the risks posed by anonymity and insufficient age verification, which exposes minors to age-inappropriate behavior and abuse. Subsequent research could explore socializing in VR compared to physical reality and its impacts on users across different age groups and gender identities.

References

Hinduja, S., & Patchin, J. W. (2024). Metaverse risks and harms among US youth: Experiences, gender differences, and prevention and response measures. *New Media & Society, 26*(1), 1–22. https://doi.org/10.1177/14614448241284413

Krell, F., & Wettmann, N. (2023). Corporeal interactions in VRChat: Situational intensity and body synchronization. *Symbolic Interaction, 46*(2), 159–181. https://doi.org/10.1002/symb.629

LaValle, S. M. (2020). *Virtual reality*. Cambridge University Press.

Quent, M. (2023). Demokratische Kultur und das nächste Internet: Chancen und Risiken virtueller immersiver Erfahrungsräume im Metaverse. In Institut für Demokratie und Zivilgesellschaft (Ed.), *Wissen schafft Demokratie: Schwerpunkt Netzkulturen und Plattformpolitiken* (Vol. 14, pp. 30–43). Institut für Demokratie und Zivilgesellschaft.

Sabri, N., Chen, B., Teoh, A., Dow, S. P., Vaccaro, K., & ElSherief, M. (2023). Challenges of Moderating Social Virtual Reality. In *CHI '23: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1–20). ACM. https://doi.org/10.1145/3544548.358132

Vogelgesang, A. (2024). *Social VR platform design, Nutzerkreativität und ästhetische Kontrolle* (Project Immersive Democracy Working Paper). European Metaverse Research Network.