What do policymakers need to know about harassment in the metaverse?

Verity McIntosh[1,2]*

[1] Bristol VR Lab, University of the West of England, Bristol, United Kingdom
[2] Digital Cultures Research Centre, University of the West of England, Bristol, United Kingdom

*** Correspondence:** Corresponding Author Verity McIntosh, verity.mcintosh@uwe.ac.uk

Keywords: metaverse, governance, abuse, harassment, policymaking.

Abstract

As immersive technologies and spatial computing paradigms move into the mainstream, public and political interest in the metaverse is growing. In some respects, the metaverse offers an exciting view of the future, one in which a global community can meaningfully connect regardless of where they are in the world. In contrast, however, early instances of 'proto-metaverse' spaces have been plagued by reports of harassment and abuse.

Policymakers around the world are now considering the role that governments might play in the regulation and governance of metaverse spaces, seeking to secure protections for citizens, and criminal accountability for offenders in this fast-evolving space.

This paper introduces some of the key issues for governments engaging with this topic, including the suitability of existing legislative frameworks, and consideration of a new category of harm that seeks to recognise the distinctive impact of 'conduct' abuses in metaverse environments.

1 Definition of terms

There is yet to form one coherent definition of the term 'metaverse' and there remains much debate about what is inferred by the term and how it is applied. This paper utilises the X Reality Safety Intelligence (XRSI) definition of the metaverse as:

"A network of interconnected virtual worlds with the following key characteristics: Presence, Persistence, Immersion and Interoperability" (XRSI, 2023)

Forms of harassment discussed in this paper are generally limited to behavioural activity occurring in real time in what might be considered a metaverse, or proto-metaverse environment. This may take the form of verbal or gestural abuse, and/or the use of embodied avatars and virtual objects against other users to enact behaviours experienced as aggressive, violating, offensive or demeaning. An understanding of harassment and abuse could reasonably be extended to include areas such as data and privacy abuses, identity cloning, social and political manipulation, fraud, theft and exploitation. For the purpose of clarity, this paper will focus solely on real time, peer-to-peer encounters involving one or more natural persons in a virtual environment. It should be noted, however that instances of harassment in metaverse contexts may form part of a wider pattern of abuse, taking place both on- and offline and should be considered in such a context when abuses are reported.

2 Harassment and abuse

In recent years, occurrences of harassment and abuse within proto-metaverse platforms, sometimes referred to as social VR platforms, have been well documented in the media, perhaps less thoroughly explored in a scholarly context, and in relation to the role and obligations that governments may have to intervene in this space.

Evidence suggests that instances of harassment tend to increase in virtual environments devoid of managed hosting or a clear purpose, with female users and minoritized people most likely to be targeted (Limina Immersive, 2018). A survey of over 600+ users in 2018 suggested that 49% of regular female VR users reported experiences of sexual harassment or abuse in virtual social spaces (Outlaw, 2018). Since then, with the rise in public adoption of VR headsets, the issue appears to have persisted and perhaps escalated. In 2021 the Center for Countering Digital Hate asserted that users of popular social VR platform, VRChat were exposed to abusive behaviour once every seven minutes (Center for Countering Digital Hate, 2021). Numerous reports of sexual harassment and abuse within the metaverse have been reported in the media (Eccles, 2022; Patel, 2021; Rifkind, 2022).

3 Impact

Although the nature of harassment and abuse in VR differs from real-world instances, the impact on individuals can be significant. Slater calls attention to the confluence of psychologically convincing Place Illusion (PI) and Plausibility Illusion (Psi) in virtual reality, giving users a strong sense of presence, and implicating their body in the virtual space. “If you are there (PI) and what appears to be happening is really happening (Psi) then *this is happening to you!* Hence you are likely to respond as if it were real. We call this ‘response-as-if-real’ RAIR. (Slater, 2009)

Several researchers have pointed to the compounding impact of “social presence” (Lee, 2004; Ratan, 2012) i.e. the awareness of being co-present with other users, conversing and taking consequential action in a shared, virtual environment. This attribute is often understood in combination with “self-presence” and “environmental presence”, cumulatively forming a powerful sense of “being there” that has been identified as particular to virtual reality (Bailenson, 2018). Ratan has suggested that social presence might be considered to be the most impactful of the three, as the participation of other natural persons in a virtual space adds complex social cuing to the simulative environment, further convincing users of the veracity, immediacy and embodied nature of their experience (Ratan, 2012).

The UK’s Cyberpsychology Research Group call attention to the contiguous emotional impact of negative experiences in metaverse environments “Just because these events happen online rather than offline doesn’t mean they are not being experienced as real” (Askham, 2022). Madary & Metzinger take it a step further, introducing the possibility that “[t]orture in a virtual environment is still torture. The fact that one’s suffering occurs while one is immersed in a virtual environment does not mitigate the suffering itself” (Madary & Metzinger, 2016).

In the context of all of the above it seems likely that the immersive and embodied nature of social, metaverse environments will significantly intensify the impact of harassment and abuse such as physical threats or simulated violence. In metaverse environments, non-consensual instances of touching, verbal harassment or invasion of personal space may put users at particular risk of psychological and emotional distress. Future developments such as haptic technology clothing may further heighten this affect by adding a physical sensation to abuse enacted in metaverse contexts in the future.

Even without the use of specific haptic technology, there is evidence to suggest that some people, using only a headset and controllers, experience uncanny physical sensations upon being touched in virtual environments. Some hypothesise that the psychologically convincing nature of metaverse environments

can lead users to partially transfer their phenomenal self model (PSM) into that of an avatar , an effect akin to the Rubber Hand Illusion (Botvinick & Cohen, 1998). Consequently, they may report feeling pronounced physical sensations when they observe their avatar being touched or harmed, even though their physical body remains uncontacted (Desnoyers-Stewart et al., 2024; Madary & Metzinger, 2016; McIntosh & Allen, 2023).

A comparable phenomenon, 'phantom touch', is frequently discussed by users of social VR. Although largely under-researched in a formal setting, this sensation appears to involve users perceiving a touch sensation on their bodies that directly corresponds to a simulated act of touch in VR. Some users actively cultivate this sensation using virtual mirrors in order to associate avatar touch with tactile sensation. There would appear to be enhanced likelihood that those who experience a form of 'phantom touch' could be at greater risk of traumatic impact in the event of harassment and abuse (McIntosh & Allen, 2023).

One often-posed question in regard to VR abuse, from those not familiar with the technology is, 'why didn't you just take the headset off?'. Preliminary research suggests that rapid disengagement from VR, particularly under stress or anxiety, can provoke panic or dissociative episodes, therefore, the solution may not be as simple as disconnecting (Allen & McIntosh, 2022). This question also signals a tendency towards victim blaming, failing to account for well understood trauma-response behaviours such as freeze and appeasement in response to high stress, high risk encounters (Cantor & Price, 2007)

4 Design responses

In response to apparent abuses in proto-metaverse spaces, many app developers and platform owners have sought design solutions to mitigate the risk or severity of potential harms. Some have turned to social science research that may not have been initially conceived in relation to technology paradigms, drawing on research exploring physical and relational behaviour as a route into understanding the needs of social, virtual spaces.

Hideaki Matsui, a design lead at Google has publicly discussed their use of Proxemics (Hall, 1966) Hall's theory of Proxemics suggests that people will maintain differing amounts of distance from one another depending on the social setting and their cultural backgrounds. Google use this framework as a schematic, encouraging designers to construct virtual environments that conserve distances between users that are appropriate to the social context and levels of intimacy that might be anticipated in a particular encounter. As per Hall's design, they distinguish between public, social, personal and intimate space and design experiences accordingly. Their approach notably does not incorporate Hall's framing of such boundaries being informed by background and cultural context.

Michelle Cortese, Design Lead Manager at Meta extends the use of Proxemics to incorporate consent frameworks inspired by the BDSM community. She writes about the significant number of people, particularly women, who reported being sexually harassed or assaulted in multi-person virtual reality spaces in the late 2010s, and calls for an approach to personal space management that involves explicit and informed mutual consent.

"we suggest designers build granular controls that are easy to access and surface before intimate interactions begin. It's important that people can customize and control the types of experiences they're willing to have with other people in these close quarters before they happen" (Cortese & Zeller, 2019)

In the intervening years, many of these recommendations have been adopted, with features such as 'personal space bubbles' now available in most social VR apps. Personal space bubbles enforce an

invisible boundary around the user, keeping other avatars at a designated distance, or rendering them invisible and inaudible when the allotted space is impinged. In some instances, users can choose only to be perceptible to those pre-designated as ‘friends’ to minimise the risk of harassment.

Whist such design features may prove useful, they can also create an imbalance of power that favours the aggressor. The onus is on the victim to apply extreme caution entering into a metaverses space, configuring complex safety features and limiting their own experience prior to entry, or attempting to do so in the moment whilst experiencing harassing behaviours. Those persistently harassing other users, notionally violating the terms of use of the platform, encounter no such barriers.

For victims of abuse; block, mute and report tools may be available, and are designed be deployed ad hoq in the event of unwanted attention or abusive behaviour. Such reporting features can be difficult to navigate in the moment, especially when abuse is ongoing. It is also generally unclear what responses or punitive measures might follow from the reporting of such instances. To date platforms are not obliged by any regulatory authority to make transparent their internal monitoring, evidentiary and justice systems, to disclose actions taken to investigate or remediate reports of abuse, or to notify the complainant of any actions taken (Allen & McIntosh, 2022; Center for Countering Digital Hate, 2021).

5 Regulation and governance

5.1 Suitability of existing laws

Around the world, governments are seeking advice on whether existing legislature is sufficient to ensure that their citizens are afforded the same rights, protections and freedoms in metaverse spaces as they might expect in comparable physical and digital spaces.

One key, anticipated challenge to efficacy, is that many legal frameworks related to abuse and harassment make clear distinctions between ‘content’ abuses which can include the posting and sharing of abusive materials such as text, imagery and video, and physical ‘contact’ abuses, which generally involve unwanted physical touch.

Several governments have sought to improve protections for citizens in 2D online platforms in recent years. New criminal designations are being written onto the statute books for online criminal behaviour such as ‘cyber-flashing’ and the posting of ‘revenge porn’(Online Safety Act 2023, 2023). In the relatively new field of multi-person, metaverse environments, there is currently little legislative provision to account for abuses that might take place in psychologically convincing, simulative environments where multiple natural persons are co-present and interacting with one another.

Given what is understood about the immersive, embodied and relational qualities of metaverse environments, governments may need to specify a new category of harm. Perhaps one that recognises certain forms of user ‘conduct’ as harassment and abuse, even where there is no physical contact, or associated production or proliferation of content.

5.1.1 Case study

As an early test of the suitability of existing legislature, UK police announced in January 2024 that they were investigating an alleged instance of ‘sexual attack’ of a girl who is under 16, and was abused by a group of men in a social VR setting. (Camber, 2024)

In an interview with LBC News, The UK's Home Secretary, James Cleverly said "I know it is easy to dismiss this as being not real, but the whole point of these virtual environments is they are incredibly

immersive. We're talking about a child here, and a child has gone through sexual trauma. It will have had a very significant psychological effect and we should be very, very careful about being dismissive of this." (Taylor, 2024)

In response to this case, the chairman of the UK's Association of Police and Crime Commissioners, Donna Jones was reported as saying "We need to update our laws because they have not kept pace with the risks of harm that are developing from artificial intelligence and offending on platforms like the metaverse." (Taylor, 2024)

The statements of two such prominent public figures signals an appetite at policy level to apply some of the principles discussed in this paper at the highest levels of governance. This specific case is understood to be ongoing at the time of publish. It will be interesting to see how existing legislation is applied and reconciled in this seemingly unprecedented case.

5.2 Accountability

Policymakers may wish to consider creating stronger links between activity in the metaverse and national law enforcement agencies. This would ensure that serious crimes committed in metaverse worlds don't remain under the exclusive jurisdiction of the platform's internal justice system, which is arguably better suited to technology-related issues than serious criminal offences. Public confidence will also need to be built such that anyone reporting abuses to civic authorities can expect to be understood, believed, and for their complaint to be acted upon.

Criminal prosecution of individuals for abusive conduct in the metaverse is one area that governments certainly need to consider. Another is the relative culpability and accountability of the companies providing metaverse apps, platforms and services. Where frequent instances of criminal activity, such as abusive behaviour are evidently taking place in a particular app or platform, regulators may wish to consider holding providers wholly or partially accountable. Particularly if they are failing to uphold terms of use, and encouraging or turning a blind eye to abusive behaviour.

In the US, holding platforms to account is likely to prove challenging. Section 230 of the Communications Act affords legal immunity for providers of interactive computer services with respect to the actions of their users (Section 230, 1934) In the UK, the new Online Safety Act (Online Safety Act 2023, 2023) has some provision for this, extending a “duty of care” to platform owners regarding what content users, particularly children should be able to encounter online. The challenge of ‘content’ versus ‘conduct’ and ‘contact’ is largely unaddressed in the Act, however the metaverse has been deemed explicitly in scope (Local Government Association, 2022). The Institute for Engineering and Technology recently called on UK government to ensure that new legislation is made fit for purpose in relation to social, spatial environments (Almond et al., 2024). The EU’s new Digital Services Act (European Union, 2023) goes further still, holding very large tech companies legally accountable for the content posted on their platforms. Again, the Act sets out a framework for addressing illegal ‘content’ online, however there is no direct provision for metaverse contexts, and it remains unclear how the more behavioural, conduct-based forms of abuse and harassment might be addressed by this new legal framework.

5.3 Jurisdiction

In most legislative frameworks, sovereign jurisdiction is determined by the geography of where an alleged crime has taken place. For many exponents of the metaverse, the promise of this new paradigm lies in its potential to be borderless and decentralised. Just as cryptocurrency could be conceived as an

alternative to centralised banking systems, so the metaverse might be imagined as an alternative to state-based territoriality for interpersonal encounters. What then for state-based authorities looking to respond to reports of criminal activity, including reports of harassment and abuse in the metaverse?

As with the internet before it, questions of jurisdiction in metaverse contexts are proving challenging. Users of such spaces may be encountering one another in what experientially is a common metaverse environment, but connecting from very different territories, each with their own particular legal contexts. To further complicate matters, the metaverse environment visited might be provided by a company in another territory, with the underpinning technology stack hosted across multiple territories. What legal frameworks should then apply when abuses are detected? And which nation(s) should have the jurisdiction to prosecute criminal behaviour?

Laws governing interpersonal behaviour vary considerably between territories, and jurisdictional ambiguity can create a vacuum of legal accountability, a lag in governmental response to evident harms, and a gulf of support for victims of criminal behaviour.

Even in instances when jurisdiction is relatively unambiguous, or where laws can be expected to be common across territories, challenges can remain. For instance, most legal systems descended from English law e.g. Australia, Canada, New Zealand, Singapore and the United States, conform to similar systems of Tort law (civil laws pertaining to interpersonal wrongdoing between private persons). However, it remains unclear whether such laws would be legally applicable in metaverse contexts as the legal ‘personhood’ of an avatar is yet to be determined. Questions remain regarding whether the actions of an avatar in a virtual world should be considered directly analogous to the action of the embodied 'natural person' controlling it. Or whether avatar behaviour would be better understood as akin to a playable video game character (Cheong, 2022). Each approach would attract a very distinct legal response, particularly in relation to acceptable levels of interpersonal violence.

In the absence of legal certainty there is concern that cases of abuse and harassment may become entrenched in costly, intractable disputes regarding which legal jurisdiction applies, risking a drain on resources in multiple territories and lessening the likelihood of successful conviction (Europol, 2022; Kalyvaji, 2023).

One approach would be to make platforms responsible for ensuring that the legal protections of each user are implemented in the design of the space before they are granted access to a given metaverse environment. Where legal frameworks in different jurisdictions prove incompatible, this may lead to citizens from certain territories being excluded, or companies running multiple instantiations of metaverse environments, the user being directed to the space that is compliant with their domestic legal system. An alternative, or addition perhaps, is to encourage closer working with international agencies such as Interpol to ensure the complementarity of different governmental approaches, and to enhance international cooperation agreements to support cross-jurisdictional prevention and response to crimes involving metaverse technologies and environments.

5.4 Stakeholder literacy

Among the most immediately actionable opportunities for government agencies engaging with this topic, is to improve stakeholder literacy. This could be achieved by training programmes, giving stakeholders direct experience of embodied metaverse platforms, providing insight into the current trajectory and pace of technological developments, and the manner in which the affordances of this medium relate to issues of abuse and harassment. Governments may wish to consider prioritising the literacy of responsible bodies such as legislators, police and the judiciary. Public literacy campaigns

may also be valuable in supporting citizens to understand their rights, and empowering them to make informed and empowered choices about their own engagement with the metaverse.

6 Conclusion

Although the metaverse is often positioned as a 'future horizon' technology, it is evident that early versions of the metaverse are already here, and that instances of harassment and abuse are taking place with potentially significant consequences for citizens. Governments have an opportunity to urgently consider the suitability and efficacy of existing legislature, and to assess whether new legal instruments are needed to reflect the distinctive experience of embodied, immersive, multi-person environments. Policymakers may also wish to consider prevention, reporting and prosecution strategies, as well as the accountability of both individuals and platforms/service providers in relation to abusive behaviours in metaverse environments. Programmes of metaverse literacy now could equip stakeholders and the wider public with the information they need to collectively design and advocate for more positive futures for the metaverse.

7 References

Allen, C., & McIntosh, V. (2022). *Safeguarding the metaverse*. https://www.theiet.org/media/9836/safeguarding-the-metaverse.pdf

Almond, E., McIntosh, V., & Allen, C. (2024, January 3). *An open letter to Ofcom on the need to urgently review how VR spaces are governed*. The IET. https://www.theiet.org/media/press-releases/press-releases-2024/press-releases-2024-january-march/3-january-2024-an-open-letter-to-ofcom-on-the-need-to-urgently-review-how-vr-spaces-are-governed

Askham, G. (2022, April 26). Metaverse: New Documentary Exposes Racial & Sexual Abuse. *Glamour*. https://www.glamourmagazine.co.uk/article/metaverse-misogyny

Bailenson, J. (2018). *Experience on demand : what virtual reality is, how it works, and what it can do* (First edit). W.W. Norton & Company,.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669). https://doi.org/10.1038/35784

Camber, R. (2024, January 1). British police probe VIRTUAL rape in metaverse. *Daily Mail*.

Cantor, C., & Price, J. (2007). Traumatic Entrapment, Appeasement and Complex Post-Traumatic Stress Disorder: Evolutionary Perspectives of Hostage Reactions, Domestic Abuse and the Stockholm Syndrome. *Australian & New Zealand Journal of Psychiatry*, *41*(5), 377–384. https://doi.org/10.1080/00048670701261178

Center for Countering Digital Hate. (2021, December 30). *New research shows Metaverse is not safe for kids*. Center for Countering Digital Hate (CCDH). https://counterhate.com/blog/new-research-shows-metaverse-is-not-safe-for-kids/

Cheong, B. C. (2022). Avatars in the metaverse: potential legal issues and remedies. *International Cybersecurity Law Review*, *3*(2), 467–494. https://doi.org/10.1365/s43439-022-00056-9

Cortese, M., & Zeller, A. (2019). *Designing Safer Social VR Using the ideology of sexual consent to make social VR a better place*. https://immerse.news/designing-safer-social-vr-76f99f0be82e

Desnoyers-Stewart, J., Bergamo Meneghini, M., Stepanova, E. R., & Riecke, B. E. (2024). Real human touch: performer-facilitated touch enhances presence and embodiment in immersive performance. *Frontiers in Virtual Reality*, *4*. https://doi.org/10.3389/frvir.2023.1336581

Eccles, L. (2022, January 22). My journey into the metaverse — already a home to sex predators. *The Sunday Times*.

European Union. (2023). *The Digital Services Act*. https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

Europol. (2022). *Policing in the metaverse: what law enforcement needs to know, an observatory report from the Europol Innovation Lab*. https://doi.org/10.2813/81062

Hall, E. T. (1966). *The Hidden Dimension*. NY Doubleday.

Kalyvaji, M. (2023). Navigating the Metaverse Business and Legal Challenges: Intellectual Property, Privacy, and Jurisdiction. *Journal of Metaverse*, *3*(1), 87–92. https://doi.org/10.57019/jmv.1238344

Lee, K. M. (2004). Presence, Explicated. *Communication Theory*, *1*(1), 27–50. https://onlinelibrary-wiley-com.ezproxy.uwe.ac.uk/doi/epdf/10.1111/j.1468-2885.2004.tb00302.x

Limina Immersive. (2018). *Immersive Content Formats for Future Audiences*. www.digicatapult.org.uk

Local Government Association. (2022, April 19). *Online Safety Bill, Second Reading, House of Commons*. https://www.local.gov.uk/parliament/briefings-and-responses/online-safety-bill-second-reading-house-commons-19-april-2022

Madary, M., & Metzinger, T. K. (2016). Recommendations for Good Scientific Practice and the Consumers of VR-Technology. *Frontiers in Robotics and AI*, *3*. https://doi.org/10.3389/frobt.2016.00003

McIntosh, V., & Allen, C. (2023). *Child Safeguarding and Immersive Technologies: Key Concepts*. https://learning.nspcc.org.uk/research-resources/2023/child-safeguarding-immersive-technologies

Online Safety Act 2023 (2023). https://www.legislation.gov.uk/ukpga/2023/50/enacted

Outlaw, J. (2018). *Survey of Social VR Users*. The Extended Mind. https://www.extendedmind.io/2018-survey-of-social-vr-users

Patel, N. J. (2021, December 21). *Reality or Fiction?* Medium. https://medium.com/kabuni/fiction-vs-non-fiction-98aa0098f3b0

Ratan, R. (2012). Self-presence, explicated: Body, emotion, and identity extension into the virtual self. In *Handbook of Research on Technoself: Identity in a Technological Society* (pp. 321–335). IGI Global. https://doi.org/10.4018/978-1-4666-2211-1.ch018

Rifkind, H. (2022, February 21). The metaverse will be an abuser's paradise. *The Sunday Times*.

Section 230 (1934). https://crsreports.congress.gov

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions. Biological Sciences*, *364*(1535), 3549–3557. https://doi.org/10.1098/rstb.2009.0138

Taylor, W. (2024, January 2). Police investigate “rape” in metaverse after group of men attack girl in virtual reality room. *LBC News*.

XRSI. (2023). *The Metaverse - X Reality Safety Intelligence (XRSI)*. XRSI. https://xrsi.org/definition/the-metaverse