Written evidence submitted by The University of Manchester

Connected Tech: Risks to children and other potentially vulnerable users in eXtended Reality

Summary

- We welcome the DCMS Call for Evidence on Connected Technology. Our submission focuses on eXtended Reality (XR) technologies (sometimes marketed as "the Metaverse") and in particular the risks to children and other potentially vulnerable groups. This addresses three topics in the Call for Evidence: **the impact on potentially vulnerable groups**, the importance of **'safe and secure by design'**; and understanding the key **short- and long-term risks and threats.**

- In a recent report, commissioned by GCHQ, we examined how child exploitation and abuse might evolve in light of the increasing adoption of immersive XR technologies. Our report synthesises research into online child sexual exploitation and abuse with current research and developments in XR, bringing together University of Manchester expertise from computer science, psychology, and criminology. Although our focus was on risks to children, our findings are relevant to other potentially vulnerable groups, and to other criminal and harmful activities in XR.

- We find that XR technologies potentially offer new routes for offenders to access, exploit, and abuse children. Anecdotal reports suggest that some offenders are already taking advantage of these opportunities. More generally, reports suggest that harassment and bullying of both adults and children is widespread on such platforms.

- Shared XR experiences exhibit many of the same safety risks and potential for abuse and bullying as other social media, however the impact can be amplified by the visceral engagement experienced in immersive environments.

- Moderation and safeguarding measures are currently inadequate to protect people in conventional online social interactions; the complex nature of interactions and activities in XR mean effective moderation is currently impossible to achieve in a scalable manner.

- We are concerned that current and proposed legislation may not be sufficient to address these threats. The Online Safety Bill, for instance, should address XR safety explicitly, detail how XR platforms can assess and mitigate risks, and mandate the development and implementation of safety standards.

XR Technology

1. XR describes a spectrum of technologies that include:

    a. Virtual Reality (VR) — an artificially mediated, immersive experience, usually simulating a 3D world with sensory input from the physical world and displayed by a stereoscopic headset that blocks out the user's surroundings.

    b. Augmented Reality (AR) — viewing digital content overlaid on the physical world, perhaps via a smartphone app or through a specially designed headset.

    c. Mixed Reality (MR) — experiencing a believable synthetically generated 3D environment blended with the physical world around the user.

2. VR hardware can create rich, believable, three-dimensional, immersive experiences. By engaging with our senses and perception at a fundamental level, this is appealing for a range of activities from training and simulation of engineering tasks, through the treatment for deep-rooted phobias and conditions such as phantom limb pain, to entertainment and games. When implemented well it feels *real*, and enables the transfer of experience and learning back to the real world. However, **VR adds an additional intensity to abuse experienced online.** Unwanted interactions can seem much more real and impactful when another person is invading close personal space, or making obscene gestures or postures with their avatar up against yours because we react instinctively in a manner similar to how we would in a real physical interaction.

3. It is critical that safety standards and **approaches to moderation should be context-dependent and nuanced** to avoid unintended consequences. For example, young teenagers have created shared, private (invite only) virtual worlds within applications to provide themselves with venues to talk openly with supportive peers about a wide range of issues, such as mental health, bereavement, their developing sexuality, experiences of abuse and trauma, and as a way to handle isolation experienced during the pandemic. A crude, automated ban on discussing certain topics in the presence of under-18s could inadvertently remove such safety nets.

4. **Practical mechanisms** to improve safety in XR should be the initial focus. While eliminating harmful behaviour may be most desirable, achieving this is extremely hard to achieve and requires more research and development. Meanwhile many XR applications have had some success with defensive measures, notably being able to mute other people and to define 'protective bubbles' around a participant that cause other people to vanish if they enter into this personal space, thereby avoiding experiencing intensely unpleasant virtual contact.

5. **Creating rich 3D content is an important digital skill** for children to develop, and the UK is renowned for utilising such abilities in its vibrant and successful video game and digital media companies. At the same time, user-created content in XR poses

obvious risks from the introduction of harmful material. The technology to automatically identify this in 3D models (which may animate, changing shape and surface appearance over time) requires further research and is considerably less mature than that for analysing photographs or other images.

Risks to children and adults

6. Recently, public attention has been drawn to risks to children in VR. For instance, the Centre for Countering Digital Hate, in their written evidence for the Online Safety Bill based on their research, highlighted how users who were (or appeared to be) under 10 years old were approached by adults, exposed to pornographic and violent content, and to abusive and explicit comments. Similar findings have been reported by journalists 'undercover' in VR social spaces, and adults also report distressing harassment and abuse in VR.

7. Reducing risks for young people in XR is challenging, not least because **many of these applications were not specifically aimed at children and thus were not primarily designed with child safety in mind.** (Children are typically not considered as part of Equality, Diversity and Inclusion.) Platform providers like Meta and Sony state that young children (under 13 or under 12 respectively) should not use their headsets. This is a crude attempt to absolve the companies of the need to ensure that applications are safe for young children, but also has the undesirable consequence of eliminating valuable opportunities to develop informative and educational virtual experiences for younger age groups.

8. Even with manufacturer-specified age limits, **XR devices are inevitably going to prove interesting and accessible to young children within households**. As with other online activities, it is not difficult for curious under-18s to seek out adult experiences, including sexually oriented content and social spaces, and for adults to use XR social spaces as a new venue to initiate grooming. Purchasers of new XR equipment should be prominently made aware of the parental controls available on the platform, how to use them, and why this might be necessary. For children, **XR should be considered alongside other online activities in digital citizenship lessons** from early in their education.

Current attempts to reduce harm are inadequate

9. As with other forms of social media, current attempts at moderation are problematic:

    a. Users can block and report other users, but that can be time-consuming and difficult, and **dangerous behaviour is not always immediately obvious or demonstrable** but can take the form of a pattern of behaviour over an extended period (grooming for example).

b. Once a user has reported someone, they **rarely get feedback** on what action has been taken. Without feedback users may question whether reporting further incidents is worth their time.

c. If a child is using someone else's headset, they may not want to draw attention to themselves by contacting the platform. Even if they did report an incident, it may appear to come from the adult owner and not the child, and therefore be dismissed as not being a child safety issue.

d. **Proving that abusive or dangerous behaviour has occurred can be impossible** if users do not proactively save a recording of their recent interactions in the short window of opportunity provided by the platforms immediately following the abuse. Unlike text chat logs, live XR interactions may leave very little evidence for subsequent investigation.

e. Social spaces can be 'policed' by human moderators (for example, in some spaces human moderators prevent trolls disrupting meetings). But it is impossible for human moderators to be present in every XR social space, and the idea that someone is around always 'listening in' is likely to reduce the appeal of the platform.

f. **Reporting abuse is frequently used as a way to attack or bully opponents** online, often by 'brigading' large groups of people to join in reporting a user. Crude measures of reputation based on the number of reports received about an individual are likely to encourage this and cause further harm to victims. More research is required, and the role of automated decision making needs to be closely monitored.

10. In tackling abuse, it may be tempting to require users of online multiuser XR environments to use verified real names to reduce the temptation to engage in abusive or bullying behaviour. However, **this puts potentially vulnerable users at risk from others being able to link their online identities with their real identity**. Typically known as 'doxing', there are numerous examples of non-immersive online gaming disputes leading to physical harassment, fraudulent fast-food orders, and most seriously, 'swatting', the causing of armed police to attend a house through fake distress calls. In relation to children, being able to link a child's online identity with real world information, such as a parent's public Facebook profile, could reveal photographs of the child, the school they attend, interests and likes, names of friends, and so on, all of which would significantly aid grooming. **Providers should be required to protect all externally linked identities**, including in other related online information such as game league tables, player profiles, responses to abuse reports, and emails. (Even without directly linking to another account, someone reusing a public nickname they have used elsewhere may provide enough information to make the connection to their real identity; this kind of danger is something that applies more broadly than just in XR and should be taught as part of a digital citizenship curriculum.)

Widening Opportunities for Abuse

11. **Our analysis suggests that risks to children go beyond being exposed to inappropriate content and potential grooming**. Recent years have seen the development of immersive video, erotic games, sexually oriented social spaces, and a burgeoning sex-tech industry, largely focused on the use of haptic devices which replicate real world sensations and create tactile user experiences. These developments are, of course, legal in most jurisdictions, and have benefits, for example, enabling quasi-realistic sexual experiences for people who might otherwise lack sexual contact, and as an outlet for creativity and play. However, they could also be exploited in grooming and child sexual abuse. For instance, numerous conventional web-based adult chat rooms enable live-streamed sexual performances that can be influenced by viewers ('camming'). This is a format exploited by CSA offenders to live-stream child abuse, and we can expect adult VR camming to serve as a model for live VR child abuse.

12. Another area of concern is computer-generated sexualised images of children in XR. Simulated abuse against child avatars has been noted in online virtual worlds for many years. We would not be surprised to see adult users adopt or create child-like avatars and personas to be used in simulated sexual activity in AR and VR, much as some already do using clothing and props in real adult encounters. Although actual children may not be harmed directly, these virtual **depictions in immersive environments can normalise the idea of sexual abuse of children.**

13. Simulated sexual activity in XR, including immersive VR camming, takes place in real time and may not result in a recording or other meaningful digital footprint. Platforms will face significant technical challenges in detecting whether such activity is taking place and whether real children are involved. Where illegal activity (such as abuse against children) is detected, how will this be evidenced for a prosecution? This challenge is exacerbated where different companies share control of different parts of interconnected virtual worlds using different software platforms. This interoperability is a core promise of the marketing surrounding the 'Metaverse' and already exists in some areas where a single shared social virtual environment can be accessed from multiple platforms simultaneously. This greatly complicates reporting experiencing abuse. We suggest that **regulators, academia and industry should cooperate on suitable technical standards for securely collecting tamper-proof evidence** while preserving privacy and masking identities of innocent parties.

Implications for regulation

14. Consumer adoption of XR technologies will continue to grow, driven by improvements in mobile augmented reality and internet capabilities, development of better hardware, reduced costs, and the increased availability of high-quality immersive content. Industry commentators predict that use of XR tools will be commonplace in a few years, but issues of user safety need addressing now. We note:

a. The Online Safety Bill places an obligation on platforms that host potentially harmful material to carry out risk assessments. Our analysis shows that **platforms need to consider several types of risk** to children, not just the widely publicised risk of grooming.

b. The Online Safety Bill requires platforms to have effective ways of mitigating risks. But **comprehensive moderation of content and activity is currently impossible in XR**. Human moderators cannot be everywhere in VR and automated detection of harmful behaviour in real-time online XR interactions is not technically possible. At present, it appears that the Bill will place an obligation on technology companies providing XR services that may be impossible to meet.

c. We should be supporting and accelerating work to develop standards and guidance, to **support XR 'safety by design'**. Technology firms are only just beginning to tackle grooming and other abuse-related activities in XR environments. Technology companies and XR app developers should anticipate and mitigate safety issues before their products are rolled out. **An effective way to include children is to include them as standard in Equality, Diversity and Inclusion.** Platforms that host VR and XR app stores should only accept apps that meet safety standards.

d. A key concern is determining how abusive and dangerous behaviour in XR can be identified and tracked and, if necessary, offenders prosecuted, **while still protecting freedom of expression and user privacy**. This will require innovative thinking and new tools for digital investigation and digital forensics, and consideration of how evidence of abuse activity in XR could be laid before a jury.

e. We are pleased to see companies like Meta begin to introduce VR **parental safety tools** for teenagers. But these must be rolled out more quickly, explained to parents more prominently, and new measures need to be developed to protect pre-teens where existing reporting mechanisms may be too complex.

f. Meanwhile, **more guidance and advice about the risks of XR** technologies is crucial. In particular, children and their caregivers need more education so they can make informed decisions about when and how to let children explore the rich opportunities of immersive virtual worlds safely.

21 July 2022
Professor Emma Barrett
Professor Steve Pettifer
Dr James Marsh