VIRRAC

Virtual Reality Risks Against Children

ROUNDTABLES REPORT













INSTITUTE FOR CONNECTED COMMUNITIES (UEL)
CENTRE FOR ABUSE AND TRAUMA STUDIES (MDX)

Professor Julia Davidson, Professor Elena Martellozzo Dr Ruby Farr, Paula Bradbury and Boglarka Meggyesfalvi



CONTENT

	_
CONTENT	2
ACKNOWLEDGEMENTS	2
INTRODUCTION	3
FINDINGS	4
1. Safety Awareness amongst children	6
2. Challenges in digital literacy	6
3. Behaviour	7
4. Psychological Impact	8
5. Neurodiversity	8
6. Legislation and policing	9
RECOMMENDATIONS	9
1. Review the applicability of current online safety legislation	10
2.1. Enhance content moderation efforts and training	10
2.2. Al-driven content moderation	11
3. Educator's resources	11
4. Safeguarding standards and evidence-based, general information	12
5. Parental involvement	13
CONCLUDING REMARKS & NEXT STEPS	14

ACKNOWLEDGEMENTS

This report shares headline findings from online roundtables, conducted as part of REPHRAIN's 'Small Impact' funding call. The team would like to formally acknowledge and thank the REPHRAIN centre for their ongoing support and encouragement. The team also acknowledges the 13 participants who took part in the Education, Industry and other roundtable interactive sessions, for their invaluable time, knowledge and expertise shared.

INTRODUCTION

Study Background

This report relates to a previous project, titled VIRRAC (Virtual Reality Risks Against Children). By placing children at the heart of our research and data collection methods, VIRRAC aimed to raise awareness of child safety in the metaverse (UNICEF, 2025). Our primary focus was to understand the risks and harms that children and young people encounter in these immersive virtual environments, ensuring their voices and experiences shaped our approach. By doing so, we sought to explore practical, child-centred solutions that can enhance safety and well-being in the metaverse. VIRRAC came to an end in March 2024, and additional funding was awarded by the REPHRAIN centre to carry out two online roundtables with professionals working in the education, and industry and other sectors.

Context and Aim

The research team proposed two interactive virtual workshop sessions with education and industry specialists. The format of these sessions was explorative and participatory (<u>Bastian, 2016</u>), whereby key findings and recommendations from the VIRRAC project were shared with attendees to inform facilitated discussions centred on applying and prioritising the VIRRAC recommendations. The virtual roundtable sessions provided a meaningful platform for engaging with key strategic stakeholders to critically examine the implications of the core VIRRAC findings and outputs.

Methods

Ethical approval was principally sought from the University of East London ethics committee and then from the REPHRAIN centre. The convenience and the snowball sampling method were used, whereby participants were recruited initially through existing networks, and were encouraged to share roundtable information with relevant colleagues and wider networks (<u>Parker et al., 2019</u>).

In order to curate a comprehensive set of questions, the two roundtable guides were developed by the team based directly on learning derived from the original VIRRAC research whilst also taking the project's findings and recommendations into account.

The VIRRAC Report Executive Summary was shared with participants before the roundtable sessions, and the team presented key findings and recommendations at the start of each event before commencing. Roundtables took place online in a secure channel, via Microsoft Teams. Transcripts were fully anonymised before being uploaded to ATLAS software for deductive thematic analysis.

FINDINGS

KEY POINT ARISING FROM THE ORIGINAL RESEARCH

Environment and Control

Behaviour modelling

Guidance

The initial findings suggest that experiences of harm, coercion to disclose personal information, and exposure to inappropriate content may be relatively prevalent among children in the metaverse. Both experts and young participants in the research unanimously acknowledged the need for more guidance, increased awareness, adequate support, and robust online safety features to better protect young users in virtual reality settings. Findings presented to participants drawn from our work with children and young people were grouped into 3 categories:

- I. **Environment and control:** Emphasising that children want and need clear boundary lines; should ideally be using devices in common areas and not hidden away; should not be on devices late at night, and need restrictions on time lengths for device usage.
- II. **Behaviour modelling:** Highlighting that children want behaviour modelling information to a) comprehend what the warnings are around app behaviours; b) learn about what positive and negative behaviours look like; c) understand how online behaviour affects others and themselves.
- III. **Guidance:** Underlining that children want guidance on the risks and how to stay safe, the safety features, how long to use the metaverse headset, and how data is being extracted.

The recommendations shared and discussed with participants varied depending on the focus of each roundtable. A total of 13 participants took part in the virtual roundtables, with 4 taking part in the education roundtable and 9 taking part in the industry roundtable. In the first session, participants ranged from schoolteachers to educational charity representatives, and to national educational safeguarding advisors. Participants of the second roundtable came from different backgrounds, mostly industry professionals, but also representatives from the police, technology sector innovators, people working for relevant non-governmental organisations and in academia.

As the first session engaged educational professionals, the main recommendations were:

- I. **Keeping Children Safe in Education (KCSIE)**: Statutory guidance provided needs to include more reference to developing technology involving Virtual Reality.
- II. **Training**: At a minimum, all Designated Safeguarding Leads and Online Safety Leads should be trained on advancements in technology used by children.
- III. **PSHE**: These lessons could be expanded to include topics relating to online behaviour such as behaviour towards others, citizenship, well-being, reporting and behaviour modelling.
- IV. **Assemblies & Safer Internet Week**: Whole school approaches to guidance on online safety as an opportunity to increase sticky learning from PSHE are needed to be included on these occasions.
- V. **Engaging with Parents**: An opportunity to educate parents to make them aware of what metaverses are and existing games that their children are using, and to advise on communication, safety features and safeguarding tips.

The second roundtable had a wide background of participants. The recommendations, focusing on how industry could improve child safety in the metaverse, were broken into three main areas:

- **I. Empowerment**: Allow for emergency exit for users; create user friendly reporting functions; increase visual messaging regarding behaviour, reporting and victimisation.
- **II. Reporting**: Implement clear adaptive, optionally anonymous, and immediate reporting functions; display explanations of the reporting process; provide provision of additional support and feedback on reporting actions taken.
- **III. Harm mitigation:** Regularly assess risks on mental health and victimisation, as well as offender networks' development, and different forms of tech-enabled abuse. Implement safety solutions for ongoing abuse.

KEY POINT ARISING FROM THE ROUNDTABLES

Safety Awareness amongst children

Digital Literacy

Behaviour

Psychological Impact on children

Neurodiversity

In response to the shared findings from the VIRRAC Project, participants identified several key themes. These were: Safety Awareness, Physical Impact, Psychological Impact, and Neurodiversity.

1. Safety Awareness amongst children

In the original research, the children voiced their understanding and awareness of online safety features, how to report other users and knowledge of some of the risks posed to them. Although educational experts in the roundtable queried whether children would act on that knowledge in the moment, they expressed a degree of relief in hearing the level of children's awareness regarding online safety, as this was an issue that is viewed as a challenge for schools. They supported the request from children for more boundaries, but expressed difficulties regarding the ability to respect them or have them enforced outside of school settings. The prevailing sentiment among participants was that addressing this complex issue is often, and wrongly, perceived as the sole responsibility of educational institutions. Many expressed concerns that schools and other educational settings are expected to develop and implement safeguarding measures without sufficient external support or guidance.

2. Challenges in digital literacy

Lack of digital literacy was viewed as a significant challenge by participants, especially, in regard to neurodiversity, socio-economic status, environment, parent and/or carer ability, volume of pastoral needs, and opportunities for learning. The need for continuous and up-to-date staff training was also emphasised, topped by the challenges experienced due to staff turnover/the significant volume of occupational movement:

"At the moment a lot of our online issues come from things like WhatsApp and communication channels that we all use. There are staff who don't use virtual reality and that kind of technology to the level that the young people are going to do. As a consequence, we're going to be talking about topics that we're not going to be able to relate to or understand straight off the top of our heads." Participant

They identified an often overlooked issue of the generational divide that exists in educational settings, not only between the students and the staff but also between newer, younger members of staff and those staff with more experience.

Participants also believed that the scale of the necessary implementation and response would pose a challenge in terms of educating children. Whilst it could be possible to deliver effective educational sessions to several hundred students through assemblies, the challenge would be if there was a requirement to run sessions with smaller groups of students or deal with individual cases. For example,

if children were to try out the headsets to achieve greater learning outcomes from online safety in the metaverse sessions, it would most likely be challenging with regards to human resources but also provide a logistical, as well as funding challenge for schools.

"The scale is the challenge: you can deliver via an assembly to a hall full of 300 students, but that's got limited impact for the kind of safeguarding you want to do. That would require children to try it out. Trying to do anything on a scale where you've got a school of so many students is really tricky." Participant

Concerns were also expressed regarding the digital literacy of parents and carers, a challenge that has been widely documented in existing literature (Haywood & Sembiante, 2023). The respondents highlighted the need for targeted educational initiatives to equip caregivers with the necessary knowledge and skills to navigate and mitigate online risks effectively. Furthermore, it was noted that in cities or towns with significant ethnic and cultural diversity, where English is often a second language, there can be challenges in effectively sharing information on safeguarding issues. Experts shared the view that it was important not to ignore parents with additional learning needs.

3. Behaviour

When exploring children's behaviour in virtual reality, themes around citizenship, risk-taking and responsible behaviour were discussed. According to the experts who participated in the research, children seem to be taking on the responsibility for their own safety when there needs to be more emphasis on the sites themselves taking responsibility for the moderation of their platforms and implementing safety-by-design tailored to the specific nature of the immersive environments. According to the research, the volume of online hate crimes being witnessed and, in part, committed by children and young people, made participants think about whether the very nature of the 3D environment could be integral to minimising the feeling of disinhibition and reduction of empathy (Corkum & Shead, 2023) even further.

"One of the reasons why someone might be mean, rude, and inappropriate towards another player is that they don't have the same sense of empathy that they might have if that person was standing in their physical presence." Participant

The concept of empathy has been explored in various studies, particularly in relation to its role in online interactions and moral decision-making. In their 2023 study, Corkum and Shead examine the

relationships between electronic communication, cognitive empathy, and antisocial behaviour on the internet. They found that compulsive internet use positively correlates with online moral disengagement, while cognitive empathy negatively correlates with moral disengagement online. These findings provide a better understanding of these dynamics in traditional online spaces, however, further research is needed to determine whether similar patterns emerge in immersive virtual environments, such as the metaverse.

4. Psychological Impact

Research suggests that the metaverse's immersive nature can lead to more intense experiences than previously experienced in 'traditional' 2D online spaces, which may lead, as a result, to lasting psychological impact (see for example Gorichanaz et al., 2023 and Kim & Kim, 2023). Practitioners involved in the research reinforced this finding, raising concerns about the potential psychological impact of virtual reality on children's mental health, particularly for those with pre-existing vulnerabilities, such as anxiety disorders, autism spectrum conditions, or a history of trauma. They highlighted the potential for increased emotional distress, difficulties in distinguishing virtual experiences from reality, and heightened susceptibility to online manipulation or cyberbullying. Furthermore, the immersive nature of virtual reality may exacerbate symptoms of social isolation or sensory overload, particularly for neurodivergent children. These concerns highlights the need for further research into the psychological effects of virtual reality on vulnerable populations.

5. Neurodiversity

From the research, it emerged that children with communication impairments use the metaverse and the importance of paying attention to children with cognitive disorders, such as Autism Spectrum Disorder (ASD), was repeatedly raised. Neurodiversity was a topic discussed in great depth by the participants as they viewed it as a challenge regarding communication, the ability to monitor, parent and or carer additional learning needs and poor risk awareness. It was argued that a child's ability to communicate effectively is a common error overlooked by safeguarding policies, protocols and responses to online harms. The triad of impairments intrinsic to autism includes difficulties in socialisation, making eye contact and communication (Bushell et al., 2017). In virtual worlds, these impairments can be mitigated by the very nature of their design, for example, in many instances children who struggle to make eye contact offline do not have to make eye contact online. Verbal communication is also not a mandatory requirement of online gaming in the metaverse, and many nonverbal children with significant communication difficulties might find it attractive to access virtual worlds for social experiences (Parsons & Cobb, 2016). The immersive online world can provide empowerment to children with special needs such as playing alongside other gamers without having to experience the discomfort of verbal challenges or the sensations of rejection. Whilst part of these experiences can be tremendously

positive, they can also be used by bad actors to manipulate and exploit (<u>Schmidt & Newbutt, 2021</u>). As a participant pointed out:

"Socially, interacting with other people, strangers, those of ill intent, children on the autism spectrum can be more at risk because they haven't got those social interaction practices. In everyday life, they might avoid it, but then they're suddenly exposed to this massive world where they can socialise with everybody, and they haven't rehearsed that previously." Participant

6. Legislation and policing

The discussion included varying opinions around the current legislative measures such as the Online Safety Act, some believing that they do not adequately address the immersive and impactful nature of metaverse environments, making it difficult for example to mirror traditional sexual offences laws in these new contexts (<u>UK Parliamentary Post, 2024</u>). The Online Safety Act was also considered an important factor in the development of more robust methods of content moderation. This pressing need for proactive moderation and policing by platforms and better cross-platform safety measures to prevent grooming and other harmful activities was widely explored as well as the need for law enforcement agencies to be better involved and informed on crime trends in the metaverse.

"There needs to be ownership taken by the technology platforms to disrupt [harmful behaviour] beforehand, and the Online Safety Act is supposed to be the vehicle for tackling this issue". Participant

RECOMMENDATIONS

The findings suggest a pressing need for revisiting and potentially revising current policies and practices to ensure they adequately cover the unique risks posed by metaverse environments. The main concern voiced by the participants regarding online child safeguarding in the metaverse lies in the implementation of the recommendations made by the VIRRAC project, specifically relating to the ones developed for practitioners. Exploring different ways of effective implementation of the original research's findings, the experts highlighted the role and potential of existing policies, content moderation and safety-by-design, training, student numbers, age parameters, timeframes and funding, as factors that can make it challenging to respond to existing and emerging online risks and harms in the

metaverse. Based on their insights, key recommendations and policy implications have been identified to enhance the safety of children and young people in immersive virtual reality.

"We need any information we can get from any source. We use it to help our internal policies because safeguarding is so important. I don't know many teachers that pay lip service to it, we do take it seriously." Participant

1. Review the applicability of current online safety legislation

While the Online Safety Act sets out to cover harms and offences in all digital spaces, there is debate about its effectiveness in addressing these issues (<u>UK Parliamentary Post, 2024</u>) and whether there's a need to review it 'from the ground up' from a new perspective. The legislation should expand its focus beyond content to include contact harms, such as cyberbullying and online harassment, which are prevalent in the metaverse. Participants in the research agreed on the vital need for widespread awareness about existing legislation regarding child safety in the metaverse as well as the clear identification of potential gaps. They agreed that it would be essential to conduct a thorough review of existing safety features implemented by metaverse platforms which should include identifying best practices, particularly beneficial for small and medium-sized enterprises (SMEs) and start-ups. Such knowledge would inform future policy, industry guidance and codes of practice introduced following the Online Safety Act, besides supporting the enforcement of the existing ones.

2.1. Enhance content moderation efforts and training

Although existing legislation mandates that platforms implement a safety-by-design approach and moderate content accessible to children, concerns persist regarding the efficacy of these measures. A widely discussed issue, echoed in the mission of the eSafety Commissioner (2023), is the increasing use of encrypted messaging platforms by offenders to transition conversations with children from virtual reality spaces to more private and less regulated environments. Given this evolving threat landscape (Bleakley et al., 2024), experts have emphasised the need for technology platforms to take greater responsibility in preventing and disrupting harmful behaviours (Meggyesfalvi, 2021), rather than relying solely on regulatory frameworks to enforce safety measures. One of the participants attributed full responsibility to the technology industry, asserting that it bears the primary obligation for addressing and mitigating the associated risks:

"I feel like it's the Internet industry, they have the power to really sort this out."

Participant

It was also argued that no moderation system can be effective if human moderators are not trained to understand and identify online harms. Without training, important nuances of human behaviour could be missed, and training staff would minimise such a risk.

2.2. Al-driven content moderation

The role of Al-driven content moderation in safeguarding children in virtual environments will be expanding (Odudu, 2024). It is important to note that while Al tools are essential in detecting harmful content, their effectiveness remains highly variable, requiring structured evaluation frameworks to assess their reliability and ethical implications (Shneiderman, 2020). Utilising proactive Al moderation tools can be incredibly useful, however, the need remains for human oversight in sophisticated problems and dilemmas. In the report 'REPHRAIN: Towards a Framework for Evaluating CSAM Prevention and Detection Tools in the Context of End-to-end encryption Environments: a Case Study' experts elaborated on the critical challenges in Al-based moderation, particularly in false positives, contextual misinterpretation, and the circumvention of detection systems by offenders. Al moderation systems, though increasingly sophisticated, struggle to identify nuanced risks, such as grooming tactics disguised through coded language or psychological manipulation within immersive virtual environments (Wood, 2024). The REPHRAIN report emphasises that 'given the tensions that arise between protecting vulnerable users like children and safeguarding privacy and security at large' on online platforms, it is essential to have a set of clear criteria when evaluating detection and prevention tools, criteria that would go beyond classification accuracy, false positive rates, and usability.

The findings from the discussions with practitioners echo these concerns, highlighting the need for metaverse platforms to adopt transparent AI auditing mechanisms. Having systematic evaluation of AI moderation tools is critical to ensure that content moderation strategies are both effective and fair. AI alone cannot replace human oversight. Instead, hybrid approaches combining AI detection with trained human moderators should be prioritised. Robust evaluation frameworks could minimise the risks of AI moderation over-policing children's online interactions, e.g. removing legitimate content, as well as the failure to detect more sophisticated, context-dependent harms. Future research could focus on developing adaptive AI models, capable of learning from real-time moderation failures, while also integrating insights from criminology, child psychology, and digital safety research.

3. Educator's resources

Both focus groups' experts identified education as the main focus area where online harms in the metaverse could be primarily tackled. When asked about what resources educators need to effectively be able to protect children in the metaverse and what would these look like, participants emphasised that educational resources designed to help those working with children who are both typically and atypically developed should include some form of *situational learning*, *training*, *improvement to policy*

and improved means of communication with parents. This was viewed as being particularly important for children with disabilities, such as Autism Spectrum Disorder as they are more reliant on following a specific order of steps in order to process and embed that learning and the actions they need to take (Nemeth et al., 2010). Starting online safety education on risks and harms in virtual reality was seen as needed from as early as primary schools, as it could assist with developing online positive behaviours at a younger age. As an expert highlighted:

"We can see from the stats of the ages of the children in the focus groups that if we're leaving this until secondary school, we've missed the window really for setting off with positive online behaviours." Participant

When it comes to developing resources for staff training, participants suggested that such topics as potential risks and harms in the metaverse, and mitigation strategies could be delivered as a part of the mandatory safeguarding training sessions that all staff are required to do every two years.

4. Safeguarding standards and evidence-based, general information

Schools need to receive more evidence-based information, where children have been a part of the consultation process, on the risks and existing harms that prevail within metaverses, including those focusing on the tracking and/or identifying processes involved in child grooming behaviours in the metaverse. More research findings needs to be shared with schools to increase awareness and enable effective safeguarding in schools. Participants stressed that it was important for practitioners to have a working knowledge of what children are seeing and using so that they could provide tangible guidance for safety. This information needs to be included in national protocols, such as the Keeping Children Safe in Education statutory guidance resource, to ensure that the information and guidance is consistent and clear. Receiving up-to-date information regularly, as well as ongoing training opportunities, would provide staff with the opportunity to ask questions and discuss any emerging concerns. This training must also adopt a situational learning approach, whereby staff get to experience virtual reality and become conversant in safeguarding, specifically reporting, features. It would also provide the opportunity for schools to share online safety guidance with parents and carers to encourage them to not only apply safety features but to also feel confident in communicating with their children about online risks and negative experiences.

"What we need to be doing is bringing the technology into schools and almost having a programme set up where they are exposed to a situation in a safe environment."

Participant

5. Parental involvement

To achieve success in addressing child safety in the metaverse, especially knowledge gaps, there is a need for greater working in partnerships, not just with experts from external organisations, but with parents. Parents need to receive support to have the ability to understand and safeguard their children in an immersive, multi-dimensional online environment, including gaining an understanding of what are the differences in terms of risks and harms from those encountered in 'traditional' 2D online spaces as these are not limited to the boundaries of school hours. Research into the importance of parent and carer engagement in their child's online safety have indicated that there is a direct correlation between a limited understanding of digital safety, and lack of engagement, with increased risk (Murray, 2018; Virag & Parti, 2011). It has also been identified that cognitive and linguistic ability, level of education, and socioeconomic standing also play a significant role in parents actively engaging with professional support. This was something emphasised by the participants who stressed that it is important not to assume that all parents have the same digital literacy level. As one expert stated:

"There are parents with English as an additional language and parents with significant learning needs themselves so we cannot come with a presumption that parents are going to understand a certain level of technological language." Participant

In such instances, it is important that resources and guidance given to parents is comprehensive and takes factors, such as diversity, into strong consideration to ensure that information is not only accessible but also understandable. Safeguarding established in schools could be informative to and supportive of parents/carers outside of school to increase the opportunity for schools/safeguarding practitioners become aware of emerging issues earlier and can take adequate steps to protect children.

CONCLUDING REMARKS & NEXT STEPS

The three primary areas identified within this project that would be helpful to focus on to maximise the effective dissemination of the VIRRAC outputs (including the <u>Toolkit report</u> and the <u>Short Videos for Children</u>), to raise awareness of the research's findings and to make a positive impact on child safety in the metaverse could be through:

- 1. Developing classroom resources;
- 2. Addressing the availability constraints; and
- 3. Engaging with experienced individuals.

Developing suitable classroom resources should include the provision of guidebooks, the provision of elearning resources, and classroom resources such as lesson plans and assembly topics. As previously mentioned, the Keeping Children Safe in Education (KCSIE) statutory guidance could be an important tool by which dissemination could be included, especially as it would ensure that the messages regarding safety in the Metaverse are consistent and clear. Utilising PSHE lessons to focus on specific topics regarding online safety, with focus towards the schools' most vulnerable children, provides another excellent opportunity, as staff, especially in secondary schools, develop rapport with their students there over the years. This can facilitate opportunities to discuss sensitive topics and disclosure of experiences that are concerning the children in an environment in which they feel safe and heard (John, 2017). These lessons can be developed to incorporate opportunities for behaviour modelling, and as a means of structuring an approach of sequence learning to support those that are neurodivergent.

During the second phase of the VIRRAC research, 'time' was a theme frequently mentioned. Concerns were voiced that delays in developing adequate educational resources and delivering training sessions on metaverse safety to both professionals and children could mean that the information these contain would be outdated by the time of implementation. On the flip side, participants considered that by rushing these processes, there was a risk that actions would be designed and implemented without proper attention. As expressed by a respondent:

"You can run the risk of it being done kind of half-heartedly just so it gets fitted in, so we can tick the box to say that we've done it." Participant

Finally, in order to enhance child safety in the metaverse any discussion and resource or policy development must include knowledgeable subject experts. However, the issue here might once again

come down to funding. The greatest impact in safeguarding can come from utilising and engaging with experienced professionals, who follow trends and actively respond to emerging threats in real-time. Such experts understand how these risks apply to the everyday lives of children and the challenges faced by practitioners. Especially, in regard to the advancements of technology to incorporate the ability to touch and be touched in a virtual world though the use of haptics.

Another reason to use subject experts is the novelty their experience holds for the children. Children at schools become familiar with teaching staff, and they might not always be willing to take on board messages for specific topics such as cybercrime in virtual reality. During the research, experts voiced the belief that it would be an effective way of engaging with children to involve external topic experts who come in to engage with the students for occasions dedicated to discussing online safety and harms in the metaverse.

"The children really sit up and take notice when they have a visiting speaker who's able to relate to them. They come in cooler clothing than we're wearing with more interesting tales to tell. They're immersed in that world and the kids instantly connect with them." Participant

This VIRRAC Report emphasised the urgency of awareness raising and developing some safeguarding measures, with a focus on educational practise. The report identified meaningful gaps in digital literacy, the psychological impacts of immersive experiences, and the unique challenges faced by neurodivergent children. It highlighted the potential limitations of current legislation, such as the Online Safety Act, in addressing these novel risks.

To mitigate these risks, the report suggests revising online safety legislation, improving educator resources, answering the need for greater cooperation between stakeholders, for example via increased parental/carer involvement, and enhancing content moderation efforts. When applying safety by design solutions, it is also essential to include technical, legal, operational, and/or contractual safeguards to prevent the re-purposing of technologies such as AI driven detection, prevention and moderation tools.

The report also stressed the importance of continuous research and policy development to adapt to the rapidly evolving metaverse worlds with regard to child safety.

"We know that it takes an online village to protect a child and we've got to be able to get everything we can to protect children online, - that is the way to go." Participant