Risks the Metaverse Poses for Children and Adolescents: An Exploratory Content Analysis

Author: Michael Hinz
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT,
The technological concept of the metaverse has the potential to change our society to an extent that can be compared with past inventions like the internet. While it encompasses the potential to bring extraordinary advances in areas such as health care, education, or e-commerce the risks and challenges it might bring for its users are often neglected. Past experiences regarding social media or online gaming showed how such technologies can have negative effects on the psychological and physiological well-being of their users and to what degree malicious individuals might abuse them. A user segment that is especially vulnerable in this context is children and adolescents. This paper addresses the types of psychological and physiological risks the usage of the metaverse can pose for children and adolescents by creating a knowledge base derived from the current state of literature and interviews with experts. Additionally, it contributes to research by developing a framework that helps to identify and categorize child-inappropriate content or behavior that occurs in metaverse spaces. These findings may facilitate collaborations between researchers, policymakers, and developers of this technology to ensure the well-being of users and to create innovative solutions that mitigate the identified risks and challenges, fostering a secure digital environment that is safe for all user segments.

Graduation Committee members:
Dr. Robin Effing
Dr. Matthias De Visser

Keywords
Metaverse, risk, harassment, inappropriate, child, adolescent, identification, categorization

1. INTRODUCTION

The emergence of the concept metaverse in recent years, a virtual world existing solely in the digital space, created unprecedented opportunities as well as challenges for organizations involved in its development and also for society as a collective. This development is greatly facilitated by revolutionary technologies such as “virtual reality (VR) which has been aided by parallel advancements in AI, the IoT, Clouds of Things, Big Data, and other technologies” (Allam et al., 2022). These virtual spaces can provide a multitude of functions ranging from simple aspects such as gaming or communication up to the creation of effective workspaces or digital-twin models of real-life scenarios (e.g., city planning or natural catastrophe prevention) (Allam et al., 2022). This range alone shows how diverse and universally applicable the concept of the metaverse will become in the upcoming years which indicates the importance for not only special interest groups but also for all members of society. The general concept of the metaverse began with Neal Stephenson's novel "Snow Crash" in which the author depicted a dystopian reality of our world in which people entered a virtual space via a head-mounted display (HMD) where they were able to construct their own personalized world according to their preferences (Abbate et al., 2022; Stephenson, N., 1992). Inspired by Neal Stephenson’s visionary concept, the metaverse is gradually taking shape with technological advancements like "Second Life" or Meta Platform Inc.’s recent venture “Horizon Worlds” (*Horizon Worlds | Virtual Reality Worlds and Communities*, 2023; *Official Site | Second Life - Virtual Worlds, Virtual Reality, VR, Avatars, and Free 3D Chat*, 2023). These immersive spaces allow users to enter digital environments in which they have the possibility to interact via customizable digital representations of themselves (Avatars), bringing us closer to Stephenson’s vision. (Dwivedi et al., 2022). As depicted in Damar (2021) a possible and inclusive definition of the metaverse could be a “shared 3D virtual world in which all activities can take place using augmented and virtual reality equipment”.

With the metaverse’s growing popularity ethical concerns naturally were raised simultaneously (Dwivedi et al., 2022). Especially the fact that the development and regulations of certain metaverse spaces are not in the control of respective government agencies, but instead of big tech companies whose hidden interests could mainly follow financial or political motives, make the ethical accountability and transparency regarding the regulation of the metaverse highly questionable (Bibri & Allam, 2022). One user group that is particularly endangered by the influence, that the metaverse can have on society, is children and adolescents. The dangers created by increased utilization of the metaverse and related technologies could include for example the publication of private data (e.g. name, address, contact information), addiction, increased violent behavior, negative impacts on psychological well-being, cyberbullying, or exposure to unsafe content or behavior that is not appropriate for children and adolescents (Kaimara et al., 2022; Lavoie et al., 2021; Lee et al., 2021; Muslihati et al., 2023; Usmani et al., 2022). Due to the still unknown impact, the metaverse will have on our society, and the lack of regulations currently defining those virtual spaces, the subject of child and adolescent protection will represent an essential future research domain. Despite the ethical importance of user protection and the prevention of aspects such as harassment or the distribution of child-inappropriate content or behavior, the current state of research regarding the metaverse is predominantly focused on technical issues and the potential positive impacts this new technology might have on our society. Aspects like the underlying technological processes, future business and monetarization possibilities, improvements regarding city planning and urbanization, or advantages for health care and education currently represent the core of literature (Abbate et al., 2022; Allam et al., 2022; Bailey & Bailenson, 2017; Bibri & Allam, 2022; Damar, 2021; Park & Kim, 2022). Only a fraction of research examines what negative effects a further implementation of the metaverse in our daily life might encompass on people's psychological and physiological health, especially with a lacking focus on minors. Considering the novel nature of this technology, it is obvious that long-term studies regarding impacts on our well-being still do not exist. Due to this clear gap in research, the aim of this paper is the provision of an initial knowledge base regarding what psychological and physiological risks exist regarding entering these virtual worlds. Furthermore, this guides the development of a conceptual framework used in answering this paper's research question: **“What types of child-inappropriate content or behavior can be identified in social or gaming spaces of the metaverse that children and adolescents might encounter during their play?”**

2. CURRENT STATE OF RESEARCH REGARDING RISKS INDUCED BY VIRTUAL ENVIRONMENTS

In order to map out the current state of research an initial literature review was conducted. The following section will provide an overview of how the current state of research examined the psychological and physiological risks the usage of the metaverse could pose for children and adolescents. The literature review took place on the interdisciplinary research databases “Scopus” and “Web of Science” via an initial keyword search of relevant concepts based on a Boolean search strategy. This was then combined with a snowballing technique in which citation tracking was utilized to further discover relevant papers. Due to the lack of research, most of the comparative studies are focused on either gaming, social media, or VR technology. A comparison that is logical because the concept of the metaverse can be seen as the next developmental step in which technologies such as gaming and social media will take place (Benrimoh et al., 2022; Lee et al., 2021; Muslihati et al., 2023; Usmani et al., 2022). This paper will follow this approach and takes research regarding relevant psychological and physiological risks of social media and gaming into account that are comparable with aspects of the metaverse. Furthermore, due to the bioethics of human research, it is a complex process to carry out research experiments with children since an extensive amount of ethical factors need to be taken into account for it (Kaimara et al., 2022; Neill, 2005). Therefore, there are papers included in the literature review that focus on young adults instead of minors which should be considered when evaluating their results. Additionally, semi-structured interviews were conducted to receive an even deeper insight into this complex research domain. As interview partners professionals participated who possess extensive expertise in either the field of VR/metaverse technology, medicine, or psychology. These findings will also be presented at the end of this section.

2.1 Literature review

2.1.1 Psychosocial and cognitive risks

2.1.1.1 Risks through addiction

The most prominent risk identified in research is the development of an addiction to these virtual spaces. This

prevalence was shown by a survey among internet users that determined that addiction to the metaverse currently represents the biggest threat of said concept to society (*Dangers of the Metaverse 2021*, 2022).

"Addiction is categorized as a chronic relapsing disorder, depicted by compulsion to seek and use either a substance or a behavior, like gambling. In addition, it includes loss of control in limiting certain behaviors or drug intake, and mostly is associated with the emergence of negative emotions (e.g., anxiety, irritability, or dysphoria) in situations where the drug or behavior is not attainable." (Korte, 2020). Zamani et al. (2009) conducted a study among students that showed a significant relationship between the degree of gaming addiction and mental health issues such as anxiety or depression. Similar findings were presented by Andreassen et al. (2017) who highlighted the relationship between pathological anxiety and depression and the excessive duration of gaming. Furthermore, another relevant finding was here that younger people are more at risk of developing such addictions since recent generations are exposed to technology from their early childhood on which increases the frequency of contact as well as the potential of becoming addicted. It is worth mentioning that the study resulted in a significant negative relationship between gaming addiction and the degree of social dysfunction the students experience. Other symptoms regarding the user's mental health described in literature are namely loneliness, social incompetence, mood swings, and a decreased level of confidence and self-esteem (Labana et al., 2020.; Lavoie et al., 2021; Muslihati et al., 2023; Paulus et al., 2018; Van Rooij et al., 2014). Additionally, adolescents that showed signs of an internet gaming disorder were significantly more likely to engage in substance abuse such as alcohol, cigarettes, or marihuana (Van Rooij et al., 2014) which aligns with the discovery that gaming addiction and substance use disorder share the same neurobiological mechanisms in the human brain (Ko et al., 2009). Furthermore, reduced academic performance could be observed due to a clear tendency of adolescents to sacrifice study time for gaming (Choo et al., 2010; Griffiths et al., 2004).

Similar effects are likely to occur due to an addiction to social media networks. This represents the other large component planned out by the metaverse with which children and adolescents are predicted to engage the most besides gaming. Mental health, being a critical factor for both types of addiction, was found to be also suffering through excessive use of social media. Hou et al. (2019) carried out a study that proved a significant negative relationship between social media use among students and their respective mental health as well as academic performance. The aspect of self-esteem was here empirically proven to have a mediating effect on the relationship which is in line with prior research indicating that social media addiction is negatively associated with the self-esteem of users (Andreassen et al., 2017; Błachnio et al., 2016). Other negative aspects that can occur through excessive use of social or other digital media are a decrease in text comprehension, feeling disconnected from peers outside of social media, as well as reduced language skills (Korte, 2020). The latter becomes evident when one examines the impact this kind of consumption of digital media can have on the human brain, especially if a child is exposed to long durations of digital media in his early years. Studies have shown that extensive use of digital media can have implications for the development of the microstructural integrity of white-matter tracks in the brain which is correlated with the language and reading capabilities of humans. In case the fiber tracks of a developing brain will not be developed to the full extent the authors expect that the development of language skills could be inhibited (Hutton et al., 2020; Korte, 2020). The brain of adolescents is especially vulnerable to certain risks regarding the usage of social media since during puberty, brain areas involved in emotional and social aspects undergo substantial changes (Korte, 2020). This was shown by Kanai et al. (2012), Meshi et al. (2015), and Pfeifer & Blakemore (2012) who identified that the usage of social media correlates with activities in the gray matter volume of the amygdala which is an area in the brain that is responsible for processing emotions.

Concerning these different types of addiction, it is essential to mention that it is unclear if these negative aspects connected to the mental health of addicts can be seen as pre-conditions or effects. It is possible that certain character traits or environmental conditions need to be present so that addiction will be developed or that these strains on the person's mental health are developing through addictive behavior. It can be hypothesized that a vicious cycle often exists between reasons, addiction, and effects which is not easy to break since the correlation might be bidirectional (Hou et al., 2019; Van Rooij et al., 2014; Zamani et al., 2009).

There are certain concepts that the metaverse and the closely related VR technology incorporate to a greater extent than traditional comparable technologies (such as desktop computers or smartphones), which are capable of increasingly fostering the occurrence of addictive behavior or other health risks that will be mentioned in the following sections. These are namely the factors of immersion and its closely connected concepts of presence, absorption, and embodiment (Cairns et al., 2014; Han et al., 2022; Lavoie et al., 2021; Shin, 2018). Immersion defines the extent to which users of a certain technology feel present in a virtual environment which subsequently increases the chance of blurring the borders between virtual reality and real life (Dwivedi et al., 2022; Goltz, 2011; Han et al., 2022). Why this factor can have a higher impact on children and adolescents playing violent video games becomes evident when taking the super-realistic properties into account that the metaverse might bring (Park & Kim, 2022). Lee et al. (2021) argue that this might lead to users attempting something they experienced during their play in real life, possibly even immoral behavior. Given the fact that younger people prefer violent aspects of games more than adults do and that additionally, studies found significant relationships between video game violence and real-life aggression one can argue that exposure of minors to virtual reality violence (amplified by a high level of immersion) creates a concerning risk for society (Anderson et al., 2010; Griffiths et al., 2004). This compelling and realistic sensory experience for users can be created through technological and design elements that can lead to immersive visual, auditory, and haptic stimuli as well as realistic interaction mechanisms (Lee et al., 2021). In detail, immersion can be described as a "confluence of different psychological faculties such as attention, planning, and perception that when unified in a game lead to a focused state of mind" (Cairns et al., 2014). Furthermore, Cairns et al. (2014) mention that a person who is engaged in an immersive state might try to remain in this world, which originates from the self-sustaining characteristic of immersion and makes them hesitant to leave (Allam et al., 2022; Dwivedi et al., 2022; Kaimara et al., 2022). In case a sufficient degree of immersion is reached, the effect can occur that users feel present in a virtual environment even though they are not physically there (Han et al., 2022; Witmer & Singer, 1998). This factor is essential for software and game developers to take into account since a great level of immersion and presence can lead to a stronger engagement and emotional connectedness to the virtual world (Lavoie et al., 2021). Absorption can be seen as closely related to immersion as both are defined by an increased mental focus and emotional investment (Lavoie et al., 2021, Shin, 2018; Teng, 2010). For the aspect of absorption, a study was conducted by Lavoie et al. (2021) that indicated that the

degree of absorption has a mediating effect on the relationship between VR experiences and negative emotions. The higher the perceived absorption was for the participants the stronger the negative emotional effects were. The last essential concept relevant in terms of immersive and addictive potential is embodiment. Embodiment describes to what extent a user of the metaverse, or other comparable virtual worlds relates to his digital avatar, embodies his characteristics, and what kind of emotional connection he forms with his virtual self. A high degree of embodiment has a positive effect on the overall sense of immersion and presence the user feels (Shin, 2018). The aspects of creating virtual representations of oneself, embodying them in the metaverse, and what risks are connected to them will be more closely evaluated in the following section.

2.1.1.2 Risks through embodying avatars

An avatar can be defined as the digital representation of an individual in the metaverse (Lee et al., 2021). As the user creates this avatar, he has the possibility to freely customize certain attributes such as the virtual appearance he wishes to embody. This ability to customize one's physical attributes in seconds and to embody a character that potentially mirrors one's profound wishes for one's appearance in real life creates a highly addictive potential while fostering the blurring of borders between virtual reality and real life (Goltz, 2011). "As the avatar grows, the user's intimacy increases, becoming more immersed in the metaverse. However, addiction (e.g., internet, video games) and excessive immersion result in confusion and lack of interest in the incongruity with the real world." (Dwivedi et al., 2022). If this addiction to the avatar becomes stronger and the user might start preferring this representation over being themselves, negligence of one's own body and immediate family, friends, and other aspects of the real world might occur (Ashour et al., 2018). Young people might be increasingly at risk here due to their tendency to favor products that fulfill their role projections and fantasy (Holsapple & Wu, 2007).

Another factor that needs to be taken into account when examining the embodiment of avatars is what implications this might have on the individual's dynamics of behavior (Lee et al., 2021). Multiple research papers mentioned the Proteus effect in connection to how the embodiment of virtual avatars can influence certain characteristics or behavioral traits in real life (Fox et al., 2013; Lee et al., 2021; Paul et al., 2022; Usmani et al., 2022; Yee et al., 2009). Fox et al. (2013) define the Proteus effect as a state of "when a user's self-representation is modified in a meaningful way that is often dissimilar to the physical self. The user then embodies the self-representation, observes him or herself behaving in this virtual form, and draws inferences regarding his or her internal beliefs or attitudes based on these observations. After embodiment occurs, the user's behavior then conforms to the modified self-representation regardless of the true physical self". The results of this study were an increase in body-related thoughts, rape-myth acceptance, and body dysmorphia among young women that embodied sexualized avatars. This aligns with Usmani et al. (2022), who mention a potential connection between the Proteus effect and the development of body dysmorphia or other related mental illnesses such as anorexia.

2.1.1.3 Risks through other users

The following section will give a closer look at how other users and their actions might pose risks for children and adolescents entering virtual worlds in the metaverse. One factor that already played a relevant role in social media and gaming environments and one that has to be prevented to ensure the safety of all user segments is (sexual) harassment (Blackwell et al., 2019; Falchuk et al., 2018; O'Keeffe et al., 2011).

While different types of online harassment and their effects are already identified in research, experts indicated that these occurrences simultaneously take place in online virtual reality as well (Blackwell et al., 2018; Chen et al., 2020; O'Keeffe et al., 2011; Shriram & Schwartz, 2017). Blackwell et al. (2019) conducted a study that shows that harassment, coupled with the immersive VR technology associated with the metaverse, can increase the psychological impact on the user's mental health compared to traditional online spaces. The paper divides the concept of harassment into three layers. The first is described as verbal harassment and includes aspects such as personal insults, hate speech, or sexualized language. The second layer is called physical harassment and refers to unwanted touching, standing too close to the user, obstructing the user's movement, or performing visible sexual gestures. Lastly, environmental harassment was mentioned as the third layer which consists of displaying sexual/violent content, drawing sexual images, or throwing objects at other users. Participants of the by Blackwell et al. (2019) conducted interviews indicated that the psychological impact they felt after the harassment took place in virtual reality was greater compared to other online mediums, such as social media. In the case of verbal harassment, a participant reasoned this through the existence of real-time voice chat in virtual spaces since one needs to actively participate in the situation instead of reading messages or posts on social media. Regarding physical harassment, the aspects of immersion, presence, and embodiment increased the feeling of realness of the situation. Since VR creates the feeling of being physically present in this space, physical altercations or harassment can induce a similar level of realness as in real life (Han et al., 2022). On the other hand, some participants indicated that such harassing behavior did not affect them to a greater extent than in other online mediums e.g., due to extensive experience of similar situations in online gaming environments that gives them a clear distinction between the virtual world and real life.

Besides other segments of users (females, people of color, or people with accents) children were specifically mentioned by participants to be one of the main targets. The metaverse can provide obvious cues regarding one's age, gender, or other characteristic due to the implemented voice chat and the user's height captured by the HMD and then expressed via the avatar. This can give hints to other users about who may be a child which creates a target for potentially vicious individuals. Jonsson et al. (2019) identified poor self-esteem and a poor relationship with parents as risk factors that lead adolescents to search for validation in online spaces and potentially become victims of sexual predators. Poor mental health was identified to be either a cause or effect of these situations. Nevertheless, Usmani et al. (2022) refer to Bragesjö et al. (2020) and Jonsson et al. (2019) while stating that "negative, violent, or abusive user experiences in the virtual world can incite similar psychological and physiological responses in an individual's real world. Such incidents can have long-term detrimental psychological effects on the victim's mental state and increase the risk for mental illness such as depression, anxiety, PTSD, or insomnia.". The risk of grooming children and adolescents in the metaverse by sexual predators describes only partially what kind of threats other users could impose on them. Research conducted by the CCDH (Center for countering digital hate) provides evidence regarding numerous child inappropriate content or behavior in one of the most popular metaverse social spaces, namely "VRChat" (Lawson, 2021). Besides the already indicated risks of harassment and exposure to explicit content, the study mentioned grooming of children to repeat racist slurs and

extremist talking points. This aligns with an interview published by the news agency “Reuters” in which they questioned Madan Oberoi, Interpol’s executive director for technology and innovation, about what kind of cybercrime could be facilitated by the metaverse. One threat that is concerning Interpol is that terrorist groups or other extremist parties could use the metaverse to groom children for their cause while additionally creating virtual worlds that can be used for planning and providing direct training opportunities (Kartit et al., 2022; Koehler et al., 2023)

2.1.2 Physiological risks

2.1.2.1 Risks through addiction

In terms of physiological threats issued by an addiction to the metaverse, various triggers or effects were found in the literature. It is argued that the association between gaming addiction and obesity is complex and indirect and that a multitude of factors are relevant to explain the different mediating processes underlying the development of obesity in children. Habits that are closely connected to excessive gaming, such as the consumption of unhealthy high-calorie drinks and food or poor sleeping patterns, were found to be correlated to the development of obesity (Kaimara et al., 2022; Kenney & Gortmaker, 2017; Turel et al., 2017).

Furthermore, Turel et al. (2016) provided a study that linked an addiction of children to information systems to the potential development of cardio-metabolic disturbances. As mentioned, bizarre sleeping patterns and insomnia are common themes related to the higher use of digital media due to their emission of short wavelength light which subsequently has an altering impact on the human's production of the sleep hormone melatonin, an effect that is even increased for children (Gottschalk, 2019; Kenney & Gortmaker, 2017). Since most of the studies regarding pre-conditions or effects of addiction are cross-sectional, reverse causation of the mentioned effects is possible.

2.1.2.2 Cybersickness

The effect of cybersickness (or simulator sickness) describes an effect that is commonly related to virtual reality technology and the usage of HMDs. Symptoms include nausea, vomiting, dizziness, and disorientation which is potentially induced by a conflict between visual stimuli and the vestibular system. While the visual influences during the VR experience indicate movement, there is none registered by the vestibular system, which subsequently leads to the mentioned effects. This conflict of information is explained by the sensory conflict theory which is also related to other motion sicknesses such as car sickness or sea sickness (Kaimara et al., 2022; Nichols & Patel, 2002; Regan, 1995). Besides other reasonings, the sensory conflict theory represents the most prominent explanation in literature (Cao et al., 2020; Kaimara et al., 2022; LaViola, 2000). It is worth noting that the effects, whose severity is connected to the duration of play, are temporary, and that children experience them to a lesser degree or after longer periods of exposure than adults (Cao et al., 2020; Kaimara et al., 2022; Nichols & Patel, 2002).

2.1.2.3 Physical injuries

How the usage of the metaverse could potentially lead to physical injuries originates from the related VR and AR technology e.g., the HMD and the handheld input devices. Facilitated by the restricted vision and level of immersion, accidents while playing in one’s own home often occur. (Park & Kim, 2022). In case users engage with the metaverse through an AR system instead of being fully immersed in VR additional risks occur if the user decides to use this technology outside of his home. Due to possible distractions or obstructing the vision of threatening objects/situations (e.g., traffic or obstacles) by digital overlays users might encounter life-threatening risks (Cloete et al., 2020). Children are especially at risk here due to their impulsive behavior and lesser experience and attention. An example from recent years was the popularity of the AR smartphone game “Pokemon GO" whose usage led to numerous injuries and even deaths due to players' lack of attention to their surroundings (Cao et al., 2020; *Pokémon GO Death Tracker*, 2016).

2.2 Preliminary interview results

The additionally conducted semi-structured interviews are of exploratory nature and shall provide an additional knowledge base regarding what kind of risks children and adolescents could encounter in the metaverse. Combined, the expertise of the questioned experts includes the areas of medicine, psychology, and VR-technology-related research. Relevant findings will be presented in the following section.

Table 1: Interview partners and their professions

Interviewee	Profession
Interviewee A	Medical researcher in the field of addiction, genetics, and neurology
Interviewee B	Lecturer with a focus on medicine and life sciences
Interviewee C	Researcher with a focus on the psychiatric evaluation and implementation of VR technology and eHealth
Interviewee D	Lecturer in health psychology and technology and researcher in VR implementations for education/research
Interviewee E	Assistant professor with a focus on medicine, life sciences, and social sciences
Interviewee F	Lecturer with a focus on (cyber) psychology, health, technology, and digital wellbeing

Interviewee A generally stated that the research regarding the metaverse is at risk of becoming an imbalanced field of study in which researchers focus too much on its potential advantages while neglecting the dangers associated with it. Due to this fact, the overarching research topic was evaluated as essential and valuable, especially due to the focus on children and adolescents as vulnerable participants in our society. It is worth noting that multiple experts indicated that it is difficult to generalize potential risks since the vulnerability of a child to the dangers of the metaverse is based on an interplay of numerous subjective factors. These are for example their relationship with friends/family, predisposition for psychological problems, or their individual level of experience and self-esteem. This is why a generalized framework applicable to all children and adolescents is difficult to develop. Furthermore, since no long-term studies regarding those risks exist yet, researchers can only speak from their experience and expertise from related study fields. Nevertheless, certain trends or probabilities of threats still exist.

Starting with the possibility of an addiction induced by the high level of immersion all questioned participants mentioned that this represents a realistic risk, especially for children and adolescents.

"The immersed feeling of presence in these virtual worlds is especially hard for children of a younger age to distinguish from reality. The risk of confusing real life with virtual reality becomes even stronger because our optical sense is the easiest to trick" indicated interviewee A. The creation of idealized avatars can foster this occurrence even more. The fact that the metaverse can provide children and adolescents with an additional representation of themselves which they potentially start preferring over their own body was mentioned by the participant as one of the greatest risks, especially for adolescents in the uncertain time of puberty. Follow-up effects that were mentioned in connection to the induced dissatisfaction with their own body were reduced self-esteem, body dysmorphia, or anorexia. Other addiction-induced effects that the experts indicated were depression, anxiety, loneliness, reduced social skills, reduced academic performance, and restlessness. Additionally, a decrease in attention skills was hypothesized by interviewee B as he indicated that "getting used to the high level of stimuli that are usually occurring in the metaverse could make it harder for children to focus on low-stimulating activities such as reading a book or following a lecture".

In terms of the increased effects of cyberbullying or other kinds of harassment, multiple experts stated that due to the higher level of immersion, children might encounter traumatic incidents that could create traumas that burden them for their whole life. Interviewee E, whose research domain includes VR-induced interventions for criminals, stated that the risk of grooming or recruitment through extremist organizations might be increasingly dangerous for adolescents in the metaverse. While adolescents in puberty tend to distance themselves from their parents to search for new peer groups, it is easy for such predators or groups to provide them with a false sense of belonging and masking their true nature and intentions through personally adjusted environments and avatars. Interviewee E indicated that "terror groups have the possibility to create a better picture of themselves by exactly tailoring their appearance in a way that is appealing to their targets which increases the chances of recruitment compared to how it would work offline".

As physical risks cybersickness and physical injuries during play were mentioned. While for cybersickness no higher threat was hypothesized for children, two experts indicated that the risk for physical injuries is significantly higher due to the impulsive and careless behavior often typical for children.

Interviewee A mentioned the fact that around 30% of the human's neurological connections restructure themselves during puberty and adapt to the current environmental conditions and circumstances. This might indicate that through extensive usage of the metaverse during these years, adolescents might develop characteristics specifically adjusted to the metaverse. This could range from deficits in facial recognition or social capabilities up to the satisfaction of human needs such as the search and interaction with potential sexual partners. Furthermore, interviewee A defined the time in which the neurological restructuring takes place as an "incredibly vulnerable phase for the brain in which traumas, that humans would normally process easily, could have immensely damaging effects".

When asked about suggestions they would have for metaverse developers as well as parents to mitigate the mentioned risks, the most prominent answer that was given here is that the focus should lay on sufficient education by parents or eventually schools about how children should behave correctly in the metaverse and how to deal with potential threats. Interviewee F stated, "You wouldn't leave your child unattended in the middle of New York City. Why then in the metaverse?". Regarding the risk of addiction, strict time restrictions and a balance with other activities (e.g., sports) were suggested. Besides co-experience and education, suggestions for developers were given as well. Here the implementation of strong governance systems was demanded that actively prevent malicious individuals to harm children in any way. Furthermore, the implementation of standardized and non-perfectionistic avatars was suggested to prevent children from preferring their virtual selves over their real bodies, while also the creation of specific child-appropriate worlds containing certain policies to protect them should ensure their safety.

2.3 Scientific framework

Following the previously highlighted issues a conceptual framework was developed to further evaluate what risks children and adolescents will face when entering the metaverse. Due to the potentially severe effects that an exposure to inappropriate content or behavior can have on minors the author decided to further investigate what types of occurrences exist in the metaverse that may cause harm for younger users.

Table 2: Overview of the conceptual framework

Research question: What types of child-inappropriate content or behavior can be identified in social or gaming spaces of the metaverse that children and adolescents might encounter during their play?	
Hypothesis	**Methodology**
Children and adolescents are at risk of encountering various types of inappropriate content or behavior while using social or gaming spaces in the metaverse	Identification and categorization of inappropriate content or behavior taking place in social or gaming spaces in the metaverse
Independent variable	**Dependent variable**
Entering social or gaming spaces in the metaverse	Encountering types of child-inappropriate content or behavior

This framework aims to guide the process of answering the research question while the respective hypothesis and independent and dependent variables provide a basis for what exact factors the chosen qualitative method is examining. While raising awareness of the potential risks children and adolescents might encounter inside the metaverse the author hopes to facilitate further research in this domain. This will also support the development and improvement of governance/safety systems that actively mitigate the identified risks.

3. METHOD

3.1 Procedure

The chosen method of this paper will provide a detailed content analysis of two different metaverse applications regarding what types of child-inappropriate content or behavior exist there. Due to the lack of research in this field an exploratory study needed to be conducted to create a knowledge basis and recommendations for future research. The sample of metaverse applications that were chosen are the social space "VRChat" and the gaming space "RecRoom". The selection of both social and gaming spaces shall be reasoned by the attempt to cover the two categories of metaverse applications with which children and

adolescents are most likely to get in contact with. Additionally, both applications were chosen under the criteria that they are free-to-play multiplayer spaces, factors that lead to higher user numbers, and subsequently to an increased amount of potentially inappropriate users and occurrences that may cause harm for minors. These spaces can be seen as predecessors to what the metaverse is aiming to be since most of its predicted and aspired functions are still in their infancy so it would be an exaggeration to present them as a viable mirror world to our real life (Allam et al., 2022).

The social media space "VRChat" especially stands out through its immense variety of worlds and the extensive customization possibilities users have for their own avatars. The latter ranges from pop culture characters to self-designed looks which gives the user the ability to embody every thinkable avatar they wish. In its virtual worlds, which can also be oriented towards existing real-life or virtual places as well as freely custom-build, users then have the possibility to engage in conversations via voice chat or share media with other participants (*VRChat*, 2023).

"RecRoom" on the other hand lays its focus on the creation of game modes that users can play among each other. The user's ability to customize their avatar is more limited compared to "VRChat". Besides the purchase of already-created clothing and the ability to create customized items, there are no other options to modify the predetermined avatar base. Examples of the numerous game modes available are fantasy roleplaying, paintball, battle royals, or sports such as basketball (*Rec Room*, n.d.).

The exact nature of the method is a practical experience performed by the author inside both chosen metaverse applications for a duration of three hours each, following a similar approach as the exploratory study conducted by the CCDH (Lawson, 2021). To ensure reliability the three hours were distributed over multiple times in a day (morning, noon, and night). The maximum duration for which the author entered the metaverse was limited to one hour per day. Also, the author did not communicate with other users in any way in order to prevent any diversion caused by e.g., gender or age. Combined, this approach averts the occurrence of errors and biases that might would've altered the reliability of the research. During these time frames, all inappropriate content or behavior which may cause physical or psychological harm to younger users was identified and categorized in a designated coding instrument which was synthesized through the respective codes of conduct of the applications (*Code of Conduct/RecRoom*, n.d.; *Community Guidelines/VRChat*, 2023) in combination with the different types of harassment described in Blackwell et al. (2019). The categorization instrument consists of three levels. The first overarching level is inspired by the research of Blackwell et al. (2019) and distinguishes each type of harassing/inappropriate content or behavior into verbal, physical, or environmental nature. This explains in what form victims encounter harassing or inappropriate activity in the metaverse while each of the three dimensions addresses characteristics unique to the metaverse that might affect each user differently. The second level refers to subcategories that further specify the reason why those situations can be seen as harassing or child inappropriate. It distinguishes between content or behavior that is insulting/offensive, sexual, violent, discriminating, self-harm promoting, or containing other child-inappropriate content or behavior (e.g., gambling, substance abuse, promotion of dangerous activities). The third level of the categorization instrument is an even more detailed specification of the previously named categories in level two. Occurrences on this level are henceforth mentioned as *instances*. See Appendix 7.1 and Appendix 7.2 for a more detailed overview of the categorization instrument's exact structure as well as definitions for its categories. This multi-level categorization instrument enables a nuanced understanding of the diverse forms of harassing/inappropriate content or behavior in the chosen metaverse spaces. This facilitates the comprehensive assessment of different threats and the development of governance/safety systems to mitigate the risks.

To create a common understanding of what criteria this paper utilizes to identify and evaluate harassing and child-inappropriate content or behavior, a foundation for certain concepts had to be laid. In terms of how exactly harassment is defined in literature, there is no clear consensus reached. Besides the broad distinction into verbal, physical, and environmental harassment Blackwell et al. (2019) emphasized the highly subjective and personal nature of how exactly types and degrees of harassment are evaluated differently depending on the person. This aligns with past research that indicates problems regarding the clear definition of what exact behavior and content can be seen as harassing and what not (Duggan, 2017; Wolak et al., 2007). In order to create a basis for how online harassment is identified this paper orientates itself to the inclusive definition described by Blackwell et al. (2019) which says, "Online harassment refers to a broad spectrum of abusive behaviors enabled by technology platforms". As examples, this and various other papers mentioned insults, the publication of personal information, shaming, violent threats or behavior, unwanted sexual conversations or actions, and more generally all types of offensive behavior and content that might make other users uncomfortable (Blackwell et al., 2018, 2019; Duggan, 2017; Wolak et al., 2007).

While counting all different types of harassment to be inappropriate for every kind of user (including children and adolescents) this paper additionally includes all other factors that can induce any type of psychological or physiological harm to a minor to be inappropriate. These contain all kinds of pornographic content, racist/discriminating behavior, or promotion of dangerous or addictive activities such as substance abuse, self-harm, or gambling (Strasburger et al., 2012; Weimann & Masri, 2023; *What Parents Need to Know About Inappropriate Content | Internet Matters*, 2021).

3.2 Results

The following section will provide an overview of situations identified during the experiments in which child-inappropriate content or behavior occurred. The overall nature of these encounters and the most relevant occurrences are described below. Additionally, the frequency of inappropriate *instances* identified was visualized via Figure 1 and Figure 2.

3.2.1 VRChat

During the three hours spent in "VRChat" a total number of 19 situations were identified in which inappropriate content or behavior took place. All of these situations fell under one of the distinctions of harassment derived from Blackwell et al. (2019) while simultaneously violating the community guidelines described on the application website (*Community Guidelines/VRChat*, 2023). In terms of verbal harassment, it became evident that verbal abuse and insults among users can be seen as common conditions in this space. The type of this verbally harassing behavior taking place in "VRChat" is often of sexual nature and originates due to sexually motivated conversations or actions among the avatars. This includes for example situations in which avatars engaged in sexual conversation or actions that subsequently provoked other users to use profane language towards them due to their own

dissatisfaction with the situation. Besides sexualized insults, other characteristics of verbal harassment included racist as well as physically threatening components. For the latter, the common theme was threatening other users in terms of revealing their real physical address and ultimately stating intentions to cause them physical harm. In some situations, this behavior is even channeled into an incitement for self-harm or even suicide. In addition to racist slurs and insults, one specific situation was defined by a user glorifying the actions performed by Adolf Hitler and the Third Reich while simultaneously trying to convince other avatars to follow his actions such as copying his national socialistic avatar appearance. When it comes to the promotion of substance abuse it is worth mentioning that the identified situations all included multiple users instead of one lone perpetrator. Those included group conversations of users that either already consumed alcohol and tried to encourage others to do the same or the glorification of drugs such as LSD.

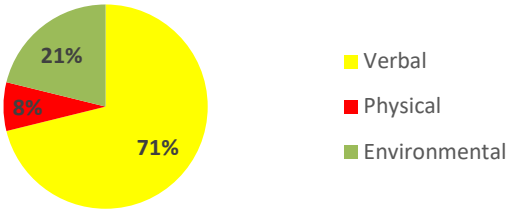

Figure 1: *Instances* of harassment and inappropriate content/behavior per overarching category for both spaces

Regarding physical harassment between avatars, one situation occurred in which one user pretended to physically assault another during an argument after which the victim left the game. Furthermore, multiple situations of clearly nonconsensual groping and sexual touching were identified where the unwillingness of the victim was obvious due to their complaints of the situation via voice chat. One extreme example of this was a user that described himself as an adult male and repeatedly engaged in sexual talk and groping towards another user who claimed to be an underage girl. This situation can be seen as an attempt for online grooming due to the fact that even after the age of the girl was known the adult man did not stop his attempts of trying to persuade her. When it comes to environmental harassment various situations were categorized in which avatars engaged in sexual interaction, conversations, and drawings for the whole space to witness. This comprised simple touching of explicit body parts as well as an "orgy" including multiple characters performing sexual positions. Besides that, another environmental harassment that took place refers to the previously mentioned situation of a user who changed the appearance of his avatar to one that incorporates swastikas for everyone to see.

3.2.2 RecRoom

Compared with the previous application, "RecRoom" contained substantially fewer inappropriate situations during the 3-hours of play. In six situations types of harassment were found. Any occurrences of physical harassment were not the case here. Since the focus of "RecRoom" represents engaging in various game modes all types of verbal harassment took place in the context of a game. These consisted of verbal insults due to defeat, teasing the opponent during a fight through profane language, and threats of physical violence. The latter occurred during a game mode after the opponent already repeatedly mentioned swear words, some of them including racist connotations. Furthermore, one situation took place in which one user vocalized a homophobic statement in the context of a group conversation regarding gender fluidity. Given the situation that one game mode provided the possibility to interact with virtual alcohol, multiple users began to drink the artificial beverages, act drunk, and encouraged others to participate. This represents a case of environmental child inappropriate behavior due to the clear glorification and incitement of alcohol consumption. Another situation worth mentioning is the broadcasting of an explicit song by a user for all other members of the space to hear. While in itself the unwelcome broadcasting of music can be already seen as a form of harassment, this song included various insults, racist slurs, glorification of illegal substances, and overall profane language.

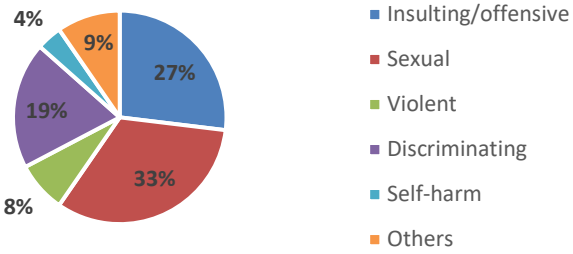

Figure 2: *Instances* of harassment and inappropriate content/behavior per subcategory for both spaces

Concluding, a significant positive relationship between the independent variable of "Entering social or gaming spaces in the metaverse" and the dependent variable of "Encountering types of child inappropriate content or behavior" became evident for both examined spaces. Due to the identified various types of inappropriate content or behavior the paper accepts the previously stated hypothesis. Further implications will be discussed in the following section.

See Appendix 7.3 for a more detailed overview regarding the frequency of *instances* per overarching category and subcategory.

4. DISCUSSION

4.1 Potential impacts on children and adolescents

This section provides a more detailed evaluation of why children and adolescents might have a higher risk potential concerning the identified types of inappropriate content or behavior and what potential impacts this might have on their psychological and physiological well-being.

Starting with the type of harassment that predominantly took place during the research one can compare the occurrence of insults and intimidation with aspects of cyberbullying prevalent in traditional social media or gaming spaces over the past two decades. Generally, research stated that children can be placed among the user segments who are targeted to a greater degree than others regarding harassing activities. This is facilitated by their easy identification either via their voice or their avatar's height (Blackwell et al., 2019). Since the concept of cyberbullying is generally more prevalent among children and adolescents it is to be assumed that similar dynamics will occur

during the usage of the metaverse (Tokunaga, 2010). According to a study published in 2006, the ages between 13 and 17 represented 89% of online harassment instances (Wolak et al., 2006). After the potential effects of cyberbullying are studied extensively in literature it becomes clear how impactful and dangerous the effects of it or other forms of harassment will be for children and adolescents in the metaverse. Known outcomes for their psychological and physiological health include stress, anxiety, depression, loneliness, substance abuse, emotional distress, anger, and most concerning a positive relationship with suicidal ideation among victims (Kowalski et al., 2014; Tokunaga, 2010). The latter might be even increased due to the identified encouragement and promotion of self-harm/suicide that occurred in the social space of "VRChat". The heightened feeling of immersion and presence induced by the technological characteristics often attributed to the metaverse (e.g., HMDs) have the potential to further amplify the mentioned effects for younger users. As mentioned by Goltz (2011) and Lee et al. (2021) the "blurring of borders" between VR and real life, which especially children are at risk of, could either intensify negative effects while simultaneously creating the risk of adapting certain malicious behaviors and projecting them into real life. This aligns with the research of Lavoie et al. (2021) which states how immersion in a virtual space can intensify negative emotions (possibly induced by online harassment) while assuming that this effect will be even more severe in the case of younger users. Since the concept of the metaverse itself aspires to become an alternative reality coexisting with real life in the form of a virtual mirror world, the extent of cyberbullying might even exceed the current impacts created by traditional social media or gaming due to its potential of becoming a highly immersive space in which children and adolescents spend a majority of their time (Dwivedi et al., 2022; Kowalski et al., 2014; Lee et al., 2021; Tokunaga, 2010).

In terms of sexual harassment similar concepts can be applied here. The occurrence of these situations especially became evident during the time the author spent in "VRChat". Multiple situations included the use of highly sexualized language, explicit actions performed amongst other users, or the attempt of older users to groom minors. These are all situations that are imperative to be seen as child-inappropriate due to their potential of causing long-term emotional distress, trauma, and problematic sexual behavior (Mori et al., 2023; Strasburger et al., 2012; Usmani et al., 2022; Weimann & Masri, 2023; *What Parents Need to Know About Inappropriate Content | Internet Matters*, 2021). Children and adolescents are at risk due to their careless and inexperienced nature. The situation worth mentioning in this regard is the situation identified in "VRChat" where a user, that mentioned to be an adult male in his thirties, attempted to perform sexual conversation and actions with a girl claiming to be a minor. Based on knowledge derived from child psychology literature one can reason that especially adolescents might be at risk of becoming victims of grooming and sexual misconduct. According to research and the content of the interviews that were conducted for this paper, adolescents in the time of puberty tend to gradually decrease their orientation and dependency on their parents while simultaneously searching for new peer groups to interact with (Remschmidt, 1994). Furthermore, the mental state during puberty can be defined by uncertainty and intensive changes in brain areas that are mainly involved in human emotional and social behavior which subsequently can make people of this age vulnerable to sexual predators (Jonsson et al., 2019; Korte, 2020; Weimann & Masri, 2023). The search for acceptance by new peer groups can also raise the risk for adolescents of becoming persuaded by extremist groups all facilitated by their careless and unexperienced nature (Kartit et al., 2022; Remschmidt, 1994; Weimann & Masri, 2023). According to Usmani et al. (2022), sexual harassment inside virtual spaces has the potential to trigger similar effects as in comparable situations occurring in real life which includes the development of PTSD, depression, anxiety, or insomnia. This corresponds with the statement given by interviewee E who indicates that if such traumas are experienced during the restructuring of an adolescent's brain, they can have life-long effects. Similar findings are presented by Goltz (2011) who mentioned that "in the years of sexual maturation, similar synapse prosperity exists in the frontal cortex of the brain, the area responsible for judgment and inhibition, but that these synapses similarly diminish to fit adult levels following sexual maturation. In both cases, the synapses that survive are those reinforced by environmental interaction." In this context, the possibility exists that adolescents become used to certain characteristics, behavioral dynamics, or preferences that are prevalent or unique to the metaverse. These might include substance abuse, aggressive behavior, racist behavior, or the development of sexual preferences linked to virtual worlds/characters. These are all aspects that were possible to encounter during the carried-out method.

Lastly, the correlation described in literature between an addiction to online gaming and an addiction to substance abuse might be further facilitated by the identified glorification of alcohol and other substances (Van Rooij et al., 2014). In this case, the proteus effect has the potential to facilitate this process even further. The embodiment of a digital representation by the younger user might induce changes in character, interests, and beliefs which makes it possible to assume that the glorification of alcohol consumption during these game modes creates similar cravings in real life (Fox et al., 2013; Goltz, 2011). The same reasoning applies to the availability of child-inappropriate avatars that users can embody in "VRChat". Ranging from highly sexualized avatars that are defined through exaggerated body features or explicit clothing up to appearances including racist symbols it is certain to assume that these can have negative psychological impacts for minors. Applying the Proteus effect, one can argue that embodying for example a sexualized avatar for a longer time might have severe effects for especially girls in their puberty potentially leading to an increase in body-related thoughts, body dysmorphia, or eventually anorexia (Fox et al., 2013; Shaffer & Kipp, 2007; Usmani et al., 2022). "RecRoom" on the other hand does not offer the same customization possibilities as "VRChat" since their avatars are bound to certain base features while no extensive customization of their appearance (e.g., exaggeration of certain body parts) is possible. Also, no explicit user-generated content was identified during the experiment which suggests that the applications governance system regarding uploaded content operates effectively (*Creator Code of Conduct/RecRoom*, n.d.).

4.2 Role of the platforms in terms of risk mitigation

Looking back on all the described inappropriate/harassing behavior and content that was identified during the method it becomes clear that an overarching and protective governance/safety system needs to be in place in order to ensure the protection of all types of user segments, especially the most vulnerable ones such as minors. The following section will give a brief description of what degree governance/safety systems are currently implemented in the examined virtual spaces and how effective they were during the experiment.

4.2.1 VRChat

One governance/safety method is a rank system to quickly identify who is a trusted user and who is not. Depending on the total hours spend inside of "VRChat", users have the possibility to gradually increase their rank over time (*VRChat Safety and Trust System*, n.d.). These ranks decide what kind of possibilities the user has in this application (e.g., uploading self-created appearances or worlds) while also playing an important role in the social dynamics within those spaces. Through this kind of system, users can assess what kind of experience other users have on this application while receiving a preliminary estimation of how trustworthy another person is. Still, users of any rank have the possibility to perform harassing actions or to share inappropriate content with others for which "VRChat" implemented additional control measures that enable each player to modify how other users can interact with them. These measures include muting, hiding the avatar's appearance/effects, or blocking them completely which restricts any kind of interaction or visibility. Furthermore, users have the opportunity to report harassing or malicious activities, that violate the applications Community Guidelines, to the applications support team which ultimately can lead to a ban of the reported user (*Community Guidelines/VRChat*, 2023.; *I Want to Report Someone/VRChat*, 2022). Real-time governance for example in the form of moderators participating in the spaces to identify and inhibit malicious behavior, was not experienced during the experiment.

4.2.2 RecRoom

One aspect that stands out for "RecRoom" is the availability of so-called Junior Accounts. These accounts are for players below the age of 13 and encompass features that increase the safety of younger users such as the prevention of sharing certain sensible data and the opportunity for parents to personally manage all settings of their child's account (*Comfort and Safety/RecRoom*, n.d.). All players are repeatedly reminded throughout the game to follow the code of conduct representing the basis of how users should behave during their duration of play (*Code of Conduct/RecRoom*, n.d.). In case somebody violates those rules two kinds of governance/safety systems are in place. Firstly, each player has the chance to mute, block, or report any inappropriate user and world while additionally, an opportunity exists to start a vote among all users present in the current space to kick one player out. Furthermore, each player can create and enter certain safe spaces in which no other player can interact with them. The second type of governance/safety system does not need to be performed by the players themselves but is an automated voice moderation system that should be able to detect and report discriminatory, sexual, harassing, and abusive language. However, it is concerning that the general code of conduct does not explicitly forbid inappropriate aspects such as substance abuse, gambling or violence among users (*Code of Conduct/RecRoom*, n.d.).

4.3 Implications, limitations, and recommendations for future research

The framework provided by this paper gives researchers an identification and categorization mechanism which can be utilized for future content analyses of any metaverse space. With the focus laying on the categorization of child-inappropriate content or behavior, this framework consists of an exhaustive catalog to differentiate between occurrences of various natures as well as to make a statement regarding their frequency during the duration of the experiment. In a practical sense, this paper hopes to facilitate other researchers, policymakers, and developers to use this categorization framework for developing governance/safety systems that actively restrain these identified risks from occurring. This fosters a safe digital environment that offers enriching experiences while mitigating potential harm.

The most relevant limitation of this framework was the amount of time spent inside each of the metaverse spaces. While the categorization instrument is easily applicable to long-term studies, the three hours spent in each space cannot be seen as representative enough to make assumptions regarding the average frequency of inappropriate content or behavior identified or the potential difference between the two spaces. Significant differences might occur if the study duration would be increased. Furthermore, inside both applications, numerous subspaces exist that might be defined by different user segments and behavioral dynamics. The goal of this study was to provide an exemplary overview of what inappropriate situations children and adolescents can possibly encounter during their time in the metaverse. Another factor that became apparent during the research was that the application's current governance/safety systems are mainly user-centric and dependent on their collaboration in terms of either temporarily blocking harassing users themselves or reporting them to the designated help desk. The real-time voice chat moderation included in "RecRoom" is the only identifiable method which does not rely on users' participation. Nevertheless, it did not fully restrict inappropriate language from occurring. The available time and scope of the framework also acted as a limitation by inhibiting the comparison of various governance/safety systems implemented in other metaverse spaces. This might have given rise to a more detailed evaluation and recommendations for improvement. While the scope of the conducted research focused on what inappropriate content or behavior can be identified in the metaverse that can directly impact the user's psychological and physiological health, it excluded the numerous amounts of privacy concerns this new technology brings with it (Allam et al., 2022; Bibri & Allam, 2022; Dwivedi et al., 2022; Falchuk et al., 2018; Fernandez & Hui, 2022; Lee et al., 2021). Similar to health risks, this topic represents another important research domain that could pose real threats to user safety and especially to minors due to their lesser knowledge and care regarding how they should handle data on the internet (O'Keeffe et al., 2011). Lastly, an obvious limitation of the chosen method was the ethical considerations when it comes to conducting experiments with the help of children. This bioethical aspect of human research makes the examination of domains related to minors a complex long-term multifactorial process, especially when it is focusing on the evaluation of potentially harmful innovative technologies such as the metaverse (Kaimara et al., 2022; Neill, 2005).

These limitations lead up to the three main recommendations this paper synthesized for future research to follow.

1. **Increasing the time frame and number of metaverse spaces that are being evaluated.**

By following this recommendation, researchers will be able to make more reliable and valid statements about both the average frequency of different kinds of inappropriate content or behavior as well as the differences between various types of metaverse spaces. This is due to a much larger sample size of different spaces being evaluated and the total duration that is spent in the metaverse. Additionally, it will be possible to evaluate and compare differences between certain countries by entering the metaverse through a VPN. Such a long-term study, lasting for example one full year, would increase the reliability and validity of the research framework.

2. **Comparison and evaluation of different governance systems existing in the metaverse.**

The metaverse as a concept exists of numerous different worlds operating independently from each other with every one of them containing their own specific governance/security system and rules (Lee et al., 2021). Since no overarching institution or government has control over all these globally available spaces the creation of a universal governance system seems nearly impossible even though standardization and interoperability among the space would create a common basis for all users to follow (Dwivedi et al., 2022). Still, the literature suggests that the developers of such virtual reality technologies and spaces have an ethical responsibility to ensure the user's well-being while entering their immersive environments (Lavoie et al., 2021). This led to the second recommendation for future research to more closely evaluate different governance/safety systems inside the metaverse, detect weaknesses and opportunities for improvement, and subsequently develop an appropriate framework to test their effectiveness. The focus here should lay especially on systems that don't depend on the collaboration of users but instead provide protection against inappropriate content or behavior in real-time. Examples of this might include active moderation by staff being present in the spaces or a detection/blocking system powered by artificial intelligence (Allam & Dhunny, 2019).

3. **Identification and categorization of privacy threats inside the metaverse and how organizations can ensure data protection for their users.**

When entering the metaverse numerous types of personal data are collected so that all functions encompassing it can be carried out. These include personal data (e.g., name, age, address, gender) to set up accounts, biometric and spatial data collected by the HDM, payment information needed to purchase items/applications, or conversations among users (Falchuk et al., 2018; Fernandez & Hui, 2022). All this information could get abused and utilized against the user by malicious intruders in case of a hacking attack or data breach. This paper suggests future researchers to develop a framework that aids the identification and categorization of privacy threats that could occur inside the metaverse to support developers and policymakers to further fortify the data protection mechanisms for users and prevent the misuse of their information.

5. CONCLUSION

This paper focused on answering the research questions: **What types of child-inappropriate content or behavior can be identified in social or gaming spaces of the metaverse that children and adolescents might encounter during their play?** The combination of an initial literature review and expert interviews offered insights into how the current state of research evaluates the risks children and adolescents face in virtual environments. Subsequently, a conceptual framework was developed that guided the identification and categorization of different types of child-inappropriate content or behavior found in the social metaverse space "VRChat" and in the gaming metaverse space "RecRoom". The findings of this research prove the existence of numerous types of child-inappropriate content or behavior. These are for example of insulting, sexual, violent, or discriminating nature while ranging from verbal over physical up to environmental types of harassment (Figure 1, Figure 2, Appendix 7.3). This sheer variety of different inappropriate aspects identified during the research provides an answer to the overarching research question of this paper. Furthermore, it is safe to say that there exists a significant positive relationship between the independent variable and the dependent variable. The stated hypothesis was proven to be correct (Table 2).

Concluding, the results of this paper prove that the metaverse currently does not provide a safe environment for children and adolescents to take part in. The prevalence of child-inappropriate content or behavior is evident and must not be ignored by parents, developers, or policymakers. Psychological or physiological harms might become the result of minors entering virtual worlds potentially leading to long-term consequences. Compared with the closely related concepts of social media and online gaming a clear tendency is observable that the metaverse recreates similar issues that were already examined in those prior online spaces. Additionally, aggravated by the heightened factors of immersion, presence, and embodiment more severe follow-up effects of harassing and inappropriate content or behavior are expected. By experiencing traumas that could profoundly impact their entire life, minors are at risk of suffering from altered brain development and lasting emotional and physical distress. Moreover, our study highlights the shortcomings of the examined governance/safety systems to provide sufficient enforceability of their own code of conduct. The lack of active moderation, coupled with ineffective enforcement of rules and a lack of immediate punishment for perpetrators, leaves children and adolescents vulnerable to harm. It is imperative for platform owners to take responsibility and implement an active governance/safety system that holds malicious users accountable and actively safeguards younger users.

However, it stays essential to mention the limitations of this research. Firstly, the duration and extensiveness of research that took place inside the metaverse do not allow it to make reliable and valid conclusions regarding the exact frequency of certain occurrences as well as the differences between multiple spaces and subspaces. Secondly, the chosen framework was solely developed to identify and categorize child-inappropriate content or behavior while neglecting the essential part of how developers and policymakers can mitigate those risks for threatened user segments. Thirdly, in terms of threats that minors could encounter in the metaverse, this paper examined how this technology might affect their psychological and physiological well-being without including potential dangers to their data privacy. To address these limitations, future researchers are recommended to extend the scope and time frame of the method. This provides them with the opportunity to conduct a more comprehensive, valid, and reliable comparative analysis of different metaverse spaces that is additionally yielding valuable insights into the effectiveness of current governance/safety systems and related risks for the user's privacy.

It has to be mentioned that the metaverse needs to be seen as a technological tool that provides users with numerous possibilities where both positive and negative implications for society need to be considered. Aspects such as an improvement of education possibilities, remote working, or health care applications represent valuable ways to utilize this new technology in ways that weren't available before. What kind of negative effects arise from it highly depends on certain circumstances, pre-conditions, or other environmental factors that lead to a negative impact on a user's life.

In conclusion, this study has laid a foundation for comprehending the risks children and adolescents face in the metaverse concerning inappropriate content or behavior. By acknowledging its limitations and offering recommendations for future researchers to follow, this paper hopes to achieve further investigations that will contribute to a more secure digital environment, especially for younger users.

6. REFERENCES

Abbate, S., Centobelli, P., Cerchione, R., Oropallo, E., & Riccio, E. (2022). *A first bibliometric literature review on Metaverse*. 254–260. https://doi.org/10.1109/TEMSCONEUROPE54743.2022.9802015

Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. *Cities*, *89*, 80–91. Scopus. https://doi.org/10.1016/j.cities.2019.01.032

Allam, Z., Sharifi, A., Bibri, S. E., Jones, D. S., & Krogstie, J. (2022). The Metaverse as a Virtual Form of Smart Cities: Opportunities and Challenges for Environmental, Economic, and Social Sustainability in Urban Futures. *Smart Cities*, *5*(3), 771–801. https://doi.org/10.3390/smartcities5030040

Anderson, C. A., Shibuya, A., Ihori, N., Swing, E. L., Bushman, B. J., Sakamoto, A., Rothstein, H. R., & Saleem, M. (2010). Violent video game effects on aggression, empathy, and prosocial behavior in eastern and western countries: A meta-analytic review. *Psychological Bulletin*, *136*(2), 151–173. https://doi.org/10.1037/a0018251

Andreassen, C. S., Pallesen, S., & Griffiths, M. D. (2017). The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. *Addict Behav*, *64*, 287–293. Medline. https://doi.org/10.1016/j.addbeh.2016.03.006

Ashour, A. S., Babo, R., Bhatnagar, V., Bouhlel, M. S., & Dey, N. (Hrsg.). (2018). *Social Networks Science: Design, Implementation, Security, and Challenges: From Social Networks Analysis to Social Networks Intelligence* (1st ed. 2018). Springer International Publishing : Imprint: Springer. https://doi.org/10.1007/978-3-319-90059-9

Bailey, J. O., & Bailenson, J. N. (2017). Considering virtual reality in children's lives. *Journal of Children and Media*, *11*, 107–113. https://doi.org/10.1080/17482798.2016.1268779

Benrimoh, D., Chheda, F. D., & Margolese, H. C. (2022). The Best Predictor of the Future—The Metaverse, Mental Health, and Lessons Learned From Current Technologies. *JMIR Mental Health*, *9*, e40410–e40410. https://doi.org/10.2196/40410

Bibri, S. E., & Allam, Z. (2022). The Metaverse as a Virtual Form of Data-Driven Smart Urbanism: On Post-Pandemic Governance through the Prism of the Logic of Surveillance Capitalism. *Smart Cities*, *5*(2), 715–727. https://doi.org/10.3390/smartcities5020037

Błachnio, A., Przepiorka, A., & Pantic, I. (2016). Association between Facebook addiction, self-esteem and life satisfaction: A cross-sectional study. *Computers in Human Behavior*, *55*, 701–705. https://doi.org/10.1016/j.chb.2015.10.026

Blackwell, L., Chen, T., Schoenebeck, S., & Lampe, C. (2018). When Online Harassment Is Perceived as Justified. *Proceedings of the International AAAI Conference on Web and Social Media*, *12*(1), Article 1. https://doi.org/10.1609/icwsm.v12i1.15036

Blackwell, L., Ellison, N., Elliott-Deflo, N., & Schwartz, R. (2019). Harassment in social virtual reality: Challenges for platform governance. *Proceedings of the ACM on Human-Computer Interaction*, *3*, 1–25. https://doi.org/10.1145/3359202

Bragesjö, M., Larsson, K., Nordlund, L., Anderbro, T., Andersson, E., & Möller, A. (2020). Early Psychological Intervention After Rape: A Feasibility Study. *Frontiers in Psychology*, *11*, 1595. https://doi.org/10.3389/fpsyg.2020.01595

Cairns, P., Cox, A., & Nordin, A. I. (2014). Immersion in Digital Games: Review of Gaming Experience Research. In M. C. Angelides & H. Agius (Hrsg.), *Handbook of Digital Games* (S. 337–361). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118796443.ch12

Cao, S., Nandakumar, K., Babu, R., & Thompson, B. (2020). Game play in virtual reality driving simulation involving head-mounted display and comparison to desktop display. *Virtual Reality*, *24*(3), 503–513. https://doi.org/10.1007/s10055-019-00412-x

Chen, G. M., Pain, P., Chen, V. Y., Mekelburg, M., Springer, N., & Troger, F. (2020). 'You really have to have a thick skin': A cross-cultural perspective on how online harassment influences female journalists. *Journalism*, *21*(7), 877–895. https://doi.org/10.1177/1464884918768500

Choo, H., Gentile, D. A., Sim, T., Li, D., Khoo, A., & Liau, A. K. (2010). Pathological video-gaming among Singaporean youth. *Annals of the Academy of Medicine, Singapore*, *39*(11), 822–829.

Cloete, R., Norval, C., & Singh, J. (2020). *A Call for Auditable Virtual, Augmented and Mixed Reality*. 1–6. https://doi.org/10.1145/3385956.3418960

Code of Conduct. (n.d.). Rec Room. Accessed on July 7th 2023, from https://recroom.com/code-of-conduct

Comfort and Safety. (n. d.). Rec Room. Accessed on July 12th 2023, from https://recroom.com/safety

Community Guidelines. (2023). VRChat. Accessed on July 7th 2023, from https://hello.vrchat.com/community-guidelines

Creator Code of Conduct. (n.d.). Rec Room. Accessed on July 11th 2023, from https://recroom.com/ccoc

Damar, M. (2021). Metaverse Shape of Your Life for Future: A bibliometric snapshot. *Journal of Metaverse*, *1*(1), 1–78.

Dangers of the metaverse 2021. (2022, July 7). Statista. Accessed on May 14th 2023 from https://www.statista.com/statistics/1288822/metaverse-dangers/

Dorn, S. D. (2015). Digital Health: Hope, Hype, and Amara's Law. *Gastroenterology*, *149*(3), 516–520. https://doi.org/10.1053/j.gastro.2015.07.024

Duggan, M. (2017, July 11). Online Harassment 2017. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2017/07/11/online-harassment-2017/

Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M. M., Dennehy, D., Metri, B., Buhalis, D., Cheung, C. M. K., Conboy, K., Doyle, R., Dubey, R., Dutot, V., Felix, R., Goyal, D. P., Gustafsson, A., Hinsch, C., Jebabli, I., … Wamba, S. F. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *66*, 102542–102542. https://doi.org/10.1016/j.ijinfomgt.2022.102542

Falchuk, B., Loeb, S., & Neff, R. (2018). The Social Metaverse: Battle for Privacy. *IEEE Technology and Society Magazine*, *37*(2), 52–61. https://doi.org/10.1109/MTS.2018.2826060

Fernandez, C. B., & Hui, P. (2022). *Life, the Metaverse and Everything: An Overview of Privacy, Ethics, and Governance in Metaverse*. 272–277. https://doi.org/10.1109/icdcsw56584.2022.00058

Fox, J., Bailenson, J. N., & Tricase, L. (2013). The embodiment of sexualized virtual selves: The Proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior*, *29*(3), 930–938. https://doi.org/10.1016/j.chb.2012.12.027

Goltz, N. (2011). ESRB Warning: Use of Virtual Worlds by Children May Result in Addiction and Blurring of Borders – The Advisable Regulations in Light of Foreseeable Damages. *Pittsburgh Journal of Technology Law and Policy*, *11*. https://doi.org/10.5195/tlp.2011.57

Gottschalk, F. (2019). Impacts of technology use on children: Exploring literature on the brain, cognition and well-being. *OECD Education Working Papers*, Article 195. https://ideas.repec.org//p/oec/eduaab/195-en.html

Griffiths, M. D., Davies, M. N. O., & Chappell, D. (2004). Online computer gaming: A comparison of adolescent and adult gamers. *Journal of Adolescence*, *27*(1), 87–96. https://doi.org/10.1016/j.adolescence.2003.10.007

Han, D. I. D., Bergs, Y., & Moorhouse, N. (2022). Virtual reality consumer experience escapes: Preparing for the metaverse. *Virtual Reality*, *26*(4), 1443–1458. https://doi.org/10.1007/s10055-022-00641-7

Holsapple, C. W., & Wu, J. (2007). User acceptance of virtual worlds. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, *38*(4), 86–89. https://doi.org/10.1145/1314234.1314250

Hou, Y., Xiong, D., Jiang, T., Song, L., & Wang, Q. (2019). Social media addiction: Its impact, mediation, and intervention. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, *13*(1). https://doi.org/10.5817/CP2019-1-4

Horizon Worlds | Virtual Reality Worlds and Communities. (2023). Accessed on July 7th 2023, from https://www.meta.com/horizon-worlds/

Hutton, J. S., Dudley, J., Horowitz-Kraus, T., DeWitt, T., & Holland, S. K. (2020). Associations Between Screen-Based Media Use and Brain White Matter Integrity in Preschool-Aged Children. *JAMA Pediatrics*, *174*(1), e193869. https://doi.org/10.1001/jamapediatrics.2019.3869

I want to report someone. (2022, Oktober 13). VRChat. Accessed on July 7th 2023. https://help.vrchat.com/hc/en-us/articles/360062658553-I-want-to-report-someone

Jonsson, L. S., Fredlund, C., Priebe, G., Wadsby, M., & Svedin, C. G. (2019). Online sexual abuse of adolescents by a perpetrator met online: A cross-sectional study. *Child and Adolescent Psychiatry and Mental Health*, *13*, 32. https://doi.org/10.1186/s13034-019-0292-1

Kaimara, P., Oikonomou, A., & Deliyannis, I. (2022). Could virtual reality applications pose real risks to children and adolescents? A systematic review of ethical issues and concerns. *Virtual Reality*, *26*(2), 697–735. https://doi.org/10.1007/s10055-021-00563-w

Kanai, R., Bahrami, B., Roylance, R., & Rees, G. (2012). Online social network size is reflected in human brain structure. *Proceedings. Biological Sciences*, *279*(1732), 1327–1334. https://doi.org/10.1098/rspb.2011.1959

Kartit, D., Howcroft, E., & Howcroft, E. (2022, Oktober 27). Interpol says metaverse opens up new world of cybercrime. *Reuters*. https://www.reuters.com/technology/interpol-says-metaverse-opens-up-new-world-cybercrime-2022-10-27/

Kenney, E. L., & Gortmaker, S. L. (2017). United States Adolescents' Television, Computer, Videogame, Smartphone, and Tablet Use: Associations with Sugary Drinks, Sleep, Physical Activity, and Obesity. *The Journal of Pediatrics*, *182*, 144–149. https://doi.org/10.1016/j.jpeds.2016.11.015

Ko, C.-H., Liu, G.-C., Hsiao, S., Yen, J.-Y., Yang, M.-J., Lin, W.-C., Yen, C.-F., & Chen, C.-S. (2009). Brain activities associated with gaming urge of online gaming addiction. *Journal of Psychiatric Research*, *43*(7), 739–747. https://doi.org/10.1016/j.jpsychires.2008.09.012

Koehler, D., Fiebig, V., & Jugl, I. (2023). From Gaming to Hating: Extreme-Right Ideological Indoctrination and Mobilization for Violence of Children on Online Gaming Platforms. *Political Psychology*, *44*(2), 419–434. Scopus. https://doi.org/10.1111/pops.12855

Korte, M. (2020). The impact of the digital revolution on human brain and behavior: Where do we stand? *Dialogues in Clinical Neuroscience*, *22*(2), 101–111. https://doi.org/10.31887/DCNS.2020.22.2/mkorte

Kowalski, R. M., Giumetti, G. W., Schroeder, A. N., & Lattanner, M. R. (2014). Bullying in the digital age: A critical review and meta-analysis of cyberbullying research among youth. *Psychological Bulletin*, *140*(4), 1073–1137. Scopus. https://doi.org/10.1037/a0035618

Labana, R. V., Hadjisaid, J. L., Imperial, A. R., Jumawid, K. E., Lupague, M. J. M., & Malicdem, D. C. (2020). Online Game Addiction and the Level of Depression Among Adolescents in Manila, Philippines. *Central Asian Journal of Global Health*, *9*(1), e369. https://doi.org/10.5195/cajgh.2020.369

LaViola, J. J. (2000). A discussion of cybersickness in virtual environments. *ACM SIGCHI Bulletin*, *32*(1), 47–56. https://doi.org/10.1145/333329.333344

Lavoie, R., Main, K., King, C., & King, D. (2021). Virtual experience, real consequences: The potential negative emotional consequences of virtual reality gameplay. *Virtual Reality*, *25*(1), 69–81. https://doi.org/10.1007/s10055-020-00440-y

Lawson, E. C. (2021, Dezember 30). *New research shows Metaverse is not safe for kids*. Center for Countering Digital Hate | CCDH. Accessed on May 15th 2023 from https://counterhate.com/blog/new-research-shows-metaverse-is-not-safe-for-kids/

Lee, L.-H., Braud, T., Zhou, P., Wang, L., Xu, D., Lin, Z., Kumar, A., Bermejo, C., & Hui, P. (2021). All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda. *Journal of Latex Class Files*, *14*(8), 1–66. https://doi.org/10.13140/RG.2.2.11200.05124/8

Meshi, D., Tamir, D. I., & Heekeren, H. R. (2015). The Emerging Neuroscience of Social Media. *Trends in Cognitive Sciences*, *19*(12), 771–782. https://doi.org/10.1016/j.tics.2015.09.004

Mori, C., Park, J., Racine, N., Ganshorn, H., Hartwick, C., & Madigan, S. (2023). Exposure to sexual content and problematic sexual behaviors in children and adolescents: A systematic review and meta-analysis. Child Abuse & Neglect, 143, 106255. https://doi.org/10.1016/j.chiabu.2023.106255

Muslihati, Hotifah, Y., Hidayat, W. N., Purwanta, E., Valdez, A. V., 'ilmi, A. M., & Saputra, N. M. A. (2023). Predicting the mental health quality of adolescents with intensive exposure to metaverse and its counseling recommendations in a multicultural context. *Jurnal Cakrawala Pendidikan*, *42*(1), 38–52. https://doi.org/10.21831/cp.v42i1.54415

Neill, S. J. (2005). Research with children: A critical review of the guidelines. *Journal of Child Health Care: For Professionals Working with Children in the Hospital and Community*, *9*(1), 46–58. https://doi.org/10.1177/1367493505049646

Nichols, S., & Patel, H. (2002). Health and safety implications of virtual reality: A review of empirical evidence. *Applied Ergonomics*, *33*(3), 251–271. https://doi.org/10.1016/s0003-6870(02)00020-0

Official Site | Second Life—Virtual Worlds, Virtual Reality, VR, Avatars, and Free 3D Chat. (2023). Accessed on July 7th 2023, from https://secondlife.com/

O'Keeffe, G. S., Clarke-Pearson, K., Mulligan, D. A., Altmann, T. R., Brown, A., Christakis, D. A., Falik, H. L., Hill, D. L., Hogan, M. J., Levine, A. E., & Nelson, K. G. (2011). The Impact of Social Media on Children, Adolescents, and Families. *Pediatrics*, *127*(4), 800–804. https://doi.org/10.1542/peds.2011-0054

Park, S. M., & Kim, Y. G. (2022). A Metaverse: Taxonomy, Components, Applications, and Open Challenges. *IEEE Access*, *10*, 4209–4251. https://doi.org/10.1109/ACCESS.2021.3140175

Paul, I., Mohanty, S., & Sengupta, R. (2022). The role of social virtual world in increasing psychological resilience during the on-going COVID-19 pandemic. *Computers in Human Behavior*, *127*, 107036–107036. https://doi.org/10.1016/j.chb.2021.107036

Paulus, F. W., Ohmann, S., von Gontard, A., & Popow, C. (2018). Internet gaming disorder in children and adolescents: A systematic review. *Developmental Medicine & Child Neurology*, *60*(7), 645–659.

Pfeifer, J. H., & Blakemore, S.-J. (2012). Adolescent social cognitive and affective neuroscience: Past, present, and future. *Social Cognitive and Affective Neuroscience*, *7*(1), 1–10. https://doi.org/10.1093/scan/nsr099

Pokémon GO Death Tracker. (2016). Accessed on June 19th 2023, from https://pokemongodeathtracker.com/

Rec Room. (n.d.). Rec Room. Accessed on July 7th 2023, from https://recroom.com

Regan, C. (1995). An investigation into nausea and other side-effects of head-coupled immersive virtual reality. *Virtual Reality*, *1*(1), 17–31. https://doi.org/10.1007/bf02009710

Remschmidt, H. (1994). Psychosocial Milestones in Normal Puberty and Adolescence. *Hormone Research*, *41*(2), 19–29. https://doi.org/10.1159/000183955

Shaffer, D. R., & Kipp, K. (2007). *Developmental psychology: Childhood and adolescence* (7th ed). Wadsworth/Thomson.

Shin, D. (2018). Empathy and embodied experience in virtual environment: To what extent can virtual reality stimulate empathy and embodied experience? *Computers in Human Behavior*, *78*, 64–73. https://doi.org/10.1016/j.chb.2017.09.012

Shriram, K., & Schwartz, R. (2017). *All are welcome: Using VR ethnography to explore harassment behavior in immersive social virtual reality* (S. 226). https://doi.org/10.1109/VR.2017.7892258

Stephenson, N. (1992). Snow Crash. New York, *Bantam Books*

Strasburger, V. C., Jordan, A. B., & Donnerstein, E. (2012). Children, Adolescents, and the Media: Health Effects. Pediatric Clinics of North America, 59(3), 533–587. https://doi.org/10.1016/j.pcl.2012.03.025

Teng, C.-I. (2010). Customization, immersion satisfaction, and online gamer loyalty. *Computers in Human Behavior*, *26*(6), 1547–1554. https://doi.org/10.1016/j.chb.2010.05.029

Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior*, *26*(3), 277–287. Scopus. https://doi.org/10.1016/j.chb.2009.11.014

Turel, O., Romashkin, A., & Morrison, K. M. (2016). Health Outcomes of Information System Use Lifestyles among Adolescents: Videogame Addiction, Sleep Curtailment and Cardio-Metabolic Deficiencies. *PloS One*, *11*(5), e0154764. https://doi.org/10.1371/journal.pone.0154764

Turel, O., Romashkin, A., & Morrison, K. M. (2017). A model linking video gaming, sleep quality, sweet drinks consumption and obesity among children and youth. *Clinical Obesity*, *7*(4), 191–198. https://doi.org/10.1111/cob.12191

Usmani, S. S., Sharath, M., & Mehendale, M. (2022). Future of mental health in the metaverse. *General Psychiatry*, *35*(4), e100825–e100825. https://doi.org/10.1136/gpsych-2022-100825

Van Rooij, A. J., Kuss, D. J., Griffiths, M. D., Shorter, G. W., Schoenmakers, M. T., & Van de Mheen, D. (2014). The (co-)occurrence of problematic video gaming, substance use, and psychosocial problems in adolescents. *Journal of Behavioral Addictions*, *3*(3), 157–165. https://doi.org/10.1556/JBA.3.2014.013

VRChat. (2023). VRChat. Accessed on July 7th 2023, from https://hello.vrchat.com

VRChat Safety and Trust System. (n.d.). VRChat. Accessed on July 11th , from https://docs.vrchat.com/docs/vrchat-safety-and-trust-system

Weimann, G., & Masri, N. (2023). Research Note: Spreading Hate on TikTok. Studies in Conflict & Terrorism, 46(5), 752–765. https://doi.org/10.1080/1057610X.2020.1780027

What parents need to know about inappropriate content. (2021, Oktober 27). Internet Matters. Accessed on July 7th 2023 from https://www.internetmatters.org/issues/inappropriate-content/learn-about-it/

Witmer, B. G., & Singer, M. J. (1998). Measuring Presence in Virtual Environments: A Presence Questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240. https://doi.org/10.1162/105474698565686

Wolak, J., Mitchell, K., & Finkelhor, D. (2006). *Online Victimization of Youth: Five Years Later*. https://doi.org/10.1037/e525892015-001

Wolak, J., Mitchell, K. J., & Finkelhor, D. (2007). Does Online Harassment Constitute Bullying? An Exploration of Online Harassment by Known Peers and Online-Only Contacts. *Journal of Adolescent Health*, *41*(6, Supplement), S51–S58. https://doi.org/10.1016/j.jadohealth.2007.08.019

Yee, N., Bailenson, J. N., & Ducheneaut, N. (2009). The proteus effect: Implications of transformed digital self-representation on online and offline behavior. *Communication Research*, *36*(2), 285–312. https://doi.org/10.1177/0093650208330254

Zamani, E., Chashmi, M., & Hedayati, N. (2009). Effect of addiction to computer games on physical and mental health of female and male students of guidance school in city of isfahan. *Addiction & health*, *1*(2), 98–104.

7. APPENDIX

7.1 Definition of categorization instrument categories

Table 3: Definition of categorization categories

Type of category	Definition of category
Verbal, physical, and environmental harassment represent the overarching categories for the categorization framework. Insulting/offensive, sexual, violent, discriminating, self-harm promoting, and other child-inappropriate content or behavior represent the subcategories distributed below the overarching categories. The aspects that are here described as definitions of the subcategories describe the sub-subcategories also meant when talking about *instances* during the paper. One situation in which harassment or other child-inappropriate content/behavior was identified during the experiment can encompass multiple *instances* e.g. an insult and act of physical aggression in one situation.	
Verbal harassment/inappropriate content or behavior (Overarching category)	Any type of harassment or inappropriate content/behavior that is performed via voice chat (excluding shared media) or text messages
Physical harassment/inappropriate content or behavior (Overarching category)	Any type of harassment or inappropriate content/behavior that is performed via hostile avatar movements toward another user
Environmental harassment/inappropriate content or behavior (Overarching category)	Any type of harassment or inappropriate content/behavior that is performed via any type of shared media (video, audio, text excluding messages, etc.) or non-hostile avatar movements
Insulting/offensive content or behavior (Subcategory)	Includes insults, profane language, sharing or distributing offensive content (videos, audio, text, etc.), and threats/intimidations with no direct intention of physical harm
Sexual content or behavior (Subcategory)	Includes unwanted explicit/pornographic language, online grooming/solicitation of sexual acts towards minors, sexual gestures, sexual touching/groping, sharing or distributing sexually explicit content, and performing (consensual) sexual behavior for other users to see
Violent content or behavior (Subcategory)	Includes threats of physical harm, encouragement or glorification of violence, acts of virtual aggression or fighting (outside of game modes that include fighting), obstruction of movement, invasion of personal space, and sharing or distributing violent or gory content
Discriminating content or behavior (Subcategory)	Includes slurs, jokes, or offensive language targeting minority segments of society, incitement of hatred or violence against specific groups, (attempting to) recruit other players for extremist groups, performing discriminating movements or gestures, and sharing or distributing racist or discriminating content
Self-harm promoting content or behavior (Subcategory)	Includes discussions or encouragement of self-harm and sharing or distributing suicide-related or triggering content
Other child-inappropriate behavior (Subcategory)	Includes substance abuse promotion, gambling or excessive spending encouragement, promotion of dangerous activities, and sharing or distributing content of the said topics

7.2 Structure of categorization instrument

Table 4: Structure of categorization instrument

Verbal harassment or other verbal inappropriate content/behavior		Physical harassment or other physical inappropriate content/behavior		Environmental harassment or other environmental inappropriate content/behavior	
Insulting or offensive content/behavior - Insulting or offensive language - Verbal threats and intimidation	**Sexual content/behavior** - Explicit or pornographic language - Online grooming or solicitation	**Sexual content/behavior** - Inappropriate sexual movements without actively touching someone - Sexual touching or groping	**Violent content/behavior** - Acts of virtual agression or fighting (outside of game-modes that include fighting)	**Insulting or offensive content/behavior** - Sharing/distributing insulting or offensive content	**Sexual content/behavior** - Sharing/distributing sexually explicit content or performing (consensual) sexual behavior for other users to see
Violent content/behavior - Threats of physical harm - Encouragement or glorification of violence	**Content/behavior including hate speech or discrimination** - Discriminating slurs, jokes, or language targeting minorities - Incitement of hatred or or language - (Attempting to) recruit other players for extremist groups	**Content/behavior including hate speech or discrimiation** - Discriminating/racist movements or gestures		**Violent content/behavior** - Sharing/distributing violent or gory content	**Content/behavior including hate speech or discrimiation** - Sharing/distributing racist or discriminating content
Self-harm or suicide promoting content/behavior - Discussion or encouragment of any kind of self-harm	**Other child-inappropriate content/behavior** - Substance abuse promotion - Gambling or excessive spending encouragment - Promotion of dangerous challenges or activities			**Self-harm or suicide promoting content** - Sharing/distributing suicide related content	**Other child-inappropriate content/behavior** - Sharing/distributing substance abuse promoting content - Sharing/distributing gambling or excessive spending promoting content - Sharing/distributing dangerous challenges or activities promoting content

7.3 Total number of instances identified per category for both spaces combined

Table 5: Total number of *instances* identified per category for both spaces combined

Type of category	Number of instances identified
This table summarizes all *instances* in which child-inappropriate content or behavior were identified for each of the overarching categories and subcategories. The presented results combine the findings of both examined virtual spaces.	
Verbal harassment/inappropriate content or behavior	37 instances identified
Physical harassment/inappropriate content or behavior	4 instances identified
Environmental harassment/inappropriate content or behavior	6 instances identified
Insulting/offensive content or behavior	14 instances identified
Sexual content or behavior	17 instances identified
Violent content or behavior	4 instances identified
Discriminating content or behavior	10 instances identified
Self-harm promoting content or behavior	2 instances identified
Other child-inappropriate content or behavior	5 instances identified