BRACKET
Foundation

GAMING AND THE METAVERSE

The Alarming Rise of Online Sexual Exploitation and Abuse of Children Within the New Digital Frontier

unicri
United Nations
Interregional Crime and Justice
Research Institute

CENTRE FOR
ARTIFICIAL
INTELLIGENCE
AND ROBOTICS

BRACKET
CAPITAL

value
for good®

CONTENTS

EXECUTIVE SUMMARY

The online sexual exploitation and abuse of children has become a troubling byproduct of expanding digital platforms, which are increasingly a part of children's daily lives. The 2019 report "Artificial Intelligence: Combating Online Sexual Abuse of Children"[1] highlighted the growth of such abuse and technology's potential to address it. A 2018 study found that more than half of U.S. teens (59%) had experienced some type of cyberbulling; in the last three years, the number of grooming instances recorded by police jumped by (nearly/over) 70% to an all-time high in 2021. Child Sexual Abuse Material (CSAM) reports have continued to increase from 1.1 million in 2014 to 29.3 million in 2021 when over 84 million CSAM images and videos were discovered – nearly a 27-fold increase in reports within seven years.

In recent years, the transformational evolution of gaming has opened new avenues of online child abuse. Most notably, there has been a shift from closed-platform gaming experiences to virtual spaces, which enable a wider array of social interactions, including, making friends, connecting with other players and consuming content. These social gaming platforms are very similar to spaces currently being marketed as "metaverses" and increasingly being accessed via virtual reality (VR) headsets. These social gaming platforms intensify the risks children face online through their distinct characteristics:

- **Immersive intensity.** The immersiveness and lifelike nature of experiences and interactions makes it easier for perpetrators and groomers to find and build trusted relationships with minors.

- **Anonymity and ease of interaction.** Social gaming platforms encourage interactions with strangers in an anonymous environment. The competitive nature of many platforms further incentivizes minors to compete with and against adults, which increases the risk of contact with predators.

- **Multitude of activities.** In social gaming platforms, users engage in multiple sensory experiences including gestures, voice and chat – creating multiple communication modes, thereby inhibiting moderation.

- **Enclosed consumption.** Content on social gaming platforms is increasingly consumed through headsets making it more challenging for someone (such as a parent) to monitor the type of content consumed by a child.

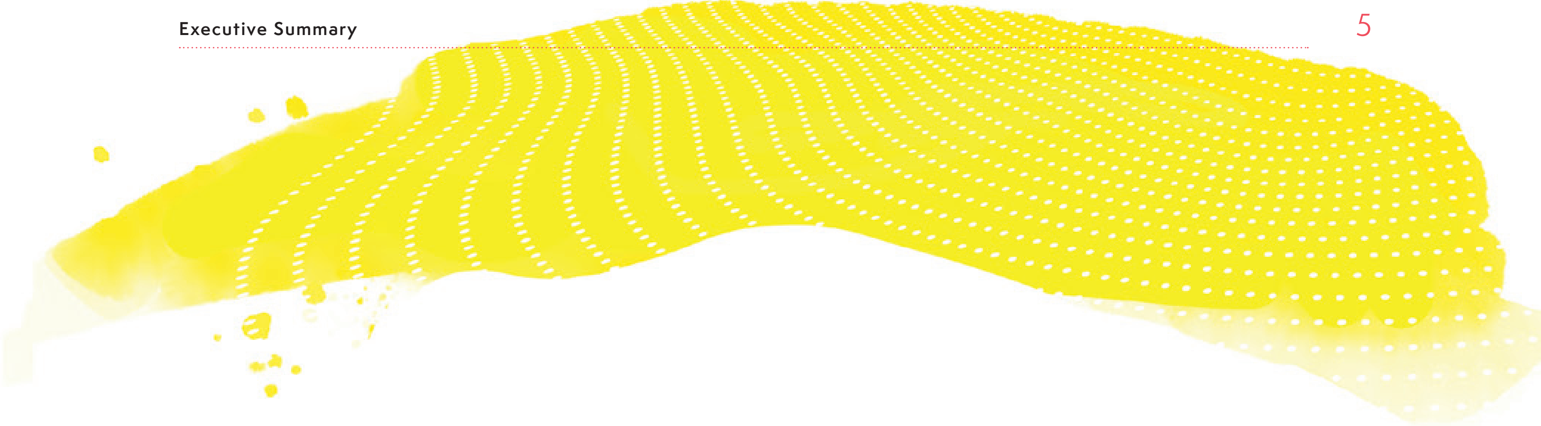
- **Financial involvement.** In-app currencies are often needed to improve the experience of so called "free-to-play" games exposing children to risks of monetary incentives for dangerous behavior.

These effects will be amplified in the future as the gaming industry – and especially the VR gaming industry – continues to grow at an accelerated pace. According to one estimate, the VR gaming market is projected to grow at a Compound Annual Growth Rate (CAGR) of 30% between 2021 and 2025, while the overall gaming market is likely to hit $279.4 billion by 2025 – nearly six times the size of the home entertainment streaming market.

Companies running and owning these platforms could do more to protect children from online sexual exploitation and abuse. This includes adopting a Safety by Design approach and developing product features that protect users. There are a number of safety solutions and ideas available. The most promising include: **(1) Age Assurance, (2) Parental Controls, (3) Reporting and Blocking, (4) AI-supported Content and User Moderation.** Many of these solutions rely on algorithms of different complexity to estimate a user's age, enable content moderation at scale or identify dangerous users.

Despite the availability of such solutions child safety largely remains an afterthought for most

[1] Bracket Foundation and Value for Good. "Artificial Intelligence – Combating Online Sexual Abuse of Children". 2019.



platforms. In fact, none of the most popular social gaming platforms have a comprehensive safety framework that includes all of the solutions currently available to social gaming platforms. This is due to:

- **Competitive disadvantage:** Companies fear being at a disadvantage to their competitors if they implement safety measures as they might increase sign-up hurdles or impact the user experience. Incentives typically stem from increasing sign-ups and monthly active users.
- **Liability:** The current legal landscape does little to impose significant legal or financial consequences on companies for neglecting child safety. Allowing children on the platform might actually increase legal liability.
- **Regulation:** The gaming industry has mostly escaped national regulations, forcing companies to implement many basic safety measures.

The largest lever to drive change in how the industry addresses child safety is the introduction of a comprehensive regulatory framework to ensure that social gaming platforms introduce standardized safety measures. Such regulations can force platforms to invest and prioritize child safety while creating a level playing field that does not punish responsible companies that do. All stakeholders involved, however, need to be engaged to drive change. Platforms need to proactively take responsibility for child safety, investors must leverage their power to channel the attention of fund-seeking companies towards the issue and parents need to be further educated on the risks of these new spaces. Meanwhile, the research community should provide insights into scale, scope and impact of child sexual exploitation and abuse in the context of emerging immersive technologies. Shared responsibility between users and platforms is paramount to usher in a new era of online safety for the most vulnerable: our children.

The insights in this report were derived through different research methodologies:

- **Aggregation of existing data and publications:** review and collection of existing published data and literature.
- **News reporting:** thorough review of news reports and documentaries on the topic of child sexual exploitation and abuse in gaming and the metaverse.
- **Expert interviews:** interviews with over 22 experts from the public and private sector, practitioners, academics and parents.
- **Internal investigations and analysis:** analysis and review of parental controls, company and platform policies, and in-game settings and safety features, as well as investigations of virtual spaces and sign-up processes.

Disclaimer:

This report has been prepared by Bracket Foundation and Value for Good in cooperation with the Centre for Artificial Intelligence and Robotics at the United Nations Interregional Crime and Justice Research Institute (UNICRI). The opinions, findings, conclusions and recommendations expressed herein do not necessarily reflect the views of the United Nations or any other national, regional or global entities involved. The designation employed and material presented in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area of its authorities, or concerning the delimitation of its frontiers or boundaries. Contents of this publication may be quoted or reproduced, provided that the source of information is acknowledged. The authors would like to receive a copy of the document in which this publication is used or quoted.

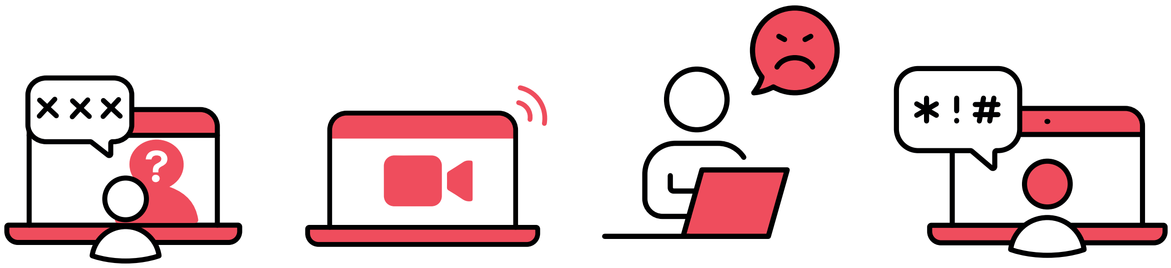
I. CHILDREN ARE INCREASINGLY THREATENED ONLINE

I. Online risks to children are increasingly acute

The rapid growth of digital technology has changed the way children interact with friends and strangers. It has allowed them to connect with peers, play games and learn. But digital technologies have also enabled children to interact with adults and have contributed to a surge in sexual exploitation and abuse of children online. The widespread adoption of new technologies has been met with little awareness or consideration of the risks posed to children.

Since the 2019 publication of the report entitled "Artificial Intelligence – Combating Online Child Sexual Abuse", there has been increased public debate about the risks for children but, fundamentally, the same patterns continue as new digital services (gaming and metaverse) and technologies (virtual reality headsets) are launched. The analysis presented in this report relies primarily on data drawn from the United States, Europe, the United Kingdom and Australia. It does not attempt to present a complete global picture but instead draws conclusions based on the data available.

FIGURE 1: **ONLINE RISKS TO CHILDREN ARE INCREASINGLY WIDESPREAD**[2]



Online grooming, including tactics such as catfishing & sextortion is on the rise – numbers of instances recorded by policy jumped by 70% in the last three years

CSAM (Child Sexual Abuse Material) continues its exponential growth – reports of child sexual abuse online have risen from 1.1 million in 2014 to 29.3 million in 2020 covering over 84 million CSAM images and videos

Self-generated images are increasing – webpages containing self-generated images increased by 168% from 68 thousand in 2020 to 182 thousand in 2021

Cyberbulling is widespread – a study found that 48% of children between nine and 17 are affected

New Gaming Platforms Combined with the Metaverse Are Making Children Online More Vulnerable.

[2] Pew Research Center. "Teen's Experiences with Online Harassment and Bullying, by Demographic Group". September 2018.

Online grooming is on the rise: Online grooming occurs when a person, often an adult, wrongfully gains the trust of a child online and then convinces the child to commit sexual acts. Data recorded by police in England and Wales show an increase in such crimes. The number of instances recorded by police jumped by about 70% from 2017/2018 to 2020/2021 to an all-time high of 5,441 Sexual Communication with a Child offences recorded in 2021.[3] Online grooming crimes disproportionately affect girls, with four in five victims being females aged 12-15.[4] Online groomers often employ tactics such as catfishing and sextortion.

Catfishing: Catfishing is a process of luring someone into a relationship by means of a fictional online persona. Adult predators may pose as a child to gain the trust of children. The FBI has seen a 20% increase in reports of catfishing from over 19,000 in 2019 to nearly 24,000 in 2020 with victims reporting losses of more than $600 million.[5] Meta alone removed more than 1.3 billion fake accounts from Facebook between October to December 2020.[6]

Sextortion: Sextortion refers to a form of extortion where children are blackmailed using self-generated sexting images to extort sexual favors including demands for more explicit imagery, under threat of sharing the images on social media or with family members. The UK's Revenge Porn Helpline reports cases of sextortion have increased by 89.5% from 2020 to 2021 reaching 1,124 reports.[7] A survey by Thorn in 2017 found that one in four victims of sextortion were 13 years or younger and two in three were female victims younger than 16.[8]

CSAM continues its exponential growth: CSAM (Child Sexual Abuse Material) is child pornography and any content that depicts sexually explicit activities involving a child. CSAM reports have also increased significantly in the last decade. Reports of child sexual exploitation and abuse online have risen from 1.1 million in 2014 to 29.3 million in 2020. These reports were made in the United States but may relate to content generated worldwide.[9]

Self-generated images: Self-generated images are images that have voluntarily or coercively been created by the child in it. These images are often known as "nudes" or "sexting" by the general public. A study by Thorn among children in the US aged between nine and 17 years in 2020 found that 17% of all minors have shared self-generated images before and 7% have re-shared someone else's images.[10] The Internet Watch Foundation found that the number of identified webpages containing self-generated images increased by 168% from 68 thousand in 2020 to 182 thousand in 2021.[11]

Cyberbullying: Cyberbullying is widespread and includes offensive name-calling, receiving unrequested explicit images, being threatened or having explicit images shared without consent. In the US, 59% of United States teens between the ages of 13 and 17 said they experienced harassment online which included offensive name-calling as well as receiving unrequested explicit images.[12]

[3] NSPCC. "Record High Number of Recorded Grooming Crimes Lead to Calls for Stronger Online Safety Legislation". August 2021.

[4] NSPCC. "New Figures Reveal Four in Five Victims of Online Grooming Crimes are Girls". October 2021.

[5] Federal Bureau of Investigation's Internet Crime Complaint Center. "Internet Crime Report 2020". March 2021.

[6] Guy Rosen. "How We're Tackling Misinformation Across Our Apps". Meta. March 2021.

[7] Akishah Rahman. "Sextortion Cases Reported to Revenge Porn Helpline Double in a Year". Sky News. May 2022.

[8] Thorn. "Sextortion". 2017.

[9] National Center for Missing & Exploited Children. "2014 Annual Report". 2014; National Center for Missing & Exploited Children. "CyberTipline 2021 Report". March 2022.

[10] Thorn. "Self-Generated Child Sexual Abuse Material: Youth Attitudes and Experiences in 2020". November 2021.

[11] Internet Watch Foundation. "IWF Annual Report 2021". April 2022.

[12] Pew Research Centre. A Majority of Teens Have Experienced Some Form of Cyberbullying | Pew Research Center. April 2018.

II. Gaming is changing and increasing the risks to children

Gaming has undergone a transformational evolution over the last decades: from closed platform experiences to online connected massive multiplayer games (MMPG), to the latest iteration: **social gaming platforms.** Social gaming platforms have become a place to hang out, connect with others, play games, consume content and buy or sell items. Playing a game is no longer essential to the experience of these platforms. In that respect, social gaming platforms are very similar to virtual reality spaces currently being marketed as "metaverses"[13] or metaverse gaming platforms. The features of these spaces and the types of experiences players and visitors are offered seem to converge. Social gaming platforms and metaverses have fundamentally changed how users interact with each other and their virtual environments.

Key features of social gaming platforms, as well as metaverses, are: (1) the player or user is embodied in a digital space through an **avatar**, (2) the player or user is able to interact **with both friends and strangers** (3) the interactions can be **multilayered** through the incorporation of multiple ways to socialize, e.g. through live communication audio chat or avatar gestures, (4) the experiences are **immersive** through high levels of engagement often, but not exclusively, through VR and AR headsets.[14]

Using the 3C framework developed by EU Kids Online[15], it was observed how social gaming platforms, and by extension the metaverse, increase the online risks for children. The 3C framework classifies risks to children in online environments into 3Cs: content risk, contact risk and conduct risk and further breaks the risks down along three dimensions: aggressive, sexual and values risks. This study will focus on sexual abuse risks but will address the other types of risk when appropriate.

Defining the metaverse

Emerging as a combination of various trends in the digital space, the term metaverse was coined by Neal Stephenson in his 1992 science-fiction novel "Snow Crash"*. Since then, the term has been defined in various ways and used in a variety of contexts ranging from literature to movies, including „The Matrix". Lately, the term has been pushed into the spotlight through Facebook's rebranding to Meta, highlighting highlighting the company's ambitions to create its own metaverse. Despite the metaverse's growing popularity and its history in the minds of futurists, it is still a relatively nascent concept. Indeed, some argue that there is no platform yet that can be described as such. It is expected, however, to be the next iteration of the internet: a live immersive network of platforms that in the future will mostly be VR or AR based, where users will be embodied in the space through an avatar, interact with other users and the environment in ways similar to real life and can engage in commercial (trade, earn money), recreational (play games, watch concerts, work out) and educational activities. Gaming, however, currently continues to be the most mature use case acting as a seed for the metaverse that other applications develop around. Social gaming platforms exemplify many of the characteristics of the metaverse and the analysis of this report draws on data available in this context.

** In the novel, people use digital avatars of themselves to explore the online world.*

[13] The term metaverse is used in the study to refer to platforms currently on the market and marketed as such (e.g., Horizon Worlds, Decentraland). This does not mean that they have already achieved the vision of the metaverse.

[14] Virtual reality (VR) describes a computer-generated 3D environment that surrounds a user and responds to an individual's actions usually through immersive head-mounted displays. Augmented reality (AR) is the real-time use of information in the form of text, graphics, audio and other virtual enhancements integrated with real-world objects. Gartner. „Information Technology Gartner Glossary". Retrieved June 2022.

[15] EU Kids Online Rep Cover_3_AD.indd (lse.ac.uk). 2009. (This was later developed into the 4C framework, for reference see Sonia Livingstone, Mariya Stoilove. "The 4Cs: Classifying Online Risk to Children". 2021.)

FIGURE 2: **RISKS TO CHILDREN IN ONLINE ENVIRONMENTS CAN BE CLASSIFIED INTO CONTENT, CONTACT AND CONDUCT RISKS**

	Content receiving mass-produced content	Contact participating in (adult-initiated) online activity	Conduct perpetrator or victim in peer-to-peer exchange
Aggressive	› Violent/gory content	› Harassment, stalking	› Bullying, hostile peer activity
Sexual	› Pornographic content and otherwise sexually explicit content	› Grooming, sexual abuse, and exploitation (blackmailing, catfishing, lured off platform)	› Sexual harassment, sexting, self-generating CSAM
Values	› Racist/hateful content, misinformation	› Ideological persuasion	› Other harmful user-generated content

Content risks are risks where a child is exposed to inappropriate or illegal content. This can include sexually explicit, pornographic, CSAM and violent or otherwise age-inappropriate images. It may also include certain racist or discriminatory material, misinformation or hate speech, as well as websites advocating harmful or dangerous behaviors, such as self-harm, suicide and anorexia.

Contact risks are where a child interacts with adults who seek inappropriate contact or solicit a child for sexual purposes. Other contact risks occur when a child engages with individuals expressing extremist views and who seek to persuade the child to take part in unhealthy or dangerous behaviors.

Conduct risks are where a child behaves such that the behavior contributes to risky content or contact. This may include children self-generating sexualized images, sharing sexual images of others without consent or creating hateful content about other children, inciting racist, violent or discriminatory material.

While many of the risks are not new, social gaming platforms, as well as the technical developments of the metaverse, significantly increase online risks to children along all aspects of the 3C framework:

(1) Immersive intensity and lifelike character of interactions make it easier to build relationships with minors, increasing contact risk

Typical grooming patterns involve building things together and gaining the child's trust. Building such a relationship is made easier when interactions are based on a shared common interest such as in a gaming context. Social gaming and metaverse platforms facilitate perpetrators developing grooming pathways where they can easily interact with children. The lifelike and immersive nature of engaging in the metaverse heightens the experience of personal social interactions, increasing the likelihood of developing personal and trusted relationships. This subsequently increases contact risks such as sexual grooming, including catfishing and blackmailing, as well as sexual exploitation and abuse, through the creation of self-generated CSAM and sextortion.

FIGURE 3: **THE CHARACTERISTICS OF SOCIAL GAMING AND METAVERSE PLATFORMS INCREASE THE RISKS CHILDREN FACE**

Characteristics	Content	Contact	Conduct
Immersive intensity/realness of interaction makes it easier to build relationships with minors	🟥		
Competitiveness of gaming increases the risk of contact with predators	🟥		
Not easy to see what children are doing from outside of VR and AR	🟥	🟥	
In-app currencies needed to improve experience	🟥		🟥
Multitude of interaction on social gaming and metaverse platforms difficult to moderate	🟥	🟥	🟥

The contact risk increasingly can spill over into the offline world with worrying repercussions. After initial contact with a child, perpetrators move conversations to private and often encrypted communication channels. This cross-platform movement can extend offline, with perpetrators and children meeting in real life – and in extreme cases – leading to kidnapping or rape.

(2) Anonymity and ease of interactions increases the risk of contact with predators

Anonymity and the ease of interaction on social gaming platforms encourage interactions with strangers, which increases the risk of contact with predators. Predators develop the mindset: "If they don't know me, they can't catch me." This mindset significantly lowers the barrier for perpetrators to engage in wrongful behavior. Whilst offline, children often play and engage with peers in the same age group (e.g., through age brackets at sports club), online this is not always the case. Online gaming is often highly competitive and children – incentivized to play their best – are likely to play with adults if they rank highly. Combined, these factors increase the risk of children coming in contact with predators and offers predators an easy route to connect with children.

(3) The many different actions in social gaming platforms are difficult to moderate, increasing content, contact and conduct risks

In social gaming platforms, users engage through their avatar using multiple sensory experiences, including gestures, voice and chat functions – creating multiple communication modes thereby complicating moderation. Users can also generate new rooms, environments and groups, further adding dimensions that need to be monitored and moderated. These multiple input feeds create a level of complexity that make content moderation in social gaming and metaverse platforms inherently more difficult than in traditional online environments. This increases the risk that children are exposed to age-inappropriate or harmful content, contact and conduct.

(4) It is difficult for parents to monitor what children are doing from the outside when they use headsets, increasing content and contact risks

The very nature of VR and AR means that there is no easy access for parents to be involved in the experiences of their children. Unlike a screen, where a parent or guardian can casually check on developments, a VR or AR headset makes it difficult for non-users to see what is going on and almost none of the interactions are stored. This poses a risk to children who engage in social gaming and metaverse platforms without guardian supervision and interact with adult strangers or watch inappropriate content.

(5) In-app currencies are needed to improve experience on the platforms, increasing contact and conduct risks

Free-to-play social gaming platforms mean there are often no costs associated with downloading the game. To continue to play, however, in-game purchases must be made, often using in-app currencies. This can include gaming add-ons, such as equipment or avatar enhancements. In these environments, children are incentivized to earn in-app currencies to make in-app purchases that improve their gaming performance or metaverse experiences. Predators use this as an opportunity to exchange monetary tokens in return for sexual favors – often by getting child players to create and share CSAM.[16]

III. The immersive experience of VR and AR increases the level of trauma experienced by children

The immersive experience of VR and AR headsets is likely to worsen the impact and trauma caused by the experience of abuse on these platforms. A study of biophilic environments[17] on adults showed that stress indicators such as heart rate and sweating reacted similarly to physical and immersive virtual exposure.[18] This suggests that virtual realities might illicit similar biological responses to real life experiences. As part of a different study,[19] VR users were shown to treat their avatars as if they are their real bodies, supporting this point. Experiences that take place in VR and AR might be influencing psychology, physiological reactions and perception in similar ways to offline experiences.

This ambiguity between what is real and what is taking place in VR or AR means that distinguishing reality from fantasy becomes more difficult, particularly for children. Research in the context of television demonstrates that children can clearly distinguish between reality and fantasy on TV from as young as five years old. However, immersive experiences might make this more challenging and the age at which children distinguish between reality and VR or AR may be much older. Studies, for example, show that elementary school children confused seeing a virtual doppelganger swimming with orca whales as seeing orca whales in real life.[20]

[16] Bowles, N and Keller, M (2019) "Video Games and Online Chats are 'Hunting Grounds' for Sexual Predators." (https://www.nytimes.com/interactive/2019/12/07/us/video-games-child-sex-abuse.html)

[17] Biophilic environments occur when natural elements are incorporated into indoor environments.

[18] Jie Yin et al. "Effects of biophilic indoor environment on stress and anxiety recovery: A between-subjects experiment in virtual reality". Environment International. December 2019.

[19] Lara Maister. "Changing bodies changes minds: owning another body affects social cognition". Trends in Cognitive Sciences. 2015.

[20] Jakki Bailey et al. "Immersive Virtual Reality and the Developing Child". Cognitive Development in Digital Contexts. 2018.

II. SOCIAL GAMING PLATFORMS ARE GROWING AND SO IS EVIDENCE OF THEIR HARM TO CHILDREN

I. A growing industry was supercharged by the COVID-19 pandemic

From 2016 to 2021 the gaming industry's global revenues grew by more than 50%[21] to reach $198.4 billion in 2021[22] – roughly four times the size of the global digital, home entertainment market.[23,24] While this exponential growth was supercharged by social distancing measures introduced during the COVID-19 pandemic, the trend is not likely to change soon. PricewaterhouseCoopers (PwC) predicts that total gaming revenue will grow by 9% annually to reach $279.4 billion by 2025[25].

Clearly, gaming has established itself as part of children's daily lives. In 2022, 71% of US children (<18 years) played video games for at least one hour per week; US-based players spend an average of 13 hours playing video games each week. In the United States, under-18-year-olds already accounted for 24% of all gamers in 2022 – up from 15% in 2017.[26,27]

This growth in revenue and users would be in line with other industries, but a sub-genre stands out: virtual reality gaming, of which social gaming and metaverse platforms form a huge part. It is expected to more than triple its revenue between 2021 and 2025. Sales of VR units are expected to grow by 500% between 2021 and 2026.[28]

Technological developments enabling the metaverse

Recent technological developments in hardware have enabled the production of cheaper and less complex VR and AR headsets, making them accessible to a wider, less tech-savvy audience. This development is based on two key drivers:

- **Faster chips:** Chips powering VR and AR headsets have gotten lighter while increasing their computing power. Today's systems thus increasingly work without external high-end gaming computers*.

- **Cheaper headsets:** The price of the leading VR headset – the Oculus VR line – ihad fallen by roughly two-thirds in only four years. The Oculus Rift was priced at $599 when it launched in 2016. Only two years later, the lighter Oculus Go was available for a mere $199. The newest Meta Quest is currently available starting at $299**.

Sustained demand for the headsets is expected to further increase investments and innovation in the field. Today most users still access social gaming and metaverse platforms through mobile phones and laptops and experience these spaces in 2D. The future, however, is expected to be dominated by VR and AR devices, 3D interactions and experiences that are characterized by their immersiveness, bringing us closer to Stephenson's version of the metaverse.

* *Logan Kugler. "The State of Virtual Reality Hardware". Communications of the ACM. February 2021*
** *Prices retrieved from the official Meta store in June 2022*

[21] PwC. "PwC Global Entertainment & Media Outlook 2021-2025". Retrieved June 2021.

[22] Mordor Intelligence. "Gaming Market – Growth, Trends, COVID-19 Impact, and Forecasts (2022-2027)".

[23] "Over-the-top" home entertainment describes all entertainment content that users access through the internet

[24] The global theatrical and home entertainment market achieved revenues of of $101 billion (Motion Picture Association. "2019 THEME Report". March 2020.)

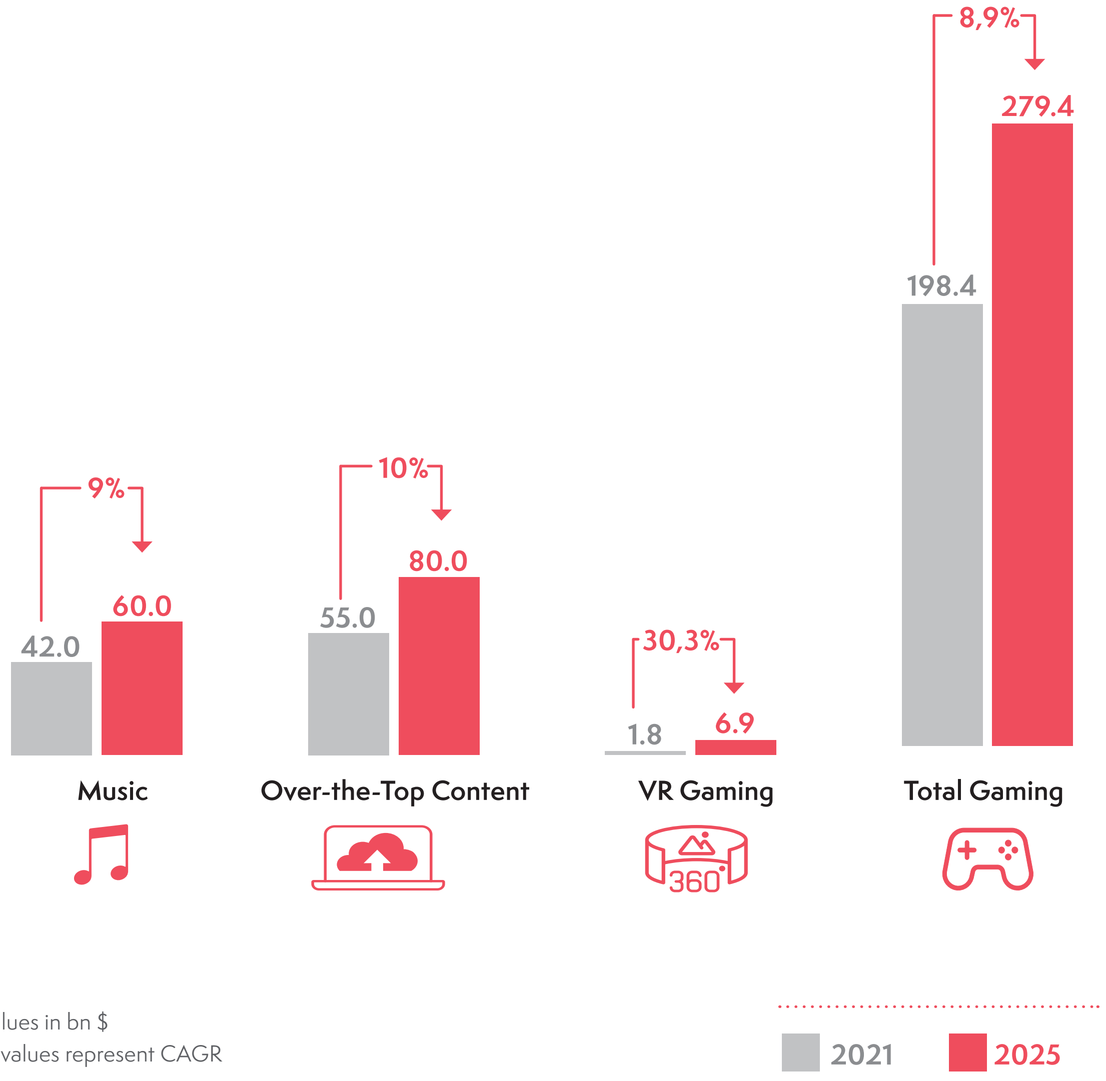
[25] PwC. "PwC Global Entertainment & Media Outlook 2021-2025".

[26] Entertainment Software Association. "2022 Essential Facts About the Video Gaming Industry". June 2022.

[27] For 2017 assuming a male female ration of 60/40 and 18% under 18 for males and 11% under 18 for females (Entertainment Software Association. "2017 Essential Facts About the Video Gaming Industry". April 2017.)

[28] IDC. "AR/VR Headset Shipments Grew Dramatically in 2021, Thanks Largely to Meta's Strong Quest 2 Volumes, with Growth Forecast to Continue, According to IDC". March 2022.

FIGURE 4: **SELECTED CURRENT AND ESTIMATED FUTURE REVENUES IN THE ENTERTAINMENT INDUSTRY**[29]

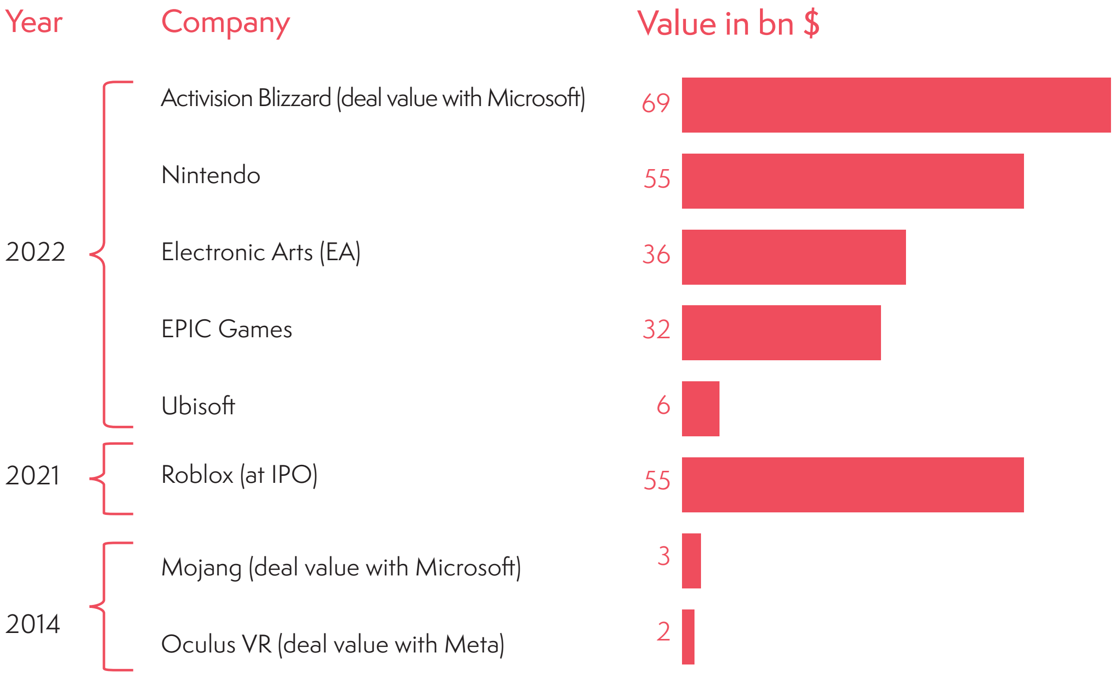


This surge in the gaming industry's relevance and the key role such virtual solutions as the metaverse play within this growth is also underlined by the recent developments of investments and valuations, as illustrated in Figure 5. As early as 2014, major tech players started to invest in (potential) VR platforms. In 2014, Meta (formerly Facebook) acquired Oculus VR for $2 billion[30]; investments in and high valuation of other platforms followed. This allocation of capital towards solutions, which can be interpreted as the foundation of what the metaverse might become, shows that businesses and investors expect a large portion of their revenue streams to shift into the realms of social gaming and metaverse platforms.

[29] Mordor Intelligence. "Gaming Market – Growth, Trends, COVID-19 Impact, and Forecasts (2022-2027)". Retrieved June 2020; PwC. "PwC Global Entertainment & Media Outlook 2021-2025". June 2021.

[30] Lucas Manfredi. "FTC Investigates Meta's Oculus VR Over Market Dominance". New York Post. January 2022

FIGURE 5: **VALUATIONS & INVESTMENTS IN SELECTED GAMING INDUSTRY PLAYERS**[31]



At this point, a comparison to the early days of social media provides interesting insights: in 2010, Facebook was at a similar stage in its lifecycle and generated approximately the same revenue as VR gaming generates today. Fast forward 11 years and Meta is generating revenues of $117.9 billion in 2021[32] – a CAGR of approximately 45%.

Another factor that has accelerated this trend is the COVID-19 pandemic. Widespread lockdowns prevented individuals from socializing offline and led to increased interaction of children on social gaming platforms.

This is illustrated by the increase in popularity of the three main social gaming platforms (Minecraft, Fortnite and Roblex) among minors during the pandemic. Ten- to 20-year-old players make up on average 38% of all gamers on those platforms but only 16% of all online gamers.[33]

For example, on Roblox, users under 16 account for as many as 67% of the total user base.[34] Children also make up a large percentage of users on adult platforms. While Fortnite is meant for adults over 18, it is widely accepted that minors play Fortnite and are likely to account for a large part of its user base. In fact, the actual total number of children on all these platforms is likely to be even higher than official numbers.

[31] Yahoo. "Market Capitalization of Largest Gaming Companies Worldwide as of May 2022 (in billion U.S. dollars)". May 2022; Ari Levy. "Microsoft Sets Record for Biggest Tech Deal Ever, Topping Dell-EMC Merger in 2016". CNBC. January 2022; PwC. "PwC Global Entertainment & Media Outlook 2021-2025". Retrieved: June 2021; Darrell Etherington. "Microsoft has Acquired Minecraft for $2.5 Billion". TechCrunch. September 2014; Lucas Manfredi. "FTC Investigates Meta's Oculus VR Over Market Dominance". New York Post. January 2022.

[32] Meta Platforms, Inc. "Annual Report 2021". February 2022.

[33] Niklas Melcher. "Deep Dive: Early Metaverse Players – Data on Demographics, Socializing, Playing, & Spending". January 2022.

[34] Roblox. "Roblox Corporation S-1 Filing". November 2020.

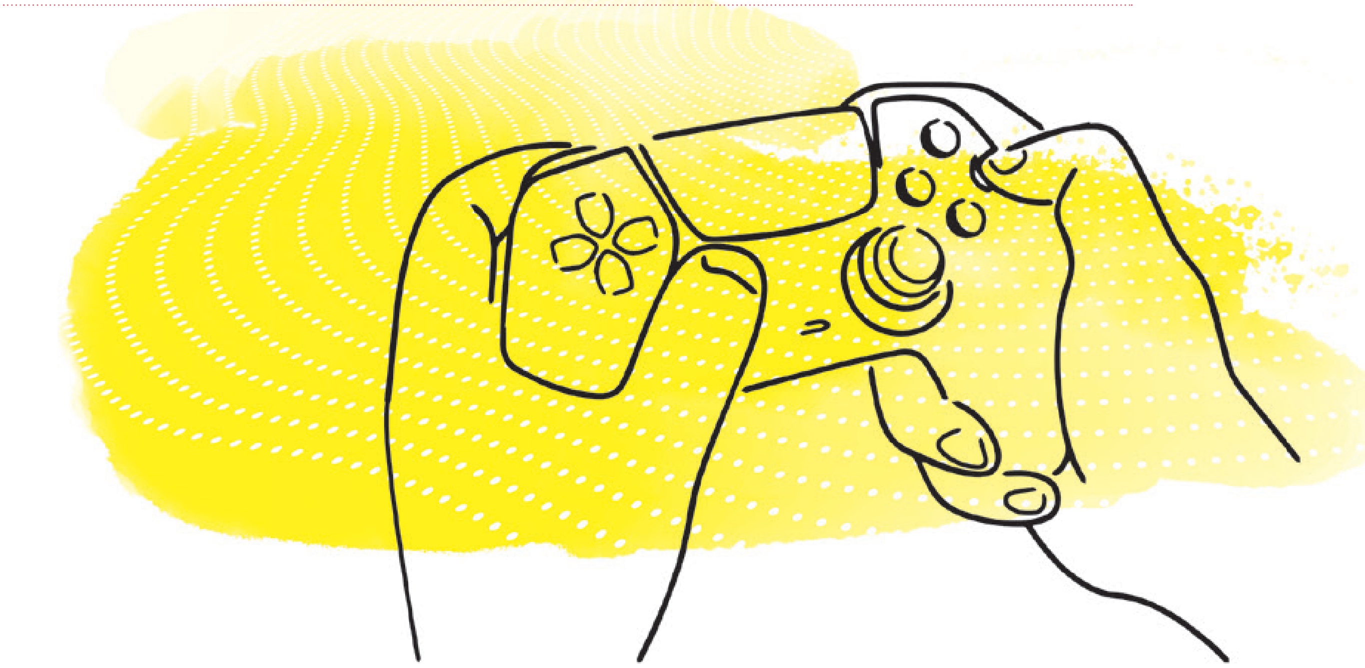
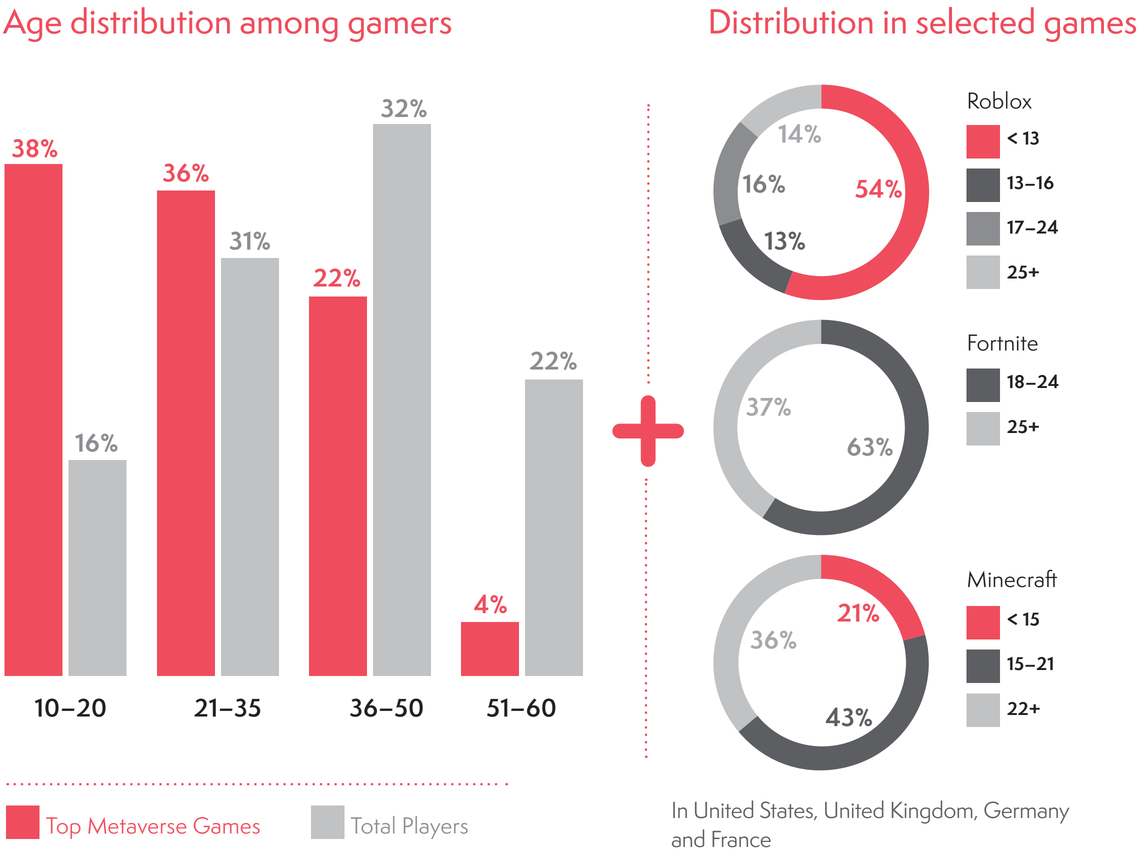


FIGURE 6: **MINORS ARE OVERREPRESENTED ON METAVERSE GAMING PLATFORMS BASED ON REPORTED NUMBERS, UNREPORTED EXPECTED TO BE HIGHER**[35]



Age distribution among gamers

Distribution in selected games

In United States, United Kingdom, Germany and France

Data from 2021

[35] Niklas Melcher. "Deep Dive: Early Metaverse Players – Data on Demographics, Socializing, Playing, & Spending". January 2022; Minecraft Seeds. "Minecraft Player Demographics". Retrieved June 2022; Roblox. "Distribution of Roblox Games Users Worldwide as of September 2020, by Age". Statista. November 2020; Vetro Analytics. "Distribution of Fortnite Players in the United States as of April 2018, by Age Group". Statista. May 2018.

FIGURE 7: ROBLOX AND FORTNITE ARE DESIGNED FOR AND ATTRACT THE MOST CHILDREN[36]

		Fortnite	Roblox	VR Chat	Minecraft
Types of activities	Play games	✓	✓	⊗	✓
	Create content	⊗	✓	✓	✓
	Interact with users	✓	✓	✓	✓
	Connect with users	✓	✓	✓	✓
	In-app purchases	✓	✓	⊗	✓
Recommended age		18+	13+	13+	10+
Monthly active users		83 mn	164 mn	0.021 mn	141 mn
Revenue (2020)		$ 5,100 mn	$ 923 mn	unknown	~$ 425 mn

II. Anecdotal evidence demonstrates severity of abuse on these platforms

While there is not enough robust quantitative data on the scale of abuse on these platforms, investigative journalists, parents and actors in law enforcement, have uncovered anecdotal evidence on the severity of abuse on social gaming and metaverse platforms. Reports that have made the headlines include examples of kidnapping, sexual grooming and abuse of children's avatars, with many incidents cited on platforms with the largest number of users: Minecraft and Roblox. In the United Kingdom, a user was jailed for two years and eight months for sexually grooming two children he met on a self-run Minecraft server;[37] he persuaded them to carry out sexual acts online and to send him intimate pictures of themselves, while he exposed himself to them online.[38] Incidents of spill-over effects into the offline world are also being reported. Earlier this year, a user in the United States was charged with sex trafficking, kidnapping and rape of a 13-year-old girl he met on Roblox.[39] This tragic incident came shortly after investigative reporting found that users on Roblox created sexually explicit rooms called "condo games" that allowed for inappropriate activity between children and adults.[40]

Other platforms also face issues with reports documenting that within minutes of using VRChat and Rec Room, an undercover reporter became surrounded by other users making sexually explicit comments.[41] The Center for Countering Digital Hate found that users, including minors, are exposed to abusive behavior every seven minutes on the platform VRChat.[42]

[36] Statista. "Gross Revenue Generated by Epic Games Worldwide from 2018 to 2025 (in Million U.S. Dollars)". May 2022; Statista. "Annual Revenue Generated by Roblox Worldwide from 2018 to 2021 (in Million U.S. Dollars)". February 2022; David Curry. "Minecraft Revenue and Usage Statistics (2022)". Business of Apps. May 2022; Paul Tassi. "Roblo's IPO Makes it Worth More than EA, Take-Two and Ubisoft". Forbes. March 2021; Steam Charts. "VRChat". Retrieved June 2022. VGChartz. "Number of Monthly Active Players of Minecraft Worldwide as of August 2021". Statista. October 2021.

[37] Minecraft servers are not run by Minecraft but run independently of Minecraft by users or businesses. Servers can be set up through software provided by Minecraft or through a hosting provider. The host of a server can establish user guidelines and enforce reporting and banning mechanisms. On those servers children can experience different virtual worlds, interact with users through the gameplay as well as chat with all users on the server through a common chat.

[38] BBC. "Minecraft Paedophile Adam Isaac Groomed Boys Online". January 2017.

[39] BBC. "Young Girl Returned After Kidnapping by Man she Met on Roblox". March 2022.

[40] BBC. "Roblox: The Children's Game with a Sex Problem". February 2022.

[41] Jamie Phillips. "Metaverse is Branded an ‚Online Wild West' by Child Safety Campaigners as Channel 4 Dispatches Uncovers Evidence of Sexual Abuse and Racism in the Virtual Reality World". April 2022.

[42] Center for Countering Digital Hate. "Facebook's Metaverse – One Incident of Abuse and Harassment Every 7 Minutes". December 2021.

Independent investigative work as part of this report has confirmed this evidence. Posing as a 16-year-old on VRChat, researchers were confronted with child-inappropriate content like swastikas on the walls within minutes.

The overall percentage and quantitative evidence of online grooming happening in games – and notably social gaming platforms – is still unclear as very little information is reported by companies. But data from Meta shows the worrying scale of online abuse: Meta acted on 83.1 million incidents of child sexual exploitation content[43] on Facebook and Instagram between April 2021 and March 2022 alone – representing over 230 thousand cases per day.[44]

III. Analysis points to increased scale of abuse and media's heightened interest

I. Number of reported incidents and media interest have increased between 2020 and 2022

For this report, a review of news articles from January 2020 to June 2022 was conducted; the review focused on 80 keywords and used the Google News platform.[45] The results show a clear increase in the number of media reports on abuse on social gaming and metaverse platforms per year. While in 2020 only six news reports were found, the number rose to 17 in 2021, with 46 reports having already been published in the first half of 2022 alone. Should this trend continue, 2022 will register 15 times more reports than 2020.

Investigative journalists uncover extent of abuse on VRChat[45]

In 2022, the sexual harassment of minors and their exposure to inappropriate content on virtual platforms were extensively documented by two journalists who investigated VRChat, one of the most popular apps within Meta's Oculus Quest Store.

BBC researcher Jess Sherwood pretended to be a 13-year-old girl while exploring virtual rooms in VRChat. As part of her investigation, she found and visited a virtual strip club where she witnessed adult men chasing a child while ordering them to remove their clothes. Condoms and sex toys were on display in many of the rooms Sherwood entered, on one occasion she even witnessed a group of adult males and adolescents performing group sex and over the course of her investigation, she observed multiple grooming incidences.*

The journalist Yinka Bokinni made very similar observations during her investigation for Channel 4 Dispatches, where she posed as a 22-year-old woman and a 13-year-old child. Within seconds of entering one of the virtual rooms in VRChat she was confronted with racist, sexist, homophobic and anti-Semitic comments by other users. Some of the comments were so extreme, that they could not be broadcast as part of the documentary.**

* *Angus Crawford et al. "Metaverse App Allows Kids into Virtual Strip Clubs". BBC News. February 2022*

** *Jonathan Rose et al. "Channel 4 Dispatches Shows Metaverse Users Boasting that they are Attracted to 'Little Girls Aged Between the Age of Nine and 12' and Joking About Rape and Racism in Virtual Reality Online". Daily Mail. April 2022.*

[43] Child sexual exploitation is widely defined as sexual abuse of a person below the age of 18 and may include grooming and trafficking.

[44] Transparency Center. "Child Sexual Exploitation, Abuse and Nudity". Meta Platforms, Inc. Retrieved June 2022.

[45] To compile the news articles for review, Google News was used. A variety of search terms were entered covering names of different metaverse platforms with child abuse terminology. The exact list of search terms can be obtained via inquiry to the authors. In addition, the time filter function was used. The following three time ranges were defined: 01.01.2020 to 12.31.2020, 01.01.2021 to 12.31.2021, 01.01.2022 to 16.06.2022.

Within these reports, Roblox and VRChat accounted for about 34% and 30% of incidents, respectively, while Meta's Horizon platforms accounted for 21% and Fortnite for 15%. Totaling the instances of times different incidents were mentioned within the analyzed reports (including multiple mentions of different incidents within one report), four incidents stood out. Together, they accounted for 52% of all instances an incident of abuse on the examined platforms was mentioned. The increase of media attention, as well as the repetitive documentation of the most prominent cases, signals an increase in public interest of the topic and how central the analyzed platforms are to this debate.

II. Tweets about sexual content on social gaming platforms have increased over the years

As part of this study, all tweets posted on Twitter between 2013 and mid-2022 containing the hashtags #robloxsex and #vrcnsfw were downloaded and analyzed. These hashtags are indicative of sexual content on Roblox and VRChat.[46] The acronym VRC stands for the platform VRChat and the acronym NSFW stands for "not safe for work," a term used in connection with child pornography. For both VRChat and Roblox the results show that cases have exponentially increased since the beginning of the COVID-19 pandemic. Between January 2021 and the first half of 2022, sexually explicit tweets about VRChat increased from 419 incidents in all of 2021 to 6,012 incidents in the first half of 2022. The most frequently used hashtags in those tweets were #vrchat, #nsfw, #vrclewd, #vrcerp and #vrchatnsfw. Sexually explicit tweets about content on Roblox, a platform primarily used by children, went from 304 in all of 2019 to 1,240 in all of 2020 to reach 9,797 tweets in the first half of 2022. All these tweets – some of which contain explicit and sometimes graphic pictures and videos of digital sexual acts – are freely accessible on Twitter.

FIGURE 8: NUMBER OF TWEETS WITH INCRIMINATING HASHTAGS[46]

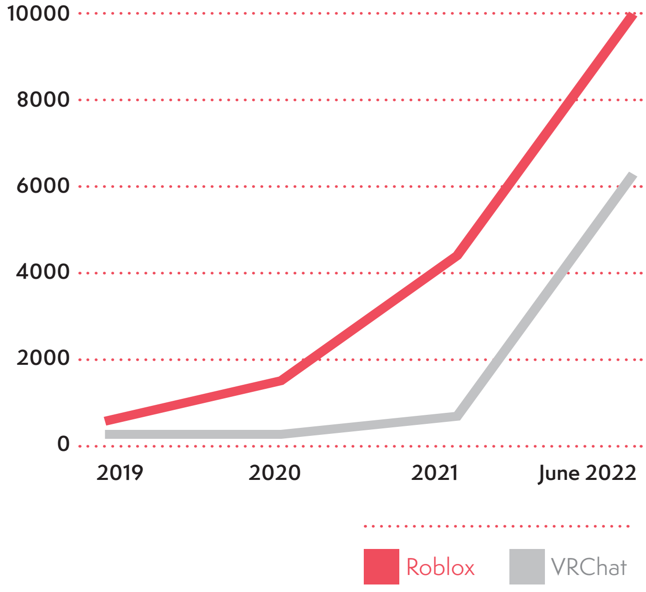
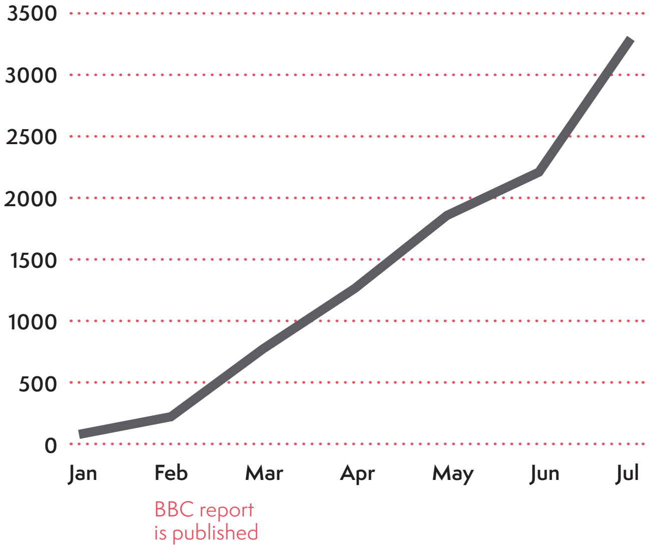
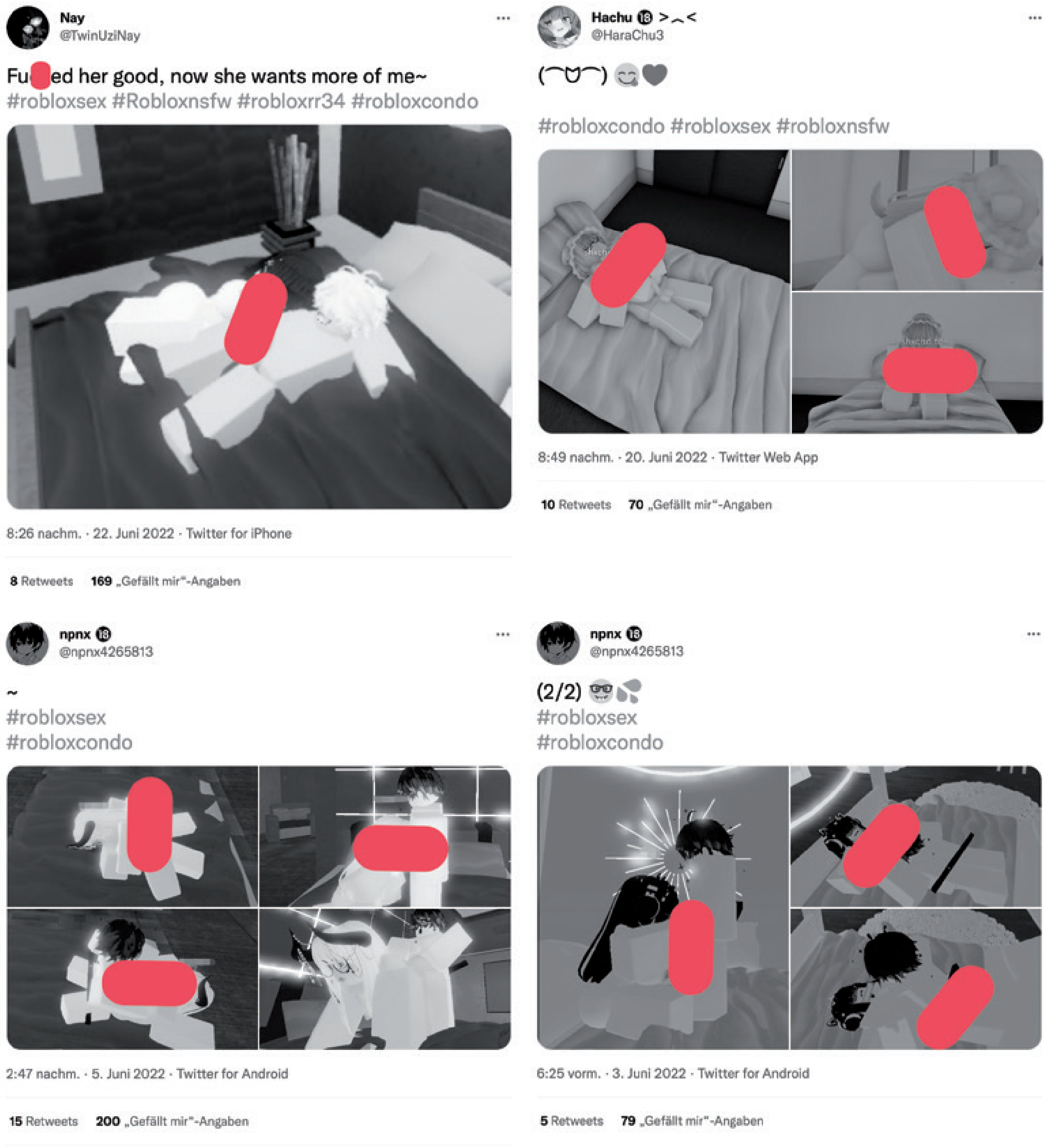


FIGURE 9: MONTHLY TWEETS CONTAINING #ROBLOXCONDO IN 2022[46]



[46] Using the official Twitter API via the twarc Python library (GitHub 2022), researchers (1) downloaded (i.e. scraped) a total of 22,819 tweets that contained the hashtags #robloxsex (16,836) and #vrcnsfw (6433). In a second and third step, researchers (2) distilled the five most frequent hashtags among these tweets and (3) analyzed the tweet volume distribution per year.

FIGURE 10: EXAMPLE TWEETS CONTAINING THE MOST FREQUENT HASHTAGS[48]



Nay
@TwinUziNay

Fu[redacted]ed her good, now she wants more of me~
#robloxsex #Robloxnsfw #robloxrr34 #robloxcondo

8:26 nachm. · 22. Juni 2022 · Twitter for iPhone

8 Retweets 169 „Gefällt mir"-Angaben

Hachu 🔞 >︿<
@HaraChu3

(⌒ᗜ⌒) 😋🖤

#robloxcondo #robloxsex #robloxnsfw

8:49 nachm. · 20. Juni 2022 · Twitter Web App

10 Retweets 70 „Gefällt mir"-Angaben

npnx 🔞
@npnx4265813

~
#robloxsex
#robloxcondo

2:47 nachm. · 5. Juni 2022 · Twitter for Android

15 Retweets 200 „Gefällt mir"-Angaben

npnx 🔞
@npnx4265813

(2/2) 🤓💦
#robloxsex
#robloxcondo

6:25 vorm. · 3. Juni 2022 · Twitter for Android

5 Retweets 79 „Gefällt mir"-Angaben

The investigative reporting by the BBC on "sex condos" on Roblox was in February 2022. The last tweet identified with the hashtag "robloxcondo" containing sexually explicit content was dated 28 June 2022. A detailed look at the number of tweets containing the hashtag "#robloxcondo" shows that since the BBC report in February 2022 the number of tweets increased drastically form 224 tweets at the time of publishing to 3,200 tweets in July 2022 – more than a 14-fold increase. This evidence clearly demonstrates the companies' inability or unwillingness to address known risks to minors on their platforms.[47]

[47] It has to be noted, that the BBC's reporting might have led to an increase in the number of identified tweets as a consequence of increased knowledge of the topic.

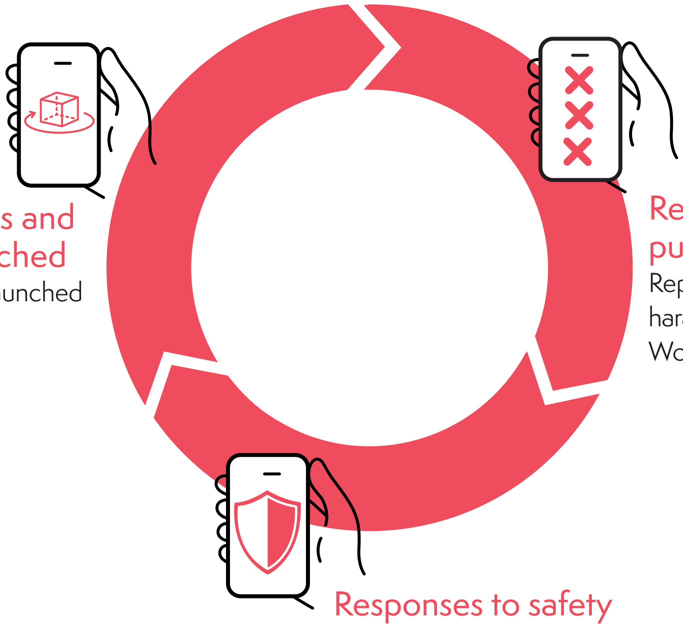
[48] Value for Good internal analysis of the frequency of tweets with incriminating hashtags on Twitter regarding Roblox and VRChat.

III. THE INDUSTRY FOCUSES ON GROWTH, NOT CHILD SAFETY

I. Company incentives

Currently implemented safety measures do not adequately protect children from potential risks on social gaming platforms and consistently fall short of the expectations of users, parents, experts and governments. While some companies invest in Trust and Safety departments and introduce safety features, child safety does not seem to be their highest priority. Social gaming platforms and metaverse companies' safety-feature development follows a reactive approach: (1) a new platform or game is launched that aims to attract as many users as possible, (2) abuse happens on their platforms or games and is uncovered by investigative journalism, class-actions and/or civil society, (3) companies react by introducing safety features and measures. A prominent example of this pattern is Meta's Horizon Venues and Worlds, where a four-foot personal boundary was introduced only after a beta tester reported being groped by other users.[49]

FIGURE 11: **HISTORICALLY, COMPANIES RESPOND TO SAFETY CONCERNS ONLY AFTER PUBLIC REPORTS OF ABUSE – MORE FIXES REQUIRED**[5]



New products and features launched

Horizon Worlds launched beta version

Reports of abuse published

Reports about first sexual harassment case on Horizon Worlds in days after launch

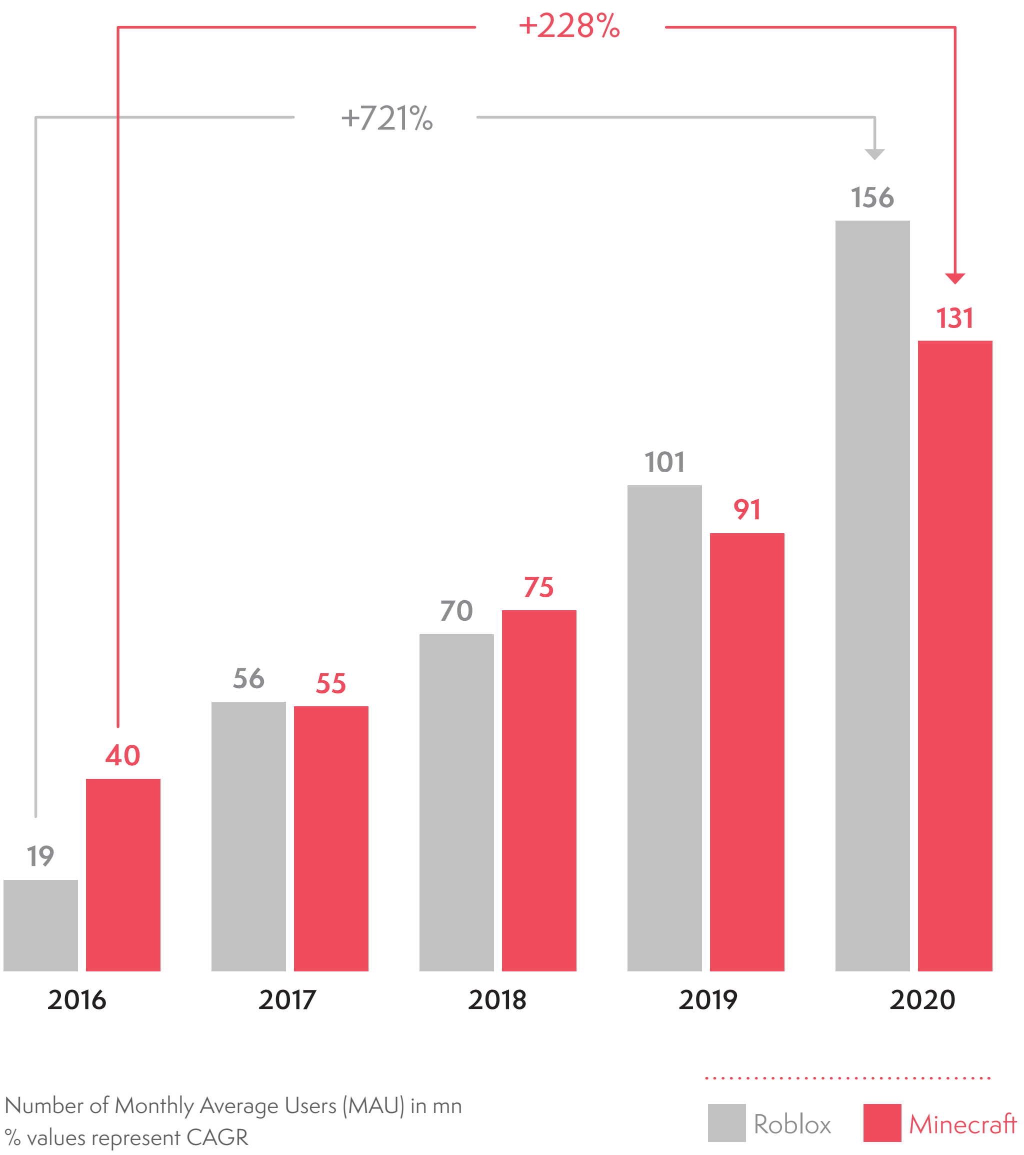
Responses to safety concerns

Horizon Worlds introduces a personal boundry, or bubble, feature

[49] Sam Tabahriti. "Meta is Putting a Stop to Virtual Groping in its Metaverse by Creating 4-Foot Safety Bubbles Around Avatars." Business Insider. February 2022.

[50] Vivek Sharma. Introducing a Personal Boundary for Horizon Worlds and Venues, Meta. February 2022.: "The roughly 4-foot distance between your avatar and others will remain on by default for non-friends, and now you'll be able to adjust your Personal Boundary from the Settings menu in Horizon Worlds."

FIGURE 12: **ROBLOX VS. MINECRAFT MAU DEVELOPMENT**[52]



Number of Monthly Average Users (MAU) in mn
% values represent CAGR

Often, companies designate themselves as only catering to adults to deny any responsibility for child abuse that might be happening on their platforms, while turning a blind eye to having children as part of their user base.

By and large, this pattern of neglecting child safety measures can be attributed to how the underlying business incentives for gaming platforms are structured.

[51] David Curry. "Minecraft Revenue and Usage Statistics (2022)". Business of Apps. May 2022; Audrey Schomer. "Roblox has Eclipsed Minecraft with 100 Million Users – Highlighting the Popularity of Digital Hangouts." Business Insider. August 2019. Brian Dean. "Roblox User and Growth Stats 2022." Backlinko. January 2022.

Reducing friction to grow the user base

Social gaming and the metaverse are new platforms, new activities and new technologies and, hence, often new businesses. Thus, as with any young business, the most important goal for these new platforms is to grow and gain traction within the market. To do so, they need to pull as many users to their platforms as possible. This creates the need to make the entry barriers to join the platform as low as possible, to guarantee a seamless sign-up proceses. If too many hurdles are added, users will perceive the cost of trying something with an uncertain level of gratification as too high and simply move on to somewhere else. Consequently, they keep the sign-up process as frictionless as possible, forgoing such frontend-screening measures as assurance, background parental consent requirements.

Companies are also eager to maximize users' time spent on their platforms to increase Monthly Average Revenue per Paying User (MARPU). Indeed, there is a strong correlation between MARPU and the average amount of time users spend on a platform. This pushes platforms to focus on user engagement instead of user safety. If competing with adults in games is more fun for children and will keep them engaged, this usually ends up being prioritized. In this competitive market, one company's decision to implement such safety measures, would put them at a competitive disadvantage. Safety measures can result in a loss of users and lower user engagement, as users search for an immediately gratifying experience.

Liability

Experts consider that many platforms prefer to stay ignorant to whether children are among their users, in order to shirk liability concerns. If they admitted to having children among their users, they would have to implement legally mandated privacy and safety measures, cumbersome to a growing business. Further, knowing the scope and scale of abuse would require that companies address the issues and perhaps make the platforms liable to parents. If the companies are unable to control the abuse easily, the necessary improvements may significantly reduce user numbers and engagement.

Lack of legal or financial ramifications

Given the current legislative landscape, most of these platforms do not have to concern themselves with legal or financial consequences for not doing more to protect children. While reputational damage may arise from major incidents and the subsequent media coverage, any financial effect due to a reduction in the user base is usually temporary. As discussed, following the several negative media reports on child safety, Roblox still managed to grow its user base.

II. Survey of implemented safety measures

I. Age verification and assurance

Most major social gaming and metaverse platforms (i.e. Roblox[52] Fortnite, Minecraft, VRChat, Horizon Worlds) do not have any official age assurance mechanisms. While Minecraft requires users to be at least 12 years old or be assisted by a parent or guardian to purchase the game, Roblox does not ask for the age of the user anywhere in the sign-up process. While Horizon Worlds officially requires its users to have an 18+ Facebook account, children seem to find an easy way around this requirement. Based on user reviews from the Oculus Store, the game appears to be full of underage players: of the 200 reviews analyzed, more than one in ten (12.6%) specifically mentioned the presence of children within Horizon Worlds.[53]

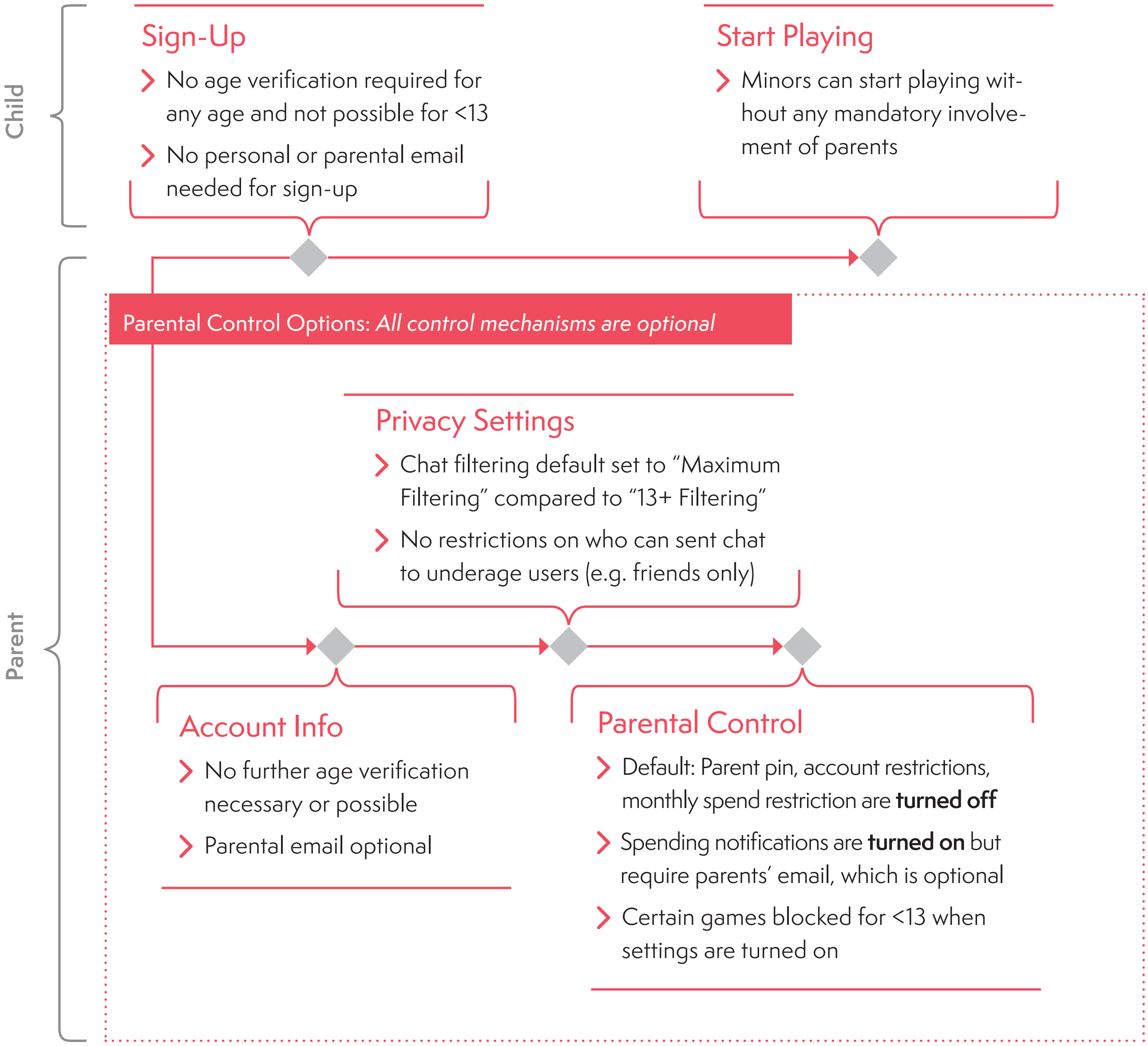
Other parts of the digital economy seem to be leading the way in emphasizing collecting age information on users. For example, users recently had to confirm their age on Instagram, and such platforms as Yubo have already implemented more sophisticated age estimation processes.

[52] Roblox offers a voluntary age assurance via ID verification, which is required to enable audio chat.

[53] Over 750 reviews of Horizon Worlds are posted on the Oculus Store. Researchers examined the 100 oldest and 100 most recent user experiences, covering a time span of almost two years from August 2020 to June 2022. The presence of minors in the game was mentioned in 25 of the 200 user reviews, or 12.6% of the analyzed comments. Most users, thereby, reported that the presence of children had an adverse effect on their gaming experience.

At Yubo, AI estimates the age by analyzing a real-time picture the user takes with their app as well as a short video for a liveliness check. The estimate is then matched against the age the user claims to be. If the distance between the estimated and stated age is too large, Yubo then forces users to go through an ID verification process. Solutions like those could soon find their way to social gaming platforms.

FIGURE 13: **DEEP DIVE ROBLOX: MINORS SIGN UP AND START PLAYING WITHOUT AGE VERIFICATION OR ANY PARENTAL CONTROLS**



Sign-Up

- No age verification required for any age and not possible for <13
- No personal or parental email needed for sign-up

Start Playing

- Minors can start playing without any mandatory involvement of parents

Parental Control Options: *All control mechanisms are optional*

Privacy Settings

- Chat filtering default set to "Maximum Filtering" compared to "13+ Filtering"
- No restrictions on who can sent chat to underage users (e.g. friends only)

Account Info

- No further age verification necessary or possible
- Parental email optional

Parental Control

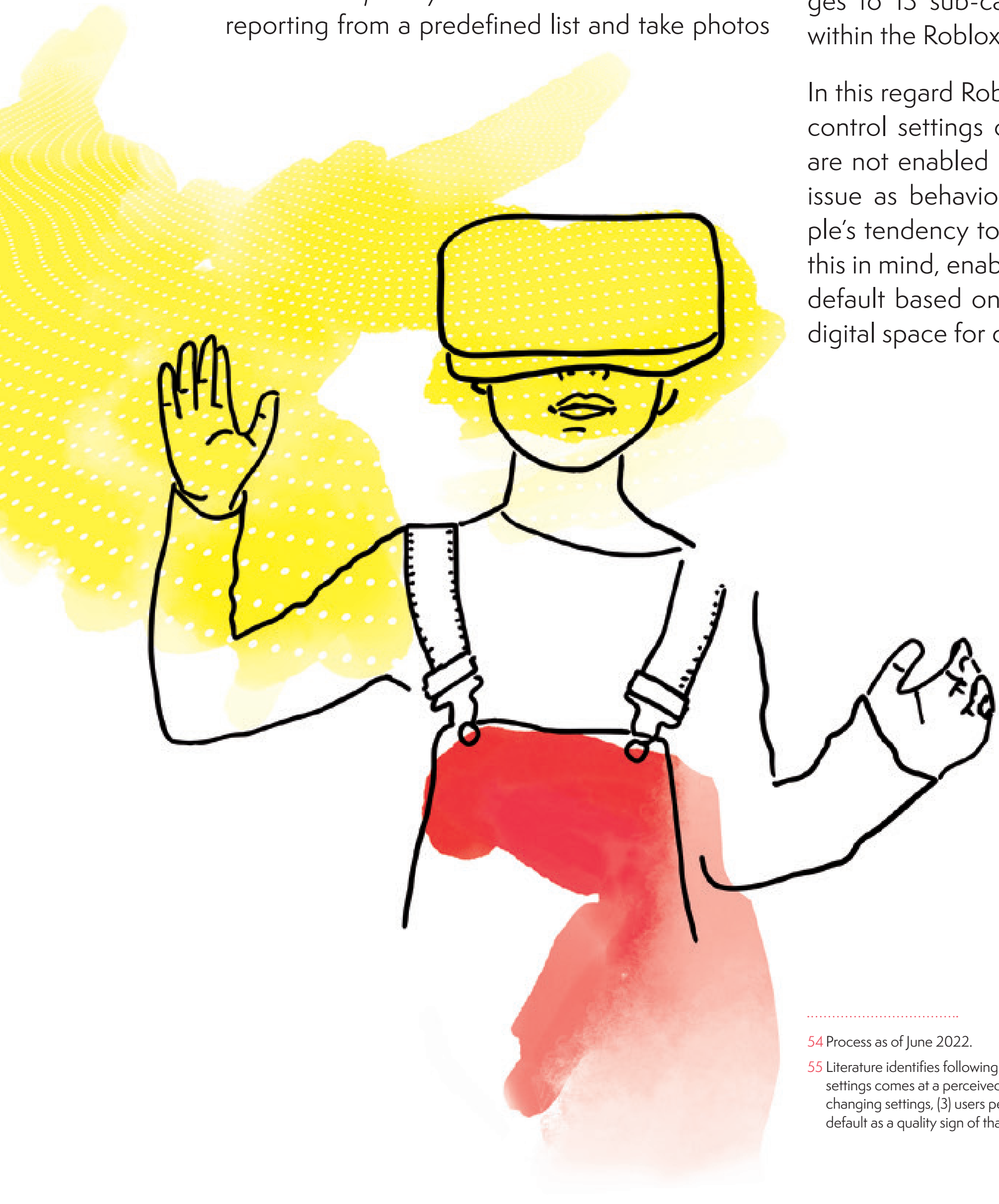
- Default: Parent pin, account restrictions, monthly spend restriction are **turned off**
- Spending notifications are **turned on** but require parents' email, which is optional
- Certain games blocked for <13 when settings are turned on

II. AI-powered content moderation

While not all the internal methods that social gaming and metaverse platforms use are publicly known, most major platforms report employing a combination of human and AI-powered moderation. Some minor platforms, however, rely on volunteers and creators to play the role of human moderator, relegating the responsibility to untrained and unverified users. On Minecraft servers, the host of a server, and not Minecraft is responsible for establishing user guidelines and enforcing them.

III. Moderating users

On the major social gaming platforms examined, reporting and blocking users was easy and straight forward. Users on VRChat, for example, can first choose what they want to report (e.g., user or item). They can then select a reason for reporting from a predefined list and take photos or videos of the issue encountered. Following that, they will be able to select the user they want to report – again from a predefined list. This same process is not always available on smaller, less popular platforms. Often a user can only report a specific communication and not overall user behavior. More importantly, on most platforms it is not always clear, what happens with these reports.



IV. Parental controls

While most platforms offer some form of parental control, they a) set the safety settings to the lowest levels for children by default, requiring parents to engage with and actively change them and b) are not easy to navigate. Roblox is a good case in point. If parents want to enable all the offered safety features, they will have to navigate through three different sections and make changes to 13 sub-categories within these sections within the Roblox settings.[54]

In this regard Roblox is not an outlier, as parental control settings on most devices and platforms are not enabled by default. This is especially an issue as behavioral science demonstrates people's tendency to stick to default settings.[55] With this in mind, enabling parental control settings by default based on user age would create a safer digital space for children.

[54] Process as of June 2022.

[55] Literature identifies following reasons for why people stick to defaults: (1) changing settings comes at a perceived "cognitive cost," (2) inertia keeps people from proactively changing settings, (3) users perceive a pre-made choice of en- or disabling a setting by default as a quality sign of that setting.

IV. COMPANIES HAVE MULTIPLE SOLUTIONS AND AVAILABLE APPROACHES TO INCREASE CHILD SAFETY

I. Solution landscape

Mapping the landscape of available solutions to protect children from abuse on social gaming platforms shows that there is no silver bullet. As social gaming platforms are newly developing spaces, as is the metaverse, there will be a continuous need to innovate and adapt when it comes to solutions to protect children. The planned interoperability between platforms on the metaverse particularly poses challenges, as

FIGURE 14: **EFFECTIVE SOLUTIONS REDUCE RISKS THROUGH PREVENTION, DETECTION AND PROSECUTION**

	Prevent	Detect	Prosecute
Content Risk	Age assurance Parental control Upload filters	Hashing Moderators Machine Learning classifiers (natural language processing , object detection, image recognition) Flagging Reporting	Banning users across platforms Creating single point of contact for law enforcement and partners Establishing frameworks and minimum standards for identified abuse with law enforcement Collecting contextual data Deprioritizing end-to-end chat encryption in communication with children Recording game play
Contact Risk	Age assurance Player matching Parental control	Age assurance ML: Groomer detection ML: Account analysis Blocking Reporting	
Conduct Risk	CSAM creation prevention system (image recognition) Smart keyboard	Moderators Behavioral analysis (sexual imitations, offensive hand signs etc.)	

it requires coordination and communication between stakeholders as well as solutions to stop inappropriate conduct across platforms. Based on experiences gained with "traditional" social media platforms and older forms of gaming, it is well established that a combination of solutions can go a long way to making online spaces safer for children. Effective solutions reduce content, contact and conduct risks, ideally before the event (prevent), contribute to the detection of the event if it occurs (detect) and enable prosecution of the offender after the event (prosecute). Figure 15 shows the different solutions along this spectrum, but protecting children in these spaces will only be possible by combining these solutions and continuously innovating to respond to new threats.

FIGURE 15: **ASSESSMENT OF POTENTIAL SOLUTIONS TO BE IMPLEMENTED FOR CHILD SAFETY**

	Potential Solution	Prevent	Detect	Prosecute
Safety by Design	**Product Design**			
	- Age assurance	✓	✓	⊗
	- Parental control	✓	✓	✓
	- Ease of reporting and blocking	⊗	✓	⊗
	Company Culture			
	- Institutional structure	✓	⊗	✓
	- Internal processes	✓	✓	✓
AI and Machine Learning	- Natural language processing (NLP)	✓	✓	⊗
	- Image recognition	✓	✓	⊗
	- Age assurance	✓	✓	⊗
	- Player Matching	✓	⊗	⊗
	- Groomer Detection	⊗	✓	✓
Transparency	- Reporting	⊗	⊗	✓

FIGURE 16: **A VARIETY OF SAFETY SOLUTIONS TO PROTECT CHILDREN ARE AVAILABLE TO PLATFORM PROVIDERS**

Tool	Description	Examples
Age assurance	Age assurance covers methods of age estimation (AI estimates age based on picture/video or behavior) and age verification (age determined e.g. through ID or credit card) Provides the basis for enforcing age-based rules and guidelines	YOTI euConsent Gamersafer
AI Classifiers	Algorithmic classifiers are used to identify abusive content and behavior text, speech, pictures and videos Allows for scaling up content moderation on platforms with many users and interactions	Thorn Safer Cease.ai Hive Appen
Moderating Users	User actions can be moderated in an effort to reduce abusive behavior Reporting and blocking other users are common tools of moderation New methods focus on matching players and detecting suspicious grooming behavior	Most social gaming platforms
Parental Controls	Umbrella term for a variety of applications and settings that allow parents to monitor and limit their children's digital activities Can be installed by the parents on the device or included as part of the application the children are accessing	Bark Net Nanny Qustodio Famisafe

Safety by Design is increasingly being hailed as an approach that can enable placing safety at the center of the design and development process of digital experiences. Safety by Design's goal is to minimize online threats by anticipating, detecting and eliminating harm before it occurs. This requires both changes to product design but also to company culture and structures that enable companies to prioritize child safety and to improve on their products. Government policies and guidelines can also enable the implementation of Safety by Design. One good practice example is Australia's eSafety Commissioner's Safety by Design approach that provides platform developers with safety principles for the design and development process as well as resources to assess and improve their own child safety approach.[56]

Safety by Design can be enabled through solutions that leverage technology for good – in particular AI-based tools – and can go a long way towards reducing risks to children on social gaming platforms.

[56] eSafety Commissioner. "Safety by Design." Retrieved June 2022.

euCONSENT

euCONSENT is an EU-funded project that aims to establish a pan-European, open-system, secure and certified interoperable age verification service for web-based apps. Users would demonstrate their age once towards a central databank where the information regarding their age will be stored as a hashed key. When visiting a website or web-based app with an age restriction, they would share a token with the website. The website can then verify the user's age using the token and the central database without the user having to disclose their identity.

II. Deep dive: Age assurance

Many of the risks that children face online stem from the anonymity of users on digital platforms. This leads to two main problems: first, children mingle freely with users of all ages while often being unaware of the age gap between themselves and the people they are interacting with. Second, minors can have unrestricted access to content that is inappropriate for their age, such as pornographic or violent content.

Age assurance solutions can help protect children from these harms by providing platforms with the information they need to enforce age-based rules and guidelines. These solutions should also be used to inform adolescent players (>18) – and in the case of minors (>13), their parents – about the ages of the other players. Currently, two main methods can be used to assess a user's age: age verification and age estimation. Age verification describes procedures that validate the user's age based on official documents (such as a government issued ID or credit card information) or through a biometric face scan combined with picture matching.

Age estimation describes tools that use AI to assess the likely age of a user based on an analysis of a user's photo, video, audio, text and/or behavior on the platform. While the accuracy of different methods varies, facial age estimation technology is becoming increasingly sophisticated and is showing promising results. One study found that a tool to analyze users' selfies to predict their age had a mean average error rate of only 1.3 years for six- to 12-year-olds and 1.55 years for 13- to 19-year-olds.[57] This technology is working so well that YOTI, a company offering age-assurance technology, was approved in 2022 by German regulators (KJM)[58] to assess the age of viewers of 18+ content. Another company, Privately, has also developed an age estimation tool based on voice, requiring users to read a text aloud. Tests have shown that the tool can differentiate between below and above 13-year-olds with a high level of accuracy and no user below 13, and clearly a minor, was classified as being 18 or older.[59]

Finally, other behavioral age estimation tools estimate age by analyzing user conduct, including the content the user consumes, the users they befriend and how they communicate and write. In the future, behavioral age estimation could also incorporate information from VR and AR headsets on eye and hand movement. At present, little is publicly known as to how accurate these models are.

The main advantage of age assurance tools is that they prevent children from getting into situations where they might be exposed to harmful content and, thus, help to reduce the total number of incidents. Moreover, as will be addressed in detail in the following sections, by ensuring that the age of every user on a given platform is known to the platform, age assurance technologies can serve as the foundation of many other child protection solutions, which rely on age as a prerequisite to other safety measures.

[57] Frank Hersey. "Global Movement Coalescing Around Age Verification and its Role in Online Safety". Biometric Update. March 2022.

[58] https://www.kjm-online.de

[59] Amy Colville. "Yoti Age Estimation Approved by German Regulator KJM for the Highest Level of Age Assurance Covering 18+ Adult Content". Yoti. May 2022.

In summary, identifying a user's age allows platforms to (1) limit access to adult-only content, rooms or spaces in virtual worlds, (2) create age-tailored experiences for adults and children, (3) inform users and their parents about the ages of their counterparts, (4) detect potential groomers by identifying adults who suspiciously seek out minors, and (5) stop pretending they do not have children on their platforms.

Limitation

Designing an age-assurance system centered around child protection that respects rights and is implementable faces a variety of challenges: (1) Privacy: Age assurance tools deal with sensitive personal data. Hence, age assurance solutions not only have to be compliant with such applicable privacy regulations as the EU's General Data Protection Regulation (GDPR), but they also raise questions about how the identity of users can be protected despite having to disclose their age. (2) Interoperability: An interoperable age-assurance system that sets out to enforce legal regulations is dependent on a large digital infrastructure and and robust implementation frameworks.

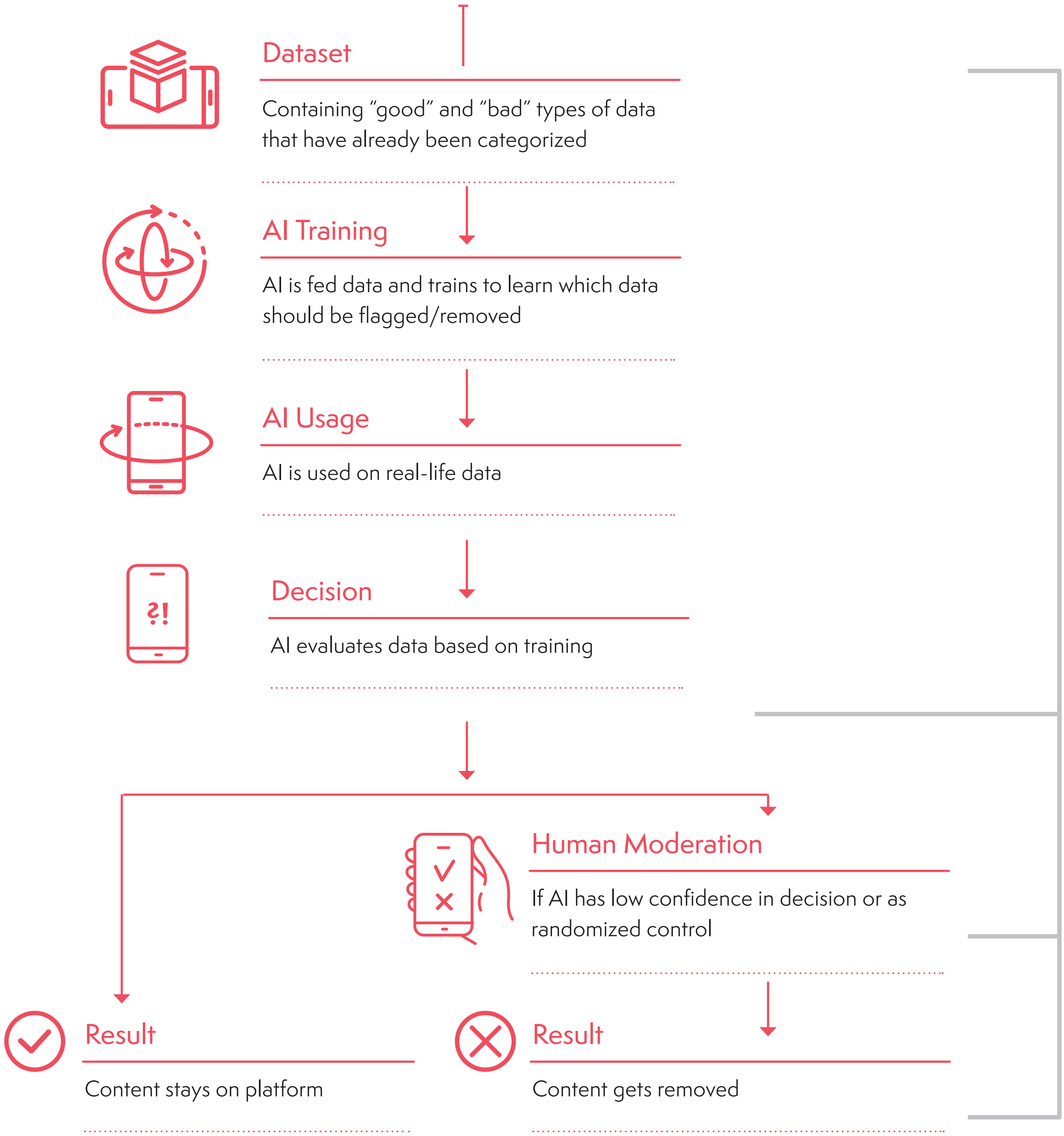
Apart from the technical and data concerns, the issue of the children's rights to freely access and leverage the digital world needs to be addressed. Through age-assurance technology, children can be restricted from entering certain platforms and interacting with certain issues. But just as children are not confined to children-only environments in the real world, the same should hold true for most digital spaces. In the end, locking children out of most platforms will not stop them from accessing content they are interested in. Many will either find a way to circumvent entry barriers or simply move to another platform, thereby, passing the problem on instead of stopping it. Rather, society has to ensure that the digital space as a whole is as safe as possible for children to participate, without risking being harmed. Only certain adult-only areas should be restricted from them – just like a nightclub or casino in the offline world.

III. Deep dive: AI-powered content moderation

Content moderation has its origins in the context of large social media platforms such as Facebook, Instagram and YouTube. Faced with the scale of content that requires moderation, they invested in AI tools to support their content moderation teams. A combination of simple content filters, AI algorithms and classifiers as well as human moderators are employed to address the issue of large-scale offensive content and behavior. The process of developing AI-assisted content moderation is described in Figure 17 and specifics related to text, audio and image moderation solutions are detailed below. On social gaming platforms, the combination of content streams means that moderation has to be able to simultaneously cover text-based content (chats), audio content (live and recorded conversations), video content and avatars' gestures, in order to determine whether a user's content or behavior violates a platform's terms. Technological innovation in effective multi-feed content moderation will be increasingly important going forward.

The underlying process of how AI-powered content monitoring works is generally the same for all the technologies touched upon in this chapter and is illustrated in Figure 17. For a deep dive into different types of AI, the first study "Artificial Intelligence – Combating Online Sexual Abuse of Children" provides an in-depth discussion of supervised and unsupervised learning.

FIGURE 17: **PROCESS OF DEVELOPING AND APPLYING AI-ASSISTED CONTENT MODERATION**



Dataset

Containing "good" and "bad" types of data that have already been categorized

AI Training

AI is fed data and trains to learn which data should be flagged/removed

AI Usage

AI is used on real-life data

Decision

AI evaluates data based on training

Human Moderation

If AI has low confidence in decision or as randomized control

Result

Content stays on platform

Result

Content gets removed

Text

Natural language processing (NLP) is the most commonly used AI tool to classify text. A machine learning algorithm is used to learn to read and understand written language. NLP can be used to identify and block particular words, phrases or expressions, such as expressions of hate speech, swear words or sexually explicit language. Not all harmful conversations are easily identified through explicit language, however. Many harmful messages are harmful through the combination of context, tone, audience, language nuance, subtlety, sarcasm or other cultural meaning. These are factors that NLP is not able to detect and thus limits its ability to identify abusive content. Another limitation of NLP is that perpetrators learn which terms AI can identify and they find ways to circumvent AI's monitoring by misspelling words or using code words. Further, most NLP is trained in English and popular Western languages making it harder to protect speakers of non-Western languages.

Contextual AI seeks to counter these issues. Contextual AI can look at broader patterns of words and phrases and consider contextual factors of message, such as who the sender and the intended audience are. This helps to more accurately identify communication that might contain harassment, grooming activities or hate speech and is seen as a promising solution for content moderation on social gaming platforms.

Audio

Increasingly, communication on online platforms is moving towards audio chat. To provide a safe environment for children, content moderation must extend to voice chats to protect children from offensive language and grooming patterns.

Voice transcription technology is starting to be used to assist human moderators with audio content, whereby voice is transferred to chat and then fed through an NLP model. This method, however, has some limitations:

(1) The lack of context recognition remains, regardless of whether the communication.

(2) Speech has more nuances than written word, including tone of voice, accents, slang, volume and tempo. Identifying potentially offensive language becomes even trickier.

(3) To be able to monitor audio, the game play or voice chat consistently needs to be recorded and stored for analysis.

(4) Due to the necessity of performing two steps for the analysis (transforming audio to text and analyzing audio) and of recording and storing the data, much more data processing and storage capacity is needed, leading to higher costs. To put it into perspective: A 10-second audio file is approximately the same size as 1,600 chat messages.[60]

Due to these difficulties, audio content monitoring is even more difficult to implement than text monitoring and is still very rarely used by social gaming and metaverse platforms.

Image

Image recognition software can be and is used to identify violent, sexual and other age-inappropriate content on platforms. It can be particularly relied on to identify CSAM. In addition to hashing[61] of known CSAM, algorithms can be trained to recognize nudity and children on images and videos. Once recognized, the content can be flagged. Currently, object detection has proven to be more effective at spotting harmful content than NLP has for spotting harmul text or audio transcripts transcripts, as images are less context-dependent and programs can rely more significantly on recognizing objects. Nevertheless, the same limitations regarding live action that apply to audio also apply in the context of visual and, especially, video content.

[60] Lee Davis. "Best Practices for Voice Chat Moderation". Spectrum Labs. June 2020.

[61] In this process an algorithm selects frames from images and videos to create a unique digital signature called a hash. These hashes are then compared against existing hashes in a database with images classified as CSAM. If there is a match with a hash in the database, the content will then be removed from the platform.

Advantages of AI-powered content moderation

The main advantage that AI can provide to content moderation lies in its speed and scalability. AI tools can work through a much larger amount of data in a much shorter period of time than human moderators ever could. The speed of AI is necessary necessary given the amounts of data that platforms have to process. Additionally, by delegating repetitive tasks to AI, human moderators can more effectively use their time to classify previously unknown content but also to identify new trends and patterns in abusive behavior and content.

Limitations of AI-powered content moderation

The most fundamental limitation of current AI solutions is that they will always only be as good as the dataset they have been trained on. Consequently, they will always be reactive by nature and unable to spot new trends, even as they are exposed to evolving methods from perpetrators. As AI solutions perform best when they are trained with data from the domain they are meant to protect, off-the-shelf solutions are often infeasible, and new platforms will struggle to implement effective AI solutions from the get-go. Lastly, relying on datasets means that AI is subject to biases embedded in the dataset, leading to potential discrimination against certain groups. This is further intensified by its tendency to struggle with nuances, whether irony or accents, leading to further discrimination against groups, who may be more likely to speak in such nuanced ways.

IV. Deep dive: Moderating users instead of content

Content moderation on social gaming platforms requires platforms to simultaneously monitor multiple actions including what the avatar is doing, saying or writing in the chat, as well as the content that is being uploaded. One solution to this challenge is to focus on moderating the users instead of the content. On social gaming platforms, companies are able to collect significant amounts of data on user behavior. These data include the content the user consumes, the users they befriend, how they communicate and write with different users and the "rooms" they seek –to name a few of the data sources. These data can be used to identify abusers based on their behavior and design children's interactions to reduce contact and conduct risks. Simple blocking and reporting have been around for years, but newer AI-based tools like matching and groomer-detection algorithms are examples of more sophisticated approaches.

Reporting and blocking users

Reporting and blocking allows children and other conscientious users to flag potential abusers to the platforms and to block them from interacting with them or others in the future. The process for reporting and blocking usually consists of the concerned user (1) selecting the type of issue that should be reported from a drop-down menu, (2) selecting the reason why it is offensive and (3) providing additional information. The report is reviewed by the platform and decision is made as to whether the reported player will be banned or blocked and for how long.

However, there are some drawbacks to reporting and blocking:

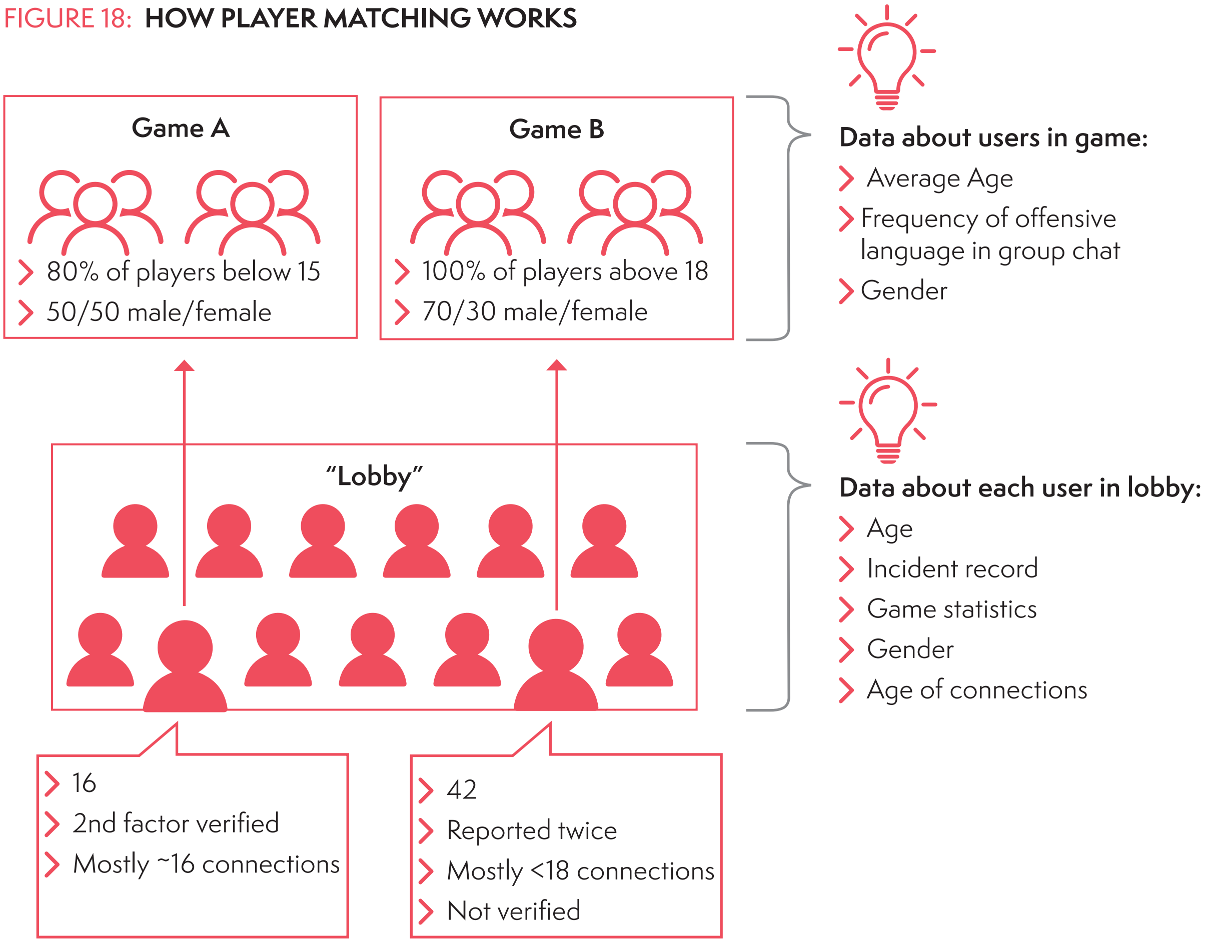
- Reported and banned users can easily re-enter platforms and circumvent a ban by creating new accounts.

- Often no action is undertaken by platforms unless a significant number of reports are received on the same issue. Individual incidents are often not sufficiently investigated.
- Reporting and blocking tools can also be used maliciously as a cyberbullying tool by repeatedly targeting specific users.
- Children tend to block and not report to continue enjoying their game without having to fill out report forms.[62]
- Reporting and blocking are reactive tools meaning that they will always only be used after the incident has happened and the damage has been done. To properly protect children online, however, proactive tools are needed that prevent incidents from occurring in the first place.

Matching

Matching describes the process of making deliberate choices regarding the users who are chosen to participate together in a given collaborative or competitive game. Instead of randomly assigning users to the different games or "rooms," an algorithm can be trained to create matches where the combination of users minimizes the risk of child abuse. Depending on the richness of data available, the algorithm could end up simply matching users in similar age groups or incorporating factors such as, skill level, past track record of adherence to platform guidelines, chat contents and any other relevant data the platform is legally allowed to use for the training. Matching particularly reduces contact risk by predicting which matches might put children in potentially dangerous situations and avoiding them

FIGURE 18: **HOW PLAYER MATCHING WORKS**



[62] Ofcom. "Children and Parents: Media Use and Attitudes Report 2020/21". April 2021.

Groomer-detection

Groomer-detection algorithms or prediction models are used to identify potentially suspicious behavior by adults on platforms. Potential data points can be adults who interact often with children or data on how an adult behaves around adults and children. Further data can be reports by other users or the exchange of in-app currencies between an adult and a child. Crucial for effectively detecting groomers and preventing them from contacting children is cross-platform collaboration. The Tech Coalition, an alliance of global tech companies working to combat child sexual exploitation and abuse online, is planning a pilot to bring tech platforms together to train and specialize a grooming detection algorithm. Pooling knowledge and data improves the accuracy of such tools and provides insights to the broader industry around grooming trends in varying contexts.

Limitations and issues of matching and groomer detection

Both matching algorithms and groomer detection models have the same set of limitations: (1) Privacy and data concerns: For effective matching and groomer detection an extensive dataset has to be collected for each account in order to create a profile of each user on which the assessment can be based. This naturally raises questions about users' data privacy. (2) Biases and risk of profiling: Relying on algorithms to identify predators and ban them or exclude certain players from gaming matches could easily lead to biases and profiling of minorities. Hence, these models would need to be independently evaluated for biases and profiling potential. (3) Data availability: As with content moderation, these algorithms are only as good as the data they are trained on. Thus, data availability and quality will be a challenge, particularly for new and young platforms.

V. Deep dive: Parental controls

One of the most prominent tools in the discussion around child safety is parental controls, an umbrella term used for a variety of different applications and settings that allow parents to monitor and limit their children's digital activities. These tools can be implemented through three different access points:

Operating systems

Parental controls can be integrated as built-in features within the operating systems of devices used by children, be it a VR or an AR headset, a mobile phone or a tablet. Across Apple, Android and Microsoft's operating systems, the overall level of control parents are given is usually relatively similar despite differences in naming and structuring of the parental control tools. For example, Apple offers parental control settings on all their devices (e.g., iPads, iPhones, iWatches), in the iTunes and AppStore as well as in the Game Center. Parents can protect their children by restricting downloads to certain apps, preventing in-app purchases, limiting access to explicit content such as movies or TV shows or restricting multiplayer and interaction features in the Game Center. In the second quarter of 2022 Oculus VR has started to roll out similar features on their VR devices, including parents' ability to block apps, request "Ask to Buy" notifications for purchases and downloads, view screentime and friends lists, and monitor the games used by their child.

Social gaming and metaverse platforms

Many platforms or games include parental control features within their settings. The platforms offer parents the option to enable settings that can limit children's ability to engage in certain age-inappropriate experiences, chat or talk to strangers or set monthly spending limits.

Third-party applications

There are many third-party providers of parental control apps and services that can be accessed by installing software on the device of the child (e.g., Bark, Qustodio, Net Nanny, SafeToNet). These services vary in their degree of intrusiveness but mostly offer tools such as screentime limits, blocking of specific games or websites and tracking the child's location. Some solutions go as far as allowing parents to read and restrict chats, listen to recordings of children's phone calls or go through pictures on the device. It has to be noted, however, that such a high level of intrusion is questionable from the perspective of children's rights. A recent focus of many apps is preventing children from creating and sharing CSAM material by building an image recognition software into the device's camera that stops the child from taking nude pictures and sharing them.

Advantages

Parental control tools enable parents to proactively protect their children and are usually flexible to allow adaption to the child's age, maturity and personal needs. Additionally, the variety of available offerings allows parents to create flexible protection mechanisms tailored to the specific platforms and devices their children are accessing and the risks they may be more prone to.

Limitations and issues

1. **Complexity of parental control settings:** To activate controls, parents have to choose from a multitude of different options. Most options only cover one device, one game or platform and only a selected number of activities and features. Furthermore, many of the settings are complex and can be challenging to navigate, even for digitally savvy parents.

2. **Differing levels of parental engagement:** Parental control tools assume parents are aware of their children's digital activities, have the required (digital) literacy to use control tools, and want and can be engaged with their children's online activities. This is often not the case, particularly in less-educated, lower-income households.

Evidence of the use of parental controls underpins their limitations as a safety tool. A 2016 survey found that only 39% of parents (two-thirds of whom were from the United States and the European Union) used some form of parental control or monitoring tool.[63] Moreover, research consistently finds that parents overstate their knowledge about their children's digital activities while underestimating their children's media usage.[64] Additionally, parents often only start implementing parental control tools after an incident has occurred. Relying on parental control tools for child safety can be seen as the platforms' way of pushing their responsibility to keep children safe on to parents.

[63] Pew Research Center. "Parents, Teens and Digital Monitoring". January 2016.

[64] Elleen Brown. "Most Parents Never check their Children's Devices." ZDNet. July 2019.

FIGURE 19: **ADVANTAGES & LIMITATIONS OF SOLUTIONS**

Tool	Advantages	Limitations
Age assurance	Provides information needed for age-based safety tools and measures	Requires sensitive personal data Missing interoperability between experiences/platforms Undermines children’s right to access digital spaces
AI classifiers	Scalability of analytical capabilities Variety of data that can be analyzed Pace of analysis and decision-making	Dependent on training data Reflects biases in datasets Unable to understand nuances and context Inadequate for spotting new trends Reacts to issues instead of being proactive
Moderating users	Avoids potentially abusive encounters Prevents incidents before they occur Creates more fun experiences for other users Detects perpetrators proactively	Requires personal data, raising privacy concerns Can easily lead to profiling of users and biases Requires amounts of data that may be unavailable
Parental controls	Equips parents with tools to protect children Adapts experience to individual needs of each user	Often settings are too complex Dependent on parental engagement Shifts responsibility to parents

More Needs to Happen for Child Safety in Online Gaming

V. REGULATORY SOLUTIONS ARE MOST IMPORTANT LEVER TO IMPROVE CHILD SAFETY IN ONLINE GAMING

I. Different countries have taken different approaches, with new proposals coming

Safety standards for consumer goods have been instrumental in incentivizing companies to ensure that new products embed child safety features as an integral part of the product development process.[65] But digital products are still not adequately regulated, and policymakers and regulators have been struggling to keep up with the rapid pace of technological and societal change. With risks to children online becoming an increasingly well-known issue, regulations remain the most potent tool for society to force companies to act and protect children in the online world. Recently, multiple legislative proposals have been under advanced discussion, particularly in the European Union and the United Kingdom and proposals worldwide are beginning to converge along some common principles, including:

- Closer collaboration between platforms and law enforcement
- Proactive risk assessment which should lead to more Safety by Design
- Transparency requirements to disclose processes
- Reduction in exposure to illegal content

FIGURE 20: **OVERVIEW OF CURRENT REGULATORY LANDSCAPE AND NEW DEVELOPMENTS**

	Current legislation and approach	New developments
EU	› Multiple EU directives and legislation cover aspects of children's lives in the digital world › Supplemented by national approaches	› Digital Services Act › Proposal for a regulation to prevent and combat child sexual abuse
UK	› Children's Code/Age-Appropriate Design Code enforced since Sep 2021, developed under Data Protection Act 2018	› Online Safety Bill (proposed)
Australia	› eSafety Commissioner (eSafety) established under the Enhancing Online Safety Act 2015	› Online Safety Act 2021, commenced in January 2022
USA	› Various pieces of legislation, including the Children's Online Privacy Protection Rule	› Kids Online Safety Act (proposed)

[65] See US Consumer Product Safety Commission for an example of federal safety rules that cover children's physical products.

The impending regulations will increase reporting and transparency requirements for companies. Companies that are attracting children have a responsibility to conduct risk assessments of their products and services, in order to identify risks for children and report on internal processes to mitigate such risks. This can be seen as a first step towards a Safety by Design approach. Regulatory debates are, however, often strongly framed by big tech companies and the perspectives and roles of smaller companies can get lost. Greater reporting and transparency obligations will also result in an increase of external service providers that monitor child risks on platforms.

The effectiveness of any legislation will ultimately, however, be determined by the enforcement and financial ramifications that the legislation will have on platforms. The level of fines and the likelihood of enforcement will encourage platforms to introduce new Safety by Design measures. The risk of reduced profits creates a financial incentive to act. Currently, most of the promising new developments are proposals and subject to debate and amendment. The proposals may change or may not pass and it remains unclear how effective enforcement will be should they pass.

Figure 21 sets out the different approaches that policymakers and governments in the European Union, the United Kingdom, Australia and the United States are taking as of summer 2022. More detail is provided around the developments in Europe and the United Kingdom, as both have comprehensive regulatory packages that are in advanced stages of discussion and are expected to have a major impact on the industry.

II. European Union is discussing multiple proposals that would impact child safety

I. Digital Services Act (DSA)

The overarching aim of the European Digital Services Act (DSA) is to protect the rights of all users online, including children, and to ensure they have less exposure to illegal content. The European Commission expects the DSA to lead to the implementation of age-assurance solutions. Very large platforms will be required to analyze any systemic risks for children stemming from children using the platforms and will be required to put in place effective content moderation mechanisms to address these risks. Large gaming companies like Roblox and Fortnite are expected to fall within the regulation's scope.

The European Commission requires that the rights of children must be protected and that targeted measures such as parental control tools, as well as tools that help children signal abuse, are introduced.[66] The level of fines for non-compliance is set at up to 6% of the company's annual turnover,[67] which is higher than the level of fines introduced by the by the GDPR. Enforcement will be dependent on the new European Digital Service Board; civil society groups have raised concerns that without proper implementation and strong enforcement mechanisms, the DSA risks becoming a paper tiger, which would mean that social gaming and metaverse platforms may largely continue with business as usual.

II. EU Commission proposal to combat CSAM

The European Commission estimates that over 60% of sexual abuse material globally is hosted in the European Union.[68] This stark statistic underpins the Commission's proposal to set out new rules for platforms to be legally required to detect, report and remove CSAM. Regulatory reporting requirements can create a huge push in the online protection of children as seen with the example of the CSAM reporting obligation to the National Center for Missing and Exploited Children in the United States. The European proposal also includes an obligation for communication services to assess the risk of CSAM spreading and of grooming occurring on their platforms.

[66] European Commission. "Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions: A Digital Decade for children and youth: the new European strategy for a better internet for kids (BIK+)". May 2022.

[67] European Commission. "Proposal for a Regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and Amending Directive 2000/31/EC". December 2020.

[68] Natasha Lomas. "Europe's CSAM Scanning Plan Unpicked". TechCrunch. May 2022.

FIGURE 21: **COMPARISON OF APPROACHES OF POLICYMAKERS ACROSS AUSTRALIA, THE UK, THE EU AND THE US**

	Australia (Online Safety Act 2021)	UK (AADC and Online Safety Bill)	EU (Digital Services Act)	US (Kids Online Safety Act)
Fines and penalties	(✓)	✓	✓	⊗
Risk mitigation measures	✓	✓	✓	(✓)
Introduction of age assurance	(✓)	(✓)	(✓)	⊗
Content moderation	(✓)	✓	✓	✓
Cooperation with law enforcement	(✓)	✓	✓	⊗
Status	In Effect	AADC: In Effect OS Bill: Pending Approval	Pending Approval	Proposed

✓ Covered (✓) Partly Covered ⊗ Not Covered

Messaging services will have to scan messages for potential CSAM materials, which some critics argue could undermine end-to-end encryption. There are worries that the proposal could, if passed, introduce mass surveillance on interpersonal communication services. Others are concerned that authorities may be overwhelmed by the number of additional reports received without having the tools or the means to follow-up on them in order to capture adult perpetrators. With sexting among teenagers on the rise, many teenagers would likely be flagged for explicit peer-to-peer conversations. The proposal is still at an early stage and will likely be amended.

III. UK has innovative approach in terms of AADC (or Children’s Code) and has proposed Online Safety Bill

One of the most advanced Safety by Design legislations is the age-appropriate design code (AADC or Children’s Code).[69] It is the first statutory code of practice in the world addressing the use of children’s data. The code applies to online services that are likely to be accessed by children under 18, even if children are not the service’s target audience. One of the limitations of the AADC is that it does not cover sites that are not likely to be accessed by children, such as pornographic websites. This will however be expected to be covered by the Online Safety Bill, once passed. Like the DSA, the Online Safety Bill is broad in scope but also introduces protections for children. The most recent draft of the proposed Online Safety Bill was intended to cover companies whose services host user-generated content, such as images, videos and comments, or which allow other user-to-user communication. The Online Safety Bill would introduce a new duty of care for online platforms towards their users, requiring them to assess and take action against illegal and also legal, but harmful, content. Social gaming platforms are expected to be within this scope and the regulatory intention is to cover metaverse platforms, too. It is uncertain, however, how it will be enforced and to what extent the legislation will evolve as it continues to go through the parliamentary process.

IV. Future legislation needs to be comprehensive and flexible to keep up with technological advances

Beyond proposals in the European Union and the United Kingdom, current proposed and enforced legislation in many countries around the globe do not adequately address the dangers children face on social gaming and metaverse platforms. Comprehensive legislation needs to explicitly address gaming platforms and formulate standards social gaming platforms, and by extension metaverse platforms, must adhere to. Governments and regulators have the greatest potential to transform the child safety space and only they can enforce a level playing field and inflict monetary consequences for neglecting child safety. Figure 22 summarizes what an ideal regulatory framework could include.

[69] Information Commissioner’s Office. ”Introduction to the Age-Appropriate Design Code”. Retrieved June 2022.

FIGURE 22: **OVERVIEW OF AN IDEAL REGULATORY FRAMEWORK**

Prevent

Key regulatory steps to prevent child abuse:

- **Age assurance** required as basis for further age-based safety tools
- **Risk assessment** of experiences for adequate safety countermeasures
- **Age-appropriate restrictions** based on individual users age-group with basic functions enabled by default
- Restricted **access for sex offenders** to platforms for children
- **Social media regulation** applied to companies that offer similar functions
- **Child safety officer** established to oversee child safety strategy and implementation

Detect

Key regulatory steps to detect child abuse:

- **Abuse identification** through detection of CSAM uploaded or distributed on such platforms and by deprioritizing end-to-end encryption in messages involving minors under 13
- **Minimum qualifications** for human moderators
- **User enablement to report violations and seek help** with "one-click-away" reporting of abusive behavior and on-platform psychological or helpline support

Prosecute

Key regulatory steps to prosecute child abuse:

- **Transparent public reporting** for all platforms actively used by children
- **Cooperation with law enforcement through** single contact points and frameworks and minimum standards for communication of identified abuse

Enforce

Key steps to enforce abuse regulation:

- **Authority** on national or supranational level to monitor and evaluate adherence to regulation
- **Power** to proactively investigate and issue fines in case of non-compliance

VI. RECOMMENDATIONS

The trends identified in this report paint a worrying picture of new technologies and their potential negative impact on child safety. Turning the tide will require a consolidated and coordinated effort by all key stakeholders from private companies, governments, parents, investors and researchers, with each having their own role to play:

1. Governments

Effective regulation is one of the most powerful tools available to enforce a consistent and holistic framework for child safety on social gaming and metaverse platforms. Governments need to use this power to level the playing field and ensure that all platforms provide the same basic level of child safety. This will require legislators and regulators to recognize the changing nature of gaming and update child safety requirements to match the realities of the new digital era. School curricula will have to be updated to teach children about responsible behavior in the metaverse and how to protect themselves.

2. Platforms

Platforms need to live up to their duty of care and place safety at the center of the design and development processes of digital experiences. With technology constantly evolving, companies should embrace a Safety by Design approach and promote child safety not as an afterthought but as an integral part of their business strategy, in order to deliver the truly joyful and beneficial experiences for children that they promise.

3. Parents

Parents need to familiarize themselves with their children's digital habits and the risks associated with them in the same way they understand the risks of the offline world. This will require them to invest time to understand and implement appropriate safety measures where necessary.

4. Investors

Investors hold a key position in the development of new platforms, as companies rely on fresh capital to realize their ideas and to grow. Recent years have shown the power investors have in shifting corporate attention to climate and social challenges. Investors further determine the Key Performance Indicators (KPIs) that are used to evaluate whether a company is deemed a success or not and to assess the potential worth of a given platform. This power can be used to include child safety KPIs as part of their assessment.

5. Researchers

The novel nature of social gaming and metaverse platforms and many of the underlying technologies (e.g., VR headsets) leaves us with many open questions: What is the impact of children's social lives shifting to these spaces? How does extensive time spent using VR headsets influence children's health? How does the immersive experience affect children's cognitive, social and psychological development, particularly in the context of sexual exploitation and abuse? What legal frameworks should be applied to crimes in the metaverse?

The research community is needed to provide governments, platforms and society with answers to many of these questions that will have a large impact on how these spaces are governed and used.

Stakeholders | Specific steps, actions, responsibilities

Governments

Prevent

1. Age-assurance

Social gaming platforms must be required to implement age-assurance tools to be able to identify whether minors are on their platforms. Social gaming platforms can choose the age-assurance method they consider adequate for their platforms. Regulators should be allowed to demand more stringent assurance methods should, for example, a social gaming platform decide to declare itself an 18+ platform.

2. Risk assessment

Information on the age distribution will allow social gaming platforms to perform age-appropriate ongoing risk assessments and formulate safety countermeasures. These assessments promote Safety by Design throughout all development stages. Platforms should be required to document and share these risk assessment measures with regulators when asked to.

3. Age-appropriate restrictions

Social gaming platforms must enforce an individualized age-group based experience to account for different safety needs of different age groups. A set of basic safety functions should always be turned on by default for all minors under 13 and only be lifted with parental consent. These should particularly cover the ability of strangers to communicate with chilren in unmoderated and encrypted spaces (e.g., unmoderated voice chat).

4. Access by sex offenders

Laws should ban registered child sex offenders from platforms that particularly attract children. While implementation is currently technically challenging, current legislation offers paths forwards. In the United Kingdom, courts already have the power to restrict a convicted sex offender's ability to access the internet without an installed computer monitoring software. Including child-focused social gaming platforms on the list of restricted activities for sex offenders additionally provides law enforcement with the legislative power to prosecute and enforce potential violations against such regulation.

5. Applying social media regulations

If gaming services offer functions such as chatting and connecting with friends, all regulation applicable to social media, such as the Children's Online Privacy Protection Act in the United States or the European Union's e-Commerce Directive, should also apply to them.

6. Child safety officer

All platforms targeting minors should be obliged to have a child safety officer that is responsible for overseeing an organization's child safety strategy and implementation.

Stakeholders | Specific steps, actions, responsibilities

Governments

Detect

1. Abuse identification

Social gaming platforms must be required to detect all CSAM and harmful content that is uploaded to or distributed on their platforms. To further reduce risks of grooming and other forms of exploitation, all text chat messages involving minors under 13 on social gaming platforms should deprioritize end-to-end encryption to allow the analysis of the conversations for potential patterns of abuse. Due to the psychologically more dangerous nature of VR, a sufficient duration of the last gameplay using VR devices must be saved at all times – if possible, on the user's device for privacy reasons – to allow users to collect evidence in the case of abuse.

2. Qualified human moderators

Social gaming platforms must be required to hire human moderators with minimum qualifications to identify and address issues. Human moderators must undergo minimum training levels to support them in understanding the risks children face in these spaces.

3. User enablement to report violations and seek help

Social gaming platforms need to empower users to protect themselves by providing easily accessible reporting and blocking tools. Social gaming platforms should be required to integrate "one-click-away" reporting and blocking options for users. They must also provide children with easy ways to find help on platforms by directly offering psychological counseling or linking to third parties and helplines to reduce the known barriers for children seeking support.

Governments

Prosecute

1. Transparency requirements

Currently proposed and enforced transparency requirements should be expanded to all platforms that are actively being used by children. The annual reporting by the Internet Watch Foundation of digital child abuse can serve as a proxy for the kind of information that should be made available (e.g., age groups of actual users, number of reported incidents).

2. Cooperation with law enforcement

To live up to their duty of care, all platforms that are used by children or on which CSAM has been detected, must by law be obliged to cooperate with law enforcement – independent of their size. They must provide law enforcement with all the data needed to prosecute perpetrators. For this purpose, they should have a single point of contact within the organization for law enforcement to communicate with and establish standards for formatting and exchanging case data with law enforcement.

Stakeholders	Specific steps, actions, responsibilities	
Governments	**Enforce**	The key to child protection legislation having an impact is effective enforcement. Hence, a national or, for the European Union, supranational supervisory authority is needed that monitors and evaluates adherence of companies to child safety regulations. The authority should be able to proactively investigate platforms and to issue fines in case of non-compliance.
Platforms		• Invest in tech solutions that improve child safety on their platforms • Be honest in recognizing and addressing safety challenges • Cooperate with other players and engage in industry alliances to share best practices, ensure common standards and facilitate dialog with other platforms, regulators, civil society (e.g., WeProtect Global Alliance, Tech Coalition)[70]
Parents		• Be aware of the games and experiences their children engage in • Familiarize themselves with the parental control tools available on the platforms their children access • Understand the dangers lurking on some of these platforms • Inform themselves and enter a dialog about risks and preventive measures with their children • Urge their representatives to push effective regulation in the realm of child safety
Investors		• Integrate child safety considerations into their investment due diligence process • Work proactively with investees by sharing knowledge among portfolio companies and facilitating the establishment of best practices • Stop the flow of money to ventures that do not consider child safety a core principle
Researchers		• Research the impact of immersive experiences (VR, AR) on child development by age and particularly the impact of traumatic experiences like child sexual exploitation and abuse • Investigate the scale of exposure to child sexual exploitation and abuse on gaming platforms • Develop legal frameworks for prosecuting crimes in the metaverse

[70] The Tech Coalition's "Project Protect" is a positive example of these corporations. It sets out to coordinate and drive collective action in establishing industry standards for innovative technology, transparency and accountability, information and knowledge sharing and independent research. Specifically, they distribute best practice resources and trainings, roadmaps for establishing industry safety standards and data sharing tools among their members.

ACKNOWLEDGMENTS

This Report has been produced by **Bracket Foundation** in collaboration with the **UNICRI Centre for AI and Robotics** and **Value for Good.**

The following individuals contributed with interviews and input:

Amanda Menkin
Trust & Safety Professional Association

Andy Burrows
NSPCC

Brenda Dvoskin
Berkman Klein Center for Internet & Society

Callum Hood
Center for Countering Digital Hate

Prof. Clare McGlynn
Durham University

Deepak Tewari
Privately

Hannah Ruschen
NSPCC

Izzy Wick
5Rights Foundation

Julie Dawson
YOTI

Leanda Barrington-Leach
5Rights Foundation

Linn Bieske
Harvard Kennedy School

Louis-Victor de Franssu
Tremau

Nichlas Hellmark
Save the Children

Phillip Amann
Europol

Rodrigo Tamellini
Gamersafer

Sean Litton
Tech Coalition

Simon Burt
Purpose

Sonia Livingston
London School of Economics

Tanya Vanden Berg
Parent

Dr. iur. Thomas-Gabriel Rüdiger
Hochschule der Polizei des Landes Brandenburg

Tom Farell
SafeToNet

ABOUT THE AUTHORS

Bracket Foundation is the philanthropic venture arm of Bracket Capital, a leading private technology investor based in Los Angeles with offices in Doha and London. Bracket Foundation's mission is to harness the power of technology for social good by leveraging technology solutions to tackle growing global challenges. In 2019, Bracket Foundation and its partners published a leading report on how Artificial Intelligence (AI) can combat online sexual abuse of children. The publication was presented on the sidelines of the United Nations (UN) General Assembly that same year and resulted in a multi-year partnership with the UNICRI (the United Nations Interregional Crime and Justice Research Institute). The partnership served as the backbone for the UN sponsored "AI for Safer Children" platform that ensued to empower law enforcement agencies worldwide with innovative tools to better detect, prevent and prosecute online sexual abuse being committed against children. Bracket Foundation is engaged with several public sector actors which include multi-lateral organizations (such as the UN, the European Union, the European Commission), NGOs and States to raise awareness on the uses of AI, building trust between the public and private sector, promoting more government investment in AI, lobbying for more data sharing commons and changes to the legislative framework around data use in order to scale a global solution to this issue. Bracket Foundation is also engaged in advocacy work especially as it relates to holding Big Tech companies accountable for the crimes being committed on their platforms.

Yalda Aoukar is Co-Founder/Managing Partner of Bracket Capital, President of Bracket Foundation and leads Bracket's nascent consulting arm. She is a fierce advocate of technology for good especially as it relates to solving the world's most pressing global challenges. She is a champion for women's empowerment and entrepreneurship in Venture Capital and other financial sectors, where women are traditionally underrepresented. In addition to investing in leading technology companies, she serves as an adviser to governments and policy makers on digital development in diverse fields such as Biotech, Food Security, Artificial Intelligence Integration and Education Technology. She sits on the board of the United Nation's backed AI for Safer Children Initiative which she helped launch, as well as the World Innovation Summit for Education (WISE). Yalda lives between London and Doha, some might say cruising at 35000 feet on an Airbus a380, with her husband and three children who also serve as her commanders in chief.

UNICRI Center for AI and Robotics was launched as a program in early 2015 by UNICRI (United Nations Interregional Crime and Justice Research Institute), and with support of the Municipality of the Hague and the Ministry of Foreign Affairs of the Netherlands UNICRI signed the host country agreement for the opening of its Centre for Artificial Intelligence and Robotics in The Hague, the Netherlands, in September 2017. This Centre is dedicated to understanding and addressing the risks and benefits of AI and robotics from the perspective of crime and security through awareness-raising, education, exchange of information, and harmonization of stakeholders.

Value for Good is a consultancy specialized in the field of social impact that envisions a world in which effective action is taken to solve societal challenges. To achieve this Value for Good inspires through insights, advises through consulting and empowers through training. Value for Good serves leaders from private sector, governments, international institutions, foundations and non-profits and equips them with the knowledge and tools to make a positive and sustainable difference.

Clara Péron is the founder and managing director of Value for Good GmbH. Originally from Montréal, Canada, she has lived and worked internationally – with longer postings in India, Cambodia, Egypt, Ukraine the US and Germany. Clara started her career in 2002 in the Canadian foreign service and after working as a strategy consultant for the Boston Consulting Group's Berlin office she founded Value for Good and the Value for Good foundation.

Ahmed Ragab is a principal at Value for Good with a focus on Tech for Good. Prior to joining Value for Good he was COO of the non-profit EdTech Kiron Open Higher Education and a senior consultant at McKinsey and Company. He studied Tech regulation and Trust and Safety in Big Tech while completing his Master's in Public Administration at the Harvard Kennedy School.

FURTHER READINGS

Common Sense. "Kids and the Metaverse: What Parents, Policymakers, and Companies Need to Know". 2022. https://www.commonsensemedia.org/kids-action/articles/what-are-kids-doing-in-the-metaverse

Microsoft Research . "Content moderation, AI, and the question of scale". 2020. https://www.researchgate.net/publication/343798653_Content_moderation_AI_and_the_question_of_scale/fulltext/5f65964c458515b7cf3edcae/Content-moderation-AI-and-the-question-of-scale.pdf?origin=publication_detail

5Rights Foundation. "But how do they know it is a child?". 2021. https://5rightsfoundation.com/uploads/But_How_Do_They_Know_It_is_a_Child.pdf

London School of Economics. "Understanding of User Needs and Problems: A Rapid Evidence Review of Age Assurance and Parental Controls". 2021. file:///C:/Users/MartenGillwald/Downloads/D2.4a%20Understanding%20user%20needs_Rapid%20evidence%20review%20Public%20version%20v2.pdf

McKinsey & Company. "Value Creation in the Metaverse". 2022. https://www.mckinsey.com/business-functions/growth-marketing-and-sales/our-insights/value-creation-in-the-metaverse

Internet Watch Foundation. "IWF Annual Report 2021". 2022. https://annualreport2021.iwf.org.uk/

Jakki Bailey and Jeremy Bailenson. "Immersive Virtual Reality and the Developing Child". 2018. https://www.stanfordvr.com/mm/2017/07/bailey-ivr-developing-child.pdf

Thorn. "Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting, and Blocking". 2021. https://info.thorn.org/hubfs/Research/Responding%20to%20Online%20Threats_2021-Full-Report.pdf

Sonia Livingston and Mariya Stoilova. "The 4Cs: Classifying Online Risk to Children". 2021. https://www.ssoar.info/ssoar/bitstream/handle/document/71817/ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf?sequence=4&isAllowed=y&lnkname=ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf

BIBLIOGRAPHY

A. (2013, December 2). Minecraft Player Demographics. Minecraft Seeds Blog. https://minecraft-seeds.net/blog/minecraft-player-demographics/

Age appropriate design: a code of practice for online services. (2021). Information Commissioner's Office. https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services-2-1.pdf

Aishah Rahman, news reporter. (2022, May 20). Sextortion cases reported to revenge porn helpline double in a year. Sky News. https://news.sky.com/story/sextortion-cases-reported-to-revenge-porn-helpline-double-in-a-year-12617111

Anderson, M. (2020, May 30). Parents, Teens and Digital Monitoring. Pew Research Center: Internet, Science & Tech. https://www.pewresearch.org/internet/2016/01/07/parents-teens-and-digital-monitoring/

AR/VR Headset Shipments Grew Dramatically in 2021, Thanks Largely to Meta's Strong Quest 2 Volumes, with Growth Forecast to Continue, According to IDC. (2022, March 21). IDC: The Premier Global Market Intelligence Company. https://www.idc.com/getdoc.jsp?containerId=prUS48969722

Bailey, J. O., & Bailenson, J. N. (2017). Immersive Virtual Reality and the Developing Child. Cognitive Development in Digital Contexts, 181–200. https://doi.org/10.1016/b978-0-12-809481-5.00009-2

BBC News. (2017, January 20). Minecraft paedophile Adam Isaac groomed boys online. https://www.bbc.com/news/uk-wales-south-east-wales-38691882

BBC News. (2022, March 3). Young girl returned after kidnapping by man she met on Roblox. https://www.bbc.com/news/world-us-canada-60607782

Bowles, N., & Keller, M. H. (2020, February 3). Video Games and Online Chats Are 'Hunting Grounds' for Sexual Predators. The New York Times. https://www.nytimes.com/interactive/2019/12/07/us/video-games-child-sex-abuse.html?mtrref=undefined&gwh=0F67BD5B84BCCE4641D563D718CF1F03&gwt=pay&assetType=PAYWALL

Bracket Foundation

Brown, E. (2019, July 29). Most parents never check their children's devices. ZDNET. https://www.zdnet.com/article/most-parents-never-check-their-childrens-devices/

Center for Countering Digital Hate. (2022, May 13). Facebook's Metaverse. Center for Countering Digital Hate | CCDH. https://counterhate.com/research/facebooks-metaverse/

Children and parents: media use and attitudes report 2020/21. (2021, April). Ofcom. https://www.ofcom.org.uk/__data/assets/pdf_file/0025/217825/children-and-parents-media-use-and-attitudes-report-2020-21.pdf

Colville, A. (2022, June 28). Yoti age estimation approved by German regulator KJM for the highest level of age assurance covering 18+ adult content ·. Yoti. https://www.yoti.com/blog/age-estimation-approved-kjm-highest-age-assurance-level-18-adult-content/

Davis, L. (n.d.). Best Practices for Voice Chat Moderation. Spectrum Labs. https://www.spectrumlabsai.com/the-blog/best-practices-for-voice-chat-moderation

Dyer, J. C. J. B. (2022, February 15). Roblox: The children's game with a sex problem. BBC News. https://www.bbc.com/news/technology-60314572

ESA. (2022). Essential Facts About the Video Game Industry. Entertainment Software Association (ESA). https://www.theesa.com/wp-content/uploads/2022/06/2022-Essential-Facts-About-the-Video-Game-Industry.pdf

Etherington, D. (2014, September 15). TechCrunch is part of the Yahoo family of brands. TC TechCrunch. https://techcrunch.com/2014/09/15/microsoft-has-acquired-minecraft/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuYnVzaW5lc3N-vZmFwcHMuY29tLw&guce_referrer_sig=AQAAAH-kEeql1yaITRrs9lSOSYH663Q_1mloTX5eAjRqdKjGlM8tn_vX1KWXaBvhU5zADgxL_Fy1ra2C5nUM9YtkENDfM6oZ-qhzJYp6yvSYuLTBuTXckR1_jiV1mYxMbJJsGrQvi84Yu4f-zo9vZDX7jDWU_Gj9yPnwyc7T3ejLrOfXGZV

EUR-Lex - 52020PC0825 - EN - EUR-Lex. (2022). EUR-Lex. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN

EUR-Lex - 52022DC0212 - EN - EUR-Lex. (2022). EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2022:212:FIN

FBI Internet Crime Complaint Center. (2021, March). 2020 Internet Crime Report. Federal Bureau of Investigation. https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf

Fox Business. (2022, January 18). FTC investigates Meta's Oculus VR over market dominance. New York Post. https://nypost.com/2022/01/18/ftc-investigates-metas-oculus-vr-business-over-market-dominance/

Galluch, M. (2022, March 7). The Case for Real Time Voice Chat Moderation Technology in the Metaverse | Speechly. Speechly. https://www.speechly.com/blog/the-case-for-real-time-voice-chat-moderation-technology-in-the-metaverse

Gartner. (n.d.-a). Definition of Augmented Reality (AR) - Gartner Information Technology Glossary. https://www.gartner.com/en/information-technology/glossary/augmented-reality-ar

Gartner. (n.d.). Definition of Virtual Reality (VR) - Gartner Information Technology Glossary. https://www.gartner.com/en/information-technology/glossary/vr-virtual-reality

Hersey, F. (2022, May 13). Global movement coalescing around age verification and its role in online safety. Biometric Update. https://www.biometricupdate.com/202203/global-movement-coalescing-around-age-verification-and-its-role-in-online-safety

Internet Watch Foundation. (2022). IWF Annual Report 2021. https://annualreport2021.iwf.org.uk/

IWF. (2021). Internet Watch Foundation Annual Report 2021 | IWF. Internet Watch Foundation. https://www.iwf.org.uk/about-us/who-we-are/annual-report-2021/

J. M. N. (2022, January 19). Metaverse Gamers: Demographics, Playing and Spending Behavior. Newzoo. https://newzoo.com/insights/articles/deep-dive-metaverse-gamers-data-on-metaverse-demographics-socializing-playing-spending-2

Kommission für Jugendmedienschutz Startseite. (2022, July 28). Kommission für Jugendmedienschutz. https://www.kjm-online.de/

Kugler, L. (2021, February 1). The State of Virtual Reality Hardware. February 2021 | Communications of the ACM. https://cacm.acm.org/magazines/2021/2/250071-the-state-of-virtual-reality-hardware/fulltext

Levy, A. (2022, January 19). Microsoft sets record for biggest tech deal ever, topping Dell-EMC merger in 2016. CNBC. https://www.cnbc.com/2022/01/18/biggest-tech-deal-ever-microsoft-activision-set-69-billion-record.html

Livingstone, S., & Stoilova, M. (2021). The 4Cs: Classifying Online Risk to Children. (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE - Children Online: Research and Evidence. https://doi.org/10.21241/ssoar.71817

Lomas, N. (2022, May 11). Europe's CSAM scanning plan unpicked. TechCrunch. https://techcrunch.com/2022/05/11/eu-csam-detection-plan/?guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAANjWCk8z9Z3EryGounby_QhsynP_jZVNowkJ1beDBsPAaEH9k1gCejkgzLTsItKAa5n-575V6BISTsVX8-WB83fNtKbxWVgoODt08AfqQZ-g2jvMSKZc4ajJrUDFLd2B9coO3luYYMluHwHC-Z1nn9bkyhXXHlrLYUSPojiElMqkS6q&guccounter=2

Maister, L., Slater, M., Sanchez-Vives, M. V., & Tsakiris, M. (2015). Changing bodies changes minds: Owning another body affects social cognition. Trends in Cognitive Sciences, 19, 6–12. https://doi.org/10.1016/j.tics.2014.11.001

Meta is putting a stop to virtual groping in its metaverse by creating 4-foot safety bubbles around avatars. (2022, February 5). Business Insider. https://www.businessinsider.com/meta-metaverse-virtual-groping-personal-boundary-safety-bubble-horizons-venues-2022-2?international=true&r=US&IR=T

Meta Reports Fourth Quarter and Full Year 2021 Results. (2022). Meta. https://investor.fb.com/investor-news/press-release-details/2022/Meta-Reports-Fourth-Quarter-and-Full-Year-2021-Results/default.aspx

Minecraft Revenue and Usage Statistics (2022). (2022, July 26). Business of Apps. https://www.businessofapps.com/data/minecraft-statistics/

Mordor Intelligence. (2022). Gaming Market Size, Value | Industry Forecast 2022 - 27. https://www.mordorintelligence.com/industry-reports/global-gaming-market?gclid=Cj0KCQjwspKUBhCvARIsAB2I-YusKdp7J8Fzuz-yrIw4vg3FTZkdzY8RaHwLiAOmrZEAqJZ1I-fuf6Qs8aAoFlEALw_wcB

Motion Picture Association of America. (2021, June 2). 2019 THEME Report. Motion Picture Association. https://www.motionpictures.org/research-docs/2019-theme-report/

National Center for Missing & Exploited Children. (2014). 2014 Annual Report - National Center for Missing and Exploited. https://studylib.net/doc/18633821/2014-annual-report---national-center-for-missing-and-expl

National Center for Missing & Exploited Children. (2021). CyberTipline Data: CyberTipline 2021 Report. https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata

NSPCC. (2021, August 24). Record high number of recorded grooming crimes lead to calls for stronger online safety legislation. https://www.nspcc.org.uk/about-us/news-opinion/2021/online-grooming-record-high/

NSPCC. (2021b, October 6). New figures reveal four in five victims of online grooming crimes are girls. https://www.nspcc.org.uk/about-us/news-opinion/2021/online-grooming-crimes-girls/

Pew Research Center (2018, September). A Majority of Teens Have Experienced Some Form of Cyberbullying.

Phillips, J. (2022, April 25). Metaverse branded an "online Wild West" as Channel 4 uncovers evidence of sexual abuse and racism. Mail Online. https://www.dailymail.co.uk/news/article-10750407/Metaverse-branded-online-Wild-West-Channel-4-uncovers-evidence-sexual-abuse-racism.html

PricewaterhouseCoopers. (n.d.). Global Entertainment & Media Outlook 2022–2026: TMT. PwC. https://www.pwc.com/outlook

Roblox - Financials - SEC Filings. (2022). Roblox. https://ir.roblox.com/financials/sec-filings/default.aspx

Roblox has eclipsed Minecraft with 100 million users — highlighting the popularity of "digital hangouts." (2019, August 7). Business Insider. https://www.businessinsider.com/roblox-user-growth-surpasses-minecraft-2019-8?international=true&r=US&IR=T

Rose, J., & Phillips, J. (2022, April 25). Channel 4 Dispatches shows Metaverse users boasting that they are attracted to "little girls." Mail Online. https://www.dailymail.co.uk/news/article-10752287/Channel-4-Dispatches-shows-Metaverse-users-boasting-attracted-little-girls.html

Rosen, V. G. P. (2021, March 22). How We're Tackling Misinformation Across Our Apps. Meta. https://about.fb.com/news/2021/03/how-were-tackling-misinformation-across-our-apps/

Safety by Design. (n.d.). eSafety Commissioner. https://www.esafety.gov.au/industry/safety-by-design

Sexueller Missbrauch, Missbrauch und Nacktdarstellung von Kindern. (2022). Facebook. https://transparency.fb.com/de-de/policies/community-standards/child-sexual-exploitation-abuse-nudity/

Smith, B. A. C. A. T. (2022, February 23). Metaverse app allows kids into virtual strip clubs. BBC News. https://www.bbc.com/news/technology-60415317

Statista. (2021, November 30). Roblox games users distribution worldwide September 2020, by age. https://www.statista.com/statistics/1190869/roblox-games-users-global-distribution-age/

Statista. (2022, August 2). Minecraft active player count worldwide 2016–2021. https://www.statista.com/statistics/680139/minecraft-active-players-worldwide/

Statista. (2022, July 27). North American sports market size 2009–2023. https://www.statista.com/statistics/214960/revenue-of-the-north-american-sports-market/

Statista. (2022a, February 22). Roblox Corporation global revenue 2018–2021. https://www.statista.com/statistics/1189990/annual-revenue-roblox-corporation/

Statista. (2022a, May 10). Epic Games annual gross revenue 2018–2025. https://www.statista.com/statistics/1234106/epic-games-annual-revenue/

Statista. (2022a, May 18). Market value of the largest gaming companies worldwide 2020–2022. https://www.statista.com/statistics/1197213/market-value-of-the-largest-gaming-companies-worldwide/

Statista. (2022b, July 27). Fortnite player distribution in the U.S. 2018, by age group. https://www.statista.com/statistics/865616/fortnite-players-age/

Stop It Now! UK and Ireland. (2021, October 12). Sexual harm prevention order SHPO (SOPO). Stop It Now. https://www.stopitnow.org.uk/concerned-about-your-own-thoughts-or-behaviour/concerned-about-use-of-the-internet/get-the-facts/consequences/being-subject-to-a-sexual-harm-prevention-order-shpo/

Tassi, P. (2021, March 12). Roblox's IPO Makes It Worth More Than EA, Take-Two And Ubisoft. Forbes. https://www.forbes.com/sites/paultassi/2021/03/12/robloxs-ipo-makes-it-worth-more-than-ea-take-two-and-ubisoft/?sh=7889317a72db

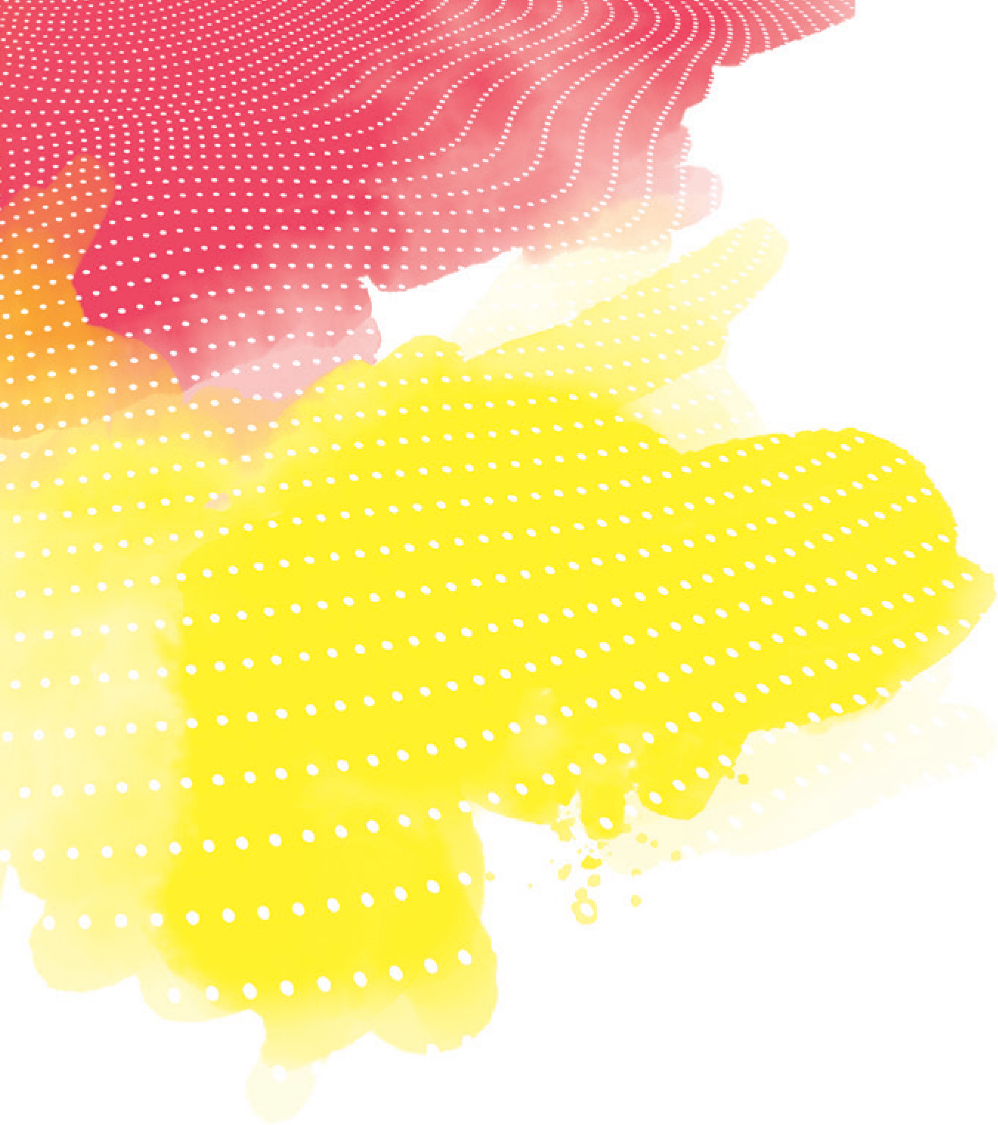
Thorn & Benenson Strategy Group. (2021, November). Self-Generated Child Sexual Abuse Material: Youth Attitudes and Experiences in 2020. Findings from 2020 quantitative research among 9–17 year olds. Thorn. https://info.thorn.org/hubfs/Research/SGCSAM_Attidues&Experiences_YouthMonitoring_FullReport_2021_FINAL%20(1).pdf

Thorn & Benenson Strategy Group. (2021a, May). Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting, and Blocking. Findings from 2020 quantitative research among 9–17 year olds. Thorn. https://info.thorn.org/hubfs/Research/Responding%20to%20Online%20Threats_2021-Full-Report.pdf

Thorn. (2022, February 22). Sextortion Research and Insights. https://www.thorn.org/sextortion/

VRChat - Steam Charts. (2022). SteamCharts. https://steamcharts.com/app/438100

Yin, J., Yuan, J., Arfaei, N., Catalano, P. J., Allen, J. G., & Spengler, J. D. (2020). Effects of biophilic indoor environment on stress and anxiety recovery: A between-subjects experiment in virtual reality. Environment International, 136, 105427. https://doi.org/10.1016/j.envint.2019.105427



IMPRINT

Authors

Bracket Foundation
8560 W. Sunset Blvd. Suite 700
West Hollywood, CA 90069
info@bracketfoundation.org

UNICRI Centre for AI & Robotics
Alexanderveld 5, 2585 The Hague
Netherlands

Ahmed Ragab and Clara Péron
Value for Good GmbH
Französische Str. 47, 10117 Berlin
Germany
mail@valueforgood.com

Funding Partner

Bracket Capital
8560 W. Sunset Blvd. Suite 700
West Hollywood, CA 90069
info@bracketcapital.org

Design, Layout and Illustrations

Malena Baum
malena@secret-monkey.com