ITUPublications

International Telecommunication Union
Standardization Sector

ITU Focus Group Technical Report

(12/2023)

ITU Focus Group on metaverse

Responsible use of AI for child protection in the metaverse

Working Group 6: Security, Data and Personally identifiable Information Protection

Technical Report ITU FGMV-13

Responsible use of AI for child protection in the metaverse

Summary

This Technical Report explores the scope for the responsible use of artificial intelligence (AI) for child protection in the metaverse as a contribution in this area to assist in the achievement of the United Nations sustainable development goals.

Keywords

AI-based systems, artificial intelligence (AI), digital threats, digital scams, metaverse, social media, technology.

Note

This Technical Report is an informative ITU-T publication. Mandatory provisions, such as those found in ITU-T Recommendations, are outside the scope of this publication. This publication should only be referenced bibliographically in ITU-T Recommendations.

Change Log

This document contains Version 1.0 of the ITU Technical Report on "Responsible Use of AI for Child Protection in the metaverse" approved at the 4th meeting of the ITU Focus Group on metaverse (ITU FG-MV), held on 4-7 December 2023 in Geneva, Switzerland.

Acknowledgements

This Technical Report was researched and written by Zaheema Iqbal (Global Foundation for Cyber Studies and Research, USA), Muhammad Khurram Khan (King Saud University; Kingdom of Saudi Arabia), Paul Grainger (University College London, United Kingdom) and Farhan Khan (Metaronical, United Arab Emirates) as a contribution to the ITU Focus Group on metaverse (ITU FG-MV). The development of this document was coordinated by Vincent Affleck (DSIT, United Kingdom), as FG-MV Working Group 6 Chair, and by Muhammad Khurram Khan (King Saud University; Kingdom of Saudi Arabia) as Chair of Task Group on child online protection.

Additional information and materials relating to this report can be found at: https://www.itu.int/go/fgmv. If you would like to provide any additional information, please contact Cristina Bueti at tsbfgmv@itu.int.

Editor & Task Group Chair:	Muhammad Khurram Khan King Saud University Kingdom of Saudi Arabia	E-mail: mkhurram@ksu.edu.sa
Editor:	Zaheema Iqbal Global Foundation for Cyber Studies and Research United States of America	E-mail: zaheemaeckbaull@gmail.com
Editor:	Paul Grainger University College London United Kingdom	E-mail: p.grainger@ucl.ac.uk
Editor:	Farhan Khan Metaronical United Arab Emirates	E-mail: farhanhbk@hotmail.com
WG6 Chair:	Vincent Affleck DSIT United Kingdom	E-mail: Vincentaffleck2@hotmail.com

Table of Contents

Technical Report ITU FGMV-13

Responsible use of AI for child protection in the metaverse

1 Scope

The purpose of this Technical Report is to formulate policy guidelines for the responsible use of artificial intelligence (AI) in protecting children in the metaverse. The scope includes:

i) Identify digital threats to children in the metaverse;

ii) Explore the role of education and AI in protecting children online;

iii) Identify international initiatives for child protection across the globe;

iv) Make policy proposals on how the responsible use of AI can help prevent children from digital threats.

2 References

None.

3 Definitions

3.1 Terms defined elsewhere

This Technical Report uses the following terms defined elsewhere:

3.1.1 artificial intelligence [b-ITU-T M.3080]: Computerized system that uses cognition to understand information and solve problems.

3.1.2 artificial Intelligence [b-ISO/IEC 2382]: Interdisciplinary field, usually regarded as a branch of computer science, dealing with models and systems for the performance of functions generally associated with human intelligence, such as reasoning and learning.

3.1.3 Internet of things (IoT) [b-ITU-T Y.4000]: A global infrastructure for the information society, enabling advanced services by interconnecting (physical and virtual) things based on existing and evolving interoperable information and communication technologies.

3.2 Terms defined in this Technical Report

None.

4 Abbreviations and acronyms

This Technical Report uses the following abbreviations and acronyms:

AI Artificial Intelligence

AWS Amazon Web Services

COP Child Online Protection

CSAM Child Sexual Abuse Material

IEEE The Institute of Electrical and Electronics Engineers

IoT Internet of Things

KYC Know Your Customer

LLM Large Language Model

NSPCC National Society for the Prevention of Cruelty to Children

OECD The Organisation for Economic Co-operation and Development

SDK Software Development Kit

UNESCO The United Nations Educational, Scientific and Cultural Organization

UNICEF The United Nations International Children's Emergency Fund

5 Background

There has been a complete shift in the way people interact, work and learn as new technological developments have infiltrated and transformed the entire world. Digital technologies are pervasive and are impacting children like never before. Children are avid users and utilize a myriad of information and communication technologies in the metaverse. This shift means that children will be interacting with digital technologies in the metaverse for the rest of their lives despite the persistence of digital inequalities. The United Nations International Children's Emergency Fund (UNICEF's) *State of the World's Children 2017: Children in a Digital World report shows that one out of three internet users is younger than eighteen years of age and seventy one percent of 15-24 years olds are using social media, making them most connected group worldwide* [b-Lancet].

Spending such time online is laden with both probable risks and rewards for children. On one hand, children are offered unparalleled opportunities for learning, self-expression and consolidating friendships, while being exposed to harmful content and other risks in the metaverse [b-Livingstone]. There are smart toys which are outfitted with sensors, cameras, microphones and software to enable personalized experiences and responses in real time. Some toys are also programmed to teach children skills such as maths, drawing, coding, or a language and have also been found to enhance communication skills, especially amongst kids with intellectual disabilities [b-Ekin]. However, children are vulnerable to the risks posed by AI which includes bias, unfairness and lack of accessibility. Furthermore, big data produces a huge amount of data per second and it is vital to process this colossal amount of information quickly [b-ECPAT]. For instance, a law enforcement agency gets hundreds of millions of files on a suspect's computer. It is nearly impossible for them to process the data to see if there are any child sexual abuse materials CSAM(s). It would take longer time for a human team which may also have a negative impact on the welfare of the investigators. In such circumstances, an AI-based system may help in identifying and finding images and videos showing child-adult content. It also helps them to build the case quickly and identify the victims. Similarly, almost every social media platform also uses this kind of systems for the detection of child sexual abuse material (CSAM) [b-ECPAT]. Facial recognition has also been used to identify missing children whereas law enforcement agencies have managed to find child abuse images on seized devices with the use of deep learning [b-ITU-Webinar]. Parental control applications use AI for scanning millions of messages which are sent by children and teens.

The Massachusetts Institute of Technology (MIT) study revealed that children aged 6-10 believe that AI assistants are more of a genius than they are. Imagine the scenario in which digital companions have years of interaction with your child, gathering data, every step, every handshake, every search, transaction, share, and breath that not even a mother knows [b-Druga]. Furthermore, AI's growing use of biometric data such as facial recognition, sweat, or heart rate can show various ranges of physical, medical and emotional states. Voice data indicates intimate health information such as intoxication, arousal, fatigue, disease or depression [b-Mullin].

It is also true that not all children benefit from the metaverse equally. Children who are exposed to and vulnerable in offline environments tend to be more vulnerable in the metaverse. Despite these challenges, children's right to play and right to information are recognized internationally through the United Nations Convention on the rights of the child (United Nations Assembly, 1989). However, governments, policymakers, practitioners, teachers and parents still struggle to find a comprehensive mechanism for the protection, inclusion and fostering of digital skills and resilience for children in

the metaverse. The recent report of the EU fining a social media platform for breaching child online privacy is another example of how technology platforms are unable to comply with the existing regulations [b-Arab News].

This Report identifies the important and most common digital threats to children and the role of education and AI on how to protect them in the metaverse. It further provides essential information about existing policies and initiatives already in place for the protection of children. Lastly, it recommends guidelines for policymakers, and technology companies on how AI can be used to safeguard children's use of metaverse without hindering their engagement in the digital environment.

5.1 What is the metaverse?

The term 'metaverse' appeared for the first time in a novel published in 1992 [b-AJET]. Since the early 2000s, researchers have started to use this term to refer to digital technologies for learners to interact with other users with avatars.

The term came to prominence around 2020 due to the rebranding of Facebook. Since then, it has been popularised by the oligarchs of high-tech platforms. It relates to the use of advanced technology to create a platform providing a credible, immersive virtual world with properties that are at the discretion of those who provide this platform. The term reflects what already exists, in part is predictive, projecting from established high-tech platforms what future changes in technology and consumer behaviour might emerge. The metaverse depends on:

- State-of-the-art equipment, such as headsets.
- Flawless, two-way web-based communications, interacting with all our senses.
- End user conduct which accepts loss of control over personal data, certain rights (such as that to privacy), and is prepared to commit time and resources to participating in the metaverse activity.

The metaverse is immersive and it mirrors reality in its vivid, interactive processes. Its advantage is that it dramatically improves communications, granting access to the full, global domain. But there is also the danger that for individuals it may come to be confused with, or replace, reality, leading to manipulation, and a range of bad behaviours including criminal, sexual and political exploitation.

Children who are still learning the dimensions of reality may be prone to manipulation. Young people are becoming dependent on the nascent metaverse to access information and socialise. However, the metaverse can, at present, avoid established safeguarding mechanisms. The metaverse has great potential for improving connectivity, but this brings the dangers of misuse and manipulation. Children are relying heavily on social media technology, but with little guidance on how to avoid potential harms, including fraud and exploitation, or through peer group bullying.

6 Digital risks to children in the metaverse

Children spend more of their time in the metaverse in contrast to the last decade. The number of 12-15 year olds who possess smartphones has increased by more than 50% [b-OECD b]. The Internet exposes children to a myriad of opportunities but it also poses greater risks which may have a detrimental impact on their human rights. These risks may include data protection issues, cyberbullying, child sexual abuse content, and online hate speech. With the appropriate mechanism, AI can help in protecting a child online.

These risks are typically the digital version of traditional threats and just as in everyday life, a completely safe digital environment is unattainable. However, setting the boundaries and conditions for secure surfing is recommended [b-OECD a].

Table 1 – An overview of online risks and opportunities in the metaverse

Type of risk	The child is the:	Opportunities	Risks
Content	Recipient	• Getting advice on health or personal issues • Getting help from educational and informative resources	• Getting advertising and spam messages • Commercial advertising masquerading as news, or embedded marketing • Getting harmful content including violent, racist, pornographic, hateful
Contact	Participant	• Being in contact with others having similar interests • Sharing ideas and experiences with others • Participating in digital activities, games	• Being harassed, stalked or bullied • Being in contact with strangers or someone who impersonates • Getting influenced by an online fraud • Breach of data privacy, harvesting of personal information, personal data misuse
Conduct	Actor	• Engagement in civic activities • Collaborating and self-initiating learning • Developing content and generating different ideas and identities	• Being engaged in illegal digital activities including downloading malicious material or hacking • Bullying or harassing others • Generating or uploading harmful material (i.e., pornography) • Furnishing harmful advice (i.e., pertaining to suicide, eating disorders).

Source: [b-Staksrud], [b-OECD a].

6.1 Conduct risks

This refers to the child being in a peer-to-peer association when their own behaviour makes them vulnerable in the metaverse. These risks include hacking, bullying or harassing others, sharing or creating harmful material, and providing harmful suggestions (eating disorders or related to suicide) [b-OECD a].

6.2 Content risks

A content risk is a risk which makes a child an active participant in the digital market. These risks include receiving digital marketing messages which are illegal for children, and security risks such as digital scams, digital frauds, and malicious codes [b-OECD a].

6.3 Contact risks

This risk relates to a child as a victim of an interactive encounter. These risks include being stalked, or bullied, tracking personal data or harvesting personal information.

6.4 Cyberbullying

Cyberbullying refers to the aggressive targeting of a victim using cyberspace. It is similar to young people who are at risk from adult exploitation; children can also get intimate photos of a peer and share them with or without the individual's consent. In today's world, bullying is no longer left at school gates; cyberbullying is a novel method for individuals to bully, hurt, and humiliate their

victims with just a click of a button [b-Lancet]. According to a survey [b-Bozzola], around 60% of children who use social media have experienced some form of bullying. Today, social media and online games are virtual playgrounds where much cyberbullying is taking place. Children can be taken into a situation in the metaverse where they are ridiculed, or in an online game where player personas can be subject to attack, leading the game from online fun to a humiliating ordeal.

6.5 Cyber predators

Cyber predators stalk children in the metaverse and take advantage of the lack of adult supervision which can end up in children being lured into personal encounters. These predators are present on gaming platforms which appeal to children. There, they not only try to pull children in, in order to exploit children's innocence [b-Kaspersky].

6.6 Sexting or revenge porn

Sexting refers to sharing or creating sexually suggestive nearly nude or nude images online. Sexting is illegal if the doer is underage. In many cases, it leads to sextortion in which the child is threatened with exposure if they do not pay or perform a desired action. Revenge porn means posting or sharing nude images without consent. Both underage sexting and revenge porn are illegal in many countries [b-OECD a].

6.7 Security and privacy

Data privacy has become a global issue, and no one knows including children, how their data is being used by technology companies and nation states. The data is stored on servers around the globe. This is subject to phishing, surveillance, data processing and behavioural advertising based on personal information. It is imperative to keep children's data secure and make them understand the importance of digital privacy.

6.8 Posting private information

Children do not understand the sensitivity of metaverse boundaries. They may share their personal information such as home address, school name, and phone number which may be detrimental to their privacy if misused.

6.9 Digital scams

Children get attracted to offers which might seem of value to them such as free access to their favourite online game. Being unaware of social media scams, young minds are easily targeted by these frauds. With the technique of phishing, cyber criminals use popular social media platforms to identify potential victims and offer them rewards in return for personal information such as parents' credit card information.

7 The role of education in protecting children in the metaverse

The perils of metaverse use are widely documented, with an emerging consensus. No doubt new dangers will evolve. With regard specifically to the potential dangers that children may encounter the national society for the prevention of cruelty to children (NSPCC) in the United Kingdom (UK) has summarised the following threats to children.

- ***Fabricated or false news stories*** that might cause worry.
- ***Viral messages*** containing false information that can easily be shared.
- ***Challenge videos*** – the more outrageous or unbelievable a video, will often mean more views and exposure for the creator.
- ***Influencers*** advertising products or competitions.
- ***Meme accounts*** quickly spreading unverified facts.

- ***Opinions*** being shared as facts.
- ***Abusive comments*** and false allegations.
- ***Scam emails or messages*** sent to a personal device asking you to provide personal information or contain blackmail demands.

This list of threats is similar to the risks identified above on the role of AI.

The Future of Education and Skills 2030 [b-OECD c], states that over the last few years, there has been discussion around areas of international public policy such as data manipulation, sophisticated AI algorithms and machine learning. New policies and regulations are urgently required to limit the extensive criminal and anti-social behaviour which lurks in social media and the developing metaverse. However, it is unlikely that automated systems, no matter how subtle and advanced will be able to provide a complete solution, however welcome they are. There is a danger, too, that excessive, mechanical protection will rob children of any sense of agency. The metaverse will host the communications channels of the future, and children will need to learn how to navigate these challenges and avoid simply becoming passive recipients of overwhelming information flows. As with well-rehearsed arguments around free speech versus censorship, there comes a point where an individual must make a judgement, often supported by cultural traditions and evolving over time. The borderline between art and pornography is a case in point, where degrees of sophistication differ. Ethical standpoints are not fixed. Opinions vary between cultures, between generations, and between individuals. Given the many dangers inherent in the metaverse, it would be unwise to introduce a further threat that of AI telling a child what their ethical standpoint should be. 'Brainwashing' comes within the domain of dictatorship, tyranny and totalitarianism.

Regulations are in the short-term, inflexible, and that can be helpful. In many instances the same rules should apply to all. But freedoms are also important as are cultural, social and artistic norms. Potentially the metaverse can extend a child's learning and sophistication of judgement. Efforts to close this down must be resisted. Children should be free to explore and challenge. An ethical attitude towards AI is essential in any area of life and should form part of a student's training. This involves being able to evaluate knowledge, knowing what is legal or not and being able to decide when to use AI systems or discard them to avoid ethical biases. Children have the right to explore elements of and be educated in the rules and protocols of safe cyberspace.

To accommodate the metaverse it will be necessary to develop a new ethical context for education, and create an environment where AI is used to support students and teachers, and where we also train students to participate fully in a future world in which AI plays an increasingly important role.

The United Nations Educational, Scientific and Cultural Organization (UNESCO) has usefully recommended a rethink of the role of education at moments of societal transformation [b-UNESCO]. This Technical Report encourages reform in curricula and teaching methods taking into account three major changes that will impact our societies: globalization, the climate challenge and the digital revolution. These changes involve both technological and ethical aspects.

The metaverse and other advanced technologies are impacting many segments of our life. As a new normal emerges, innovation in skills training needs to encompass responsible use of AI, increasing the use of data mining, machine learning, robotics, digitization, and other specific branches of AI engineering. This should be included in the curriculum from an early age.

In addition, an increase in improvised student-teacher interactions on social media platforms and applications without prior rules or accepted protocols leaves open the potential for vulnerability regarding personal data and privacy protection. This may disparately and negatively affect the lives of children who are born in already difficult economic and adverse social conditions. Evidence of children's experiences in the face of permanent risks of cyberspace brings to the fore the urgency of policy development and decision-making that strengthen the best interests of children as defined in the convention on the rights of the child, the proposed legal framework on children's rights.

Recently, the institute of electrical and electronics engineers (IEEE) published a standard [b-IEEE 2089], that addresses age-appropriate design for children's digital services. This standard creates a framework for organizations to recognize and respond to the needs of children and youth. It is the first in a series of guidelines that will allow enterprises such as social media platforms, to design age-appropriate digital products and services to become safer and more secure. Recently California introduced an age-appropriate design code act [b-California Act]. These developments are entirely appropriate.

8 Role of AI for children safety

Artificial intelligence (AI) is critically important in child online safety especially in the metaverse. In the metaverse industry or market, where every organization has its own definition of the metaverse, it becomes critical to consider how children's safety will be addressed in these innovations.

8.1 Types of metaverses and technological constraints

In the universe of metaverse, several types of metaverses exist. In fact, metaverse has been treated very differently and became a subjective word between tech influencers, experts and organizations. This always happens with every technology that has stepped into the world of retail, banking or in consumer markets for the first time [b-Juego].

Regulators usually do not put any restrictions in the initial days of any technology, and they allow technology to blossom to see how it will work between different audiences. This is exactly the same phase where we are currently living because there is no absolute definition of the metaverse. Hence, there is no pattern, framework, protocol, or best practices that exist at the moment. This technology is currently being explored by different sectors and industries so the future might be different as compared to today's version of the metaverse.

An understanding of how AI can help in children's safety varies, based on the type of the metaverse. Currently, the following are the types of metaverses that exist.

1) Decentralized metaverse
2) Centralized metaverse
3) Hybrid metaverse
4) Private metaverse

8.1.1 Decentralized metaverse

These types of metaverses are fully operated without any restrictions or boundaries. Such metaverses include "Decentraland" which is run by a community. This is a total Web 3.0 approach where no central company holds any power to change anything inside the metaverse. It is driven by a community and their aim is to make this a global Internet browsing simulation world [b-Decentraland a].

However, restriction with respect to the age of participants is just the traditional paper-based format, where a user can input their fake date of birth and enter the metaverse easily. Further, there is no restriction or user know your customer (KYC) which makes every user anonymous in the world of metaverse because this is a decentralized autonomous organization (DAO) [b-Decentraland b].

Artificial intelligence can play a vital role in this type of metaverses, however, the community needs to construct a standard to address child safety using AI. Since this system is built using peer-to-peer blockchain technology, the implementation of an AI is extremely technically challenging.

8.1.2 Centralized metaverse

A centralized metaverse or permissioned metaverse is a virtual world that is owned and operated by a single entity. This means that the entity has complete control over the platform, including the rules,

regulations and content. The "Roblox" and "Meta Quest of Facebook" are the centralized type of metaverse which is far easier to streamline and maintained in terms of regulations [b-Ogundare]. The decentralized community usually considered such metaverses as just a game instead of a metaverse. So, the dispute on the word "metaverse" is very subjective.

Artificial intelligence can easily be implemented in this kind of metaverse in addition to KYC or strict verification. AI can be much faster and protect user privacy in much greater detail as compared to the public blockchain based metaverses. However, any kind of data leak in a single entity ownership can raise major concerns as well. This means AI does not need to work only on the frontend side but also it must put restriction boundaries on the backend side of the system e.g., renderer, database, files, metadata, etc.

8.1.3 Hybrid metaverse

A hybrid metaverse is a virtual 3D world that blends the elements of both centralized and decentralized metaverses. This means that it has some features that are controlled by a single entity, such as a company or organization, and other features that are controlled by the users or its decentralized community. Such a model demonstrates the power of Web 3.0 with traditional Web 2.0 control on its users. The metaverse "Sandbox" is an example of a hybrid model where a single entity controls the power over its users and can manually ban any user [b-Medium]. To ensure a safe environment, the metaverse must prioritize security, privacy and governance [b-Sandbox]. However, the digital assets created by users will be 100% under the users' control using the Web 3.0 philosophy.

Such metaverses can implement AI bots much more easily especially in terms of regulating children's privacies. These bots can operate under the company's server with high speed as compared to peer-to-peer.

8.1.4 Private metaverse

A private metaverse is one where the public is not allowed or is very restricted to some part of the services. This is fully governed by an entity and such examples could include corporate metaverses, educational metaverses, governmental metaverses and private social metaverses.

Educational metaverses are metaverses that are created and owned by educational institutions for teaching and learning purposes. They are often used to create immersive learning experiences that are not possible in the real world. For example, the University of California, Irvine has created a metaverse called CAVE2™ where students can learn about anatomy and physiology.

Such metaverses can also utilize AI technology to verify user data especially using biometrics to avoid any child interaction. However, in the future there may be educational institutes like schools providing metaverse based education around the world. In this case, there must be an AI based verification of children's attendance using parents' permission through some sort of a Web or mobile application.

8.2 Central and decentralized AI limitation and possibilities

The centralized aspect is much easier in terms of the implementation of AI. However, even in the decentralized metaverse some part of the AI can be implemented. Following is a Table which depicts the AI feature with respect to major two types of metaverses.

Table 2 – Central and decentralized AI

Implementing feature description	Centralized metaverse	Decentralized metaverse
Content moderation scanning to protect children	Yes	Yes
Age verification using an AI bot or process	Yes	No
Personalized safety recommendations or behavioural restriction	Yes	No
Safety education for children	Yes	Yes
Smart contracts to track the addresses of the users	Yes	Yes

Source: Author.

8.3 Cloud based service providers with metaverse on-demand services

The cloud-based services provider especially Amazon Web Services (AWS), Azure and Google can provide metaverse child protection services if implemented with some kind of protocols or framework. These services are the major foundation in supporting metaverse rendering and execution. For instance, many blockchains are being implemented using cloud services from AWS or Azure [b-101 Blockchains]. Due to its vast requirements in the market, AWS and other cloud providers have offered services like managed blockchain networks where the end customer or client can implement their services without worrying about the network implementations from scratch.

These service providers can play a vital role collectively by implementing smooth KYC verification to detect the user type. This could include strict or non-strict verification using AI technologies.

They can provide their software development kit (SDK) to detect and verify content and user interaction including behavioural monitoring. Currently, such systems do not exist in the managed technological stacks.

8.4 Cybersecurity attacks and issues in metaverses directed against children

Like any digital cloud or system, metaverses are also not immune from cyber-attacks. These attacks can be stealing children's data from the database, importing children's pictures which were used for any digital twin or avatar, email addresses, locations or IP addresses, etc.

A centralized metaverse has the benefit that its security protocols can be easily maintained and implemented. However, in the event of a cyberattack, the centralized data can be seriously compromised if it is not properly encrypted. Furthermore, even encryption can be vulnerable if outdated encryption methods are used because quantum computing technology is advancing rapidly. This is why it is important to use quantum-resistant encryption algorithms to protect the centralized data more effectively. Nevertheless, data protection measures for children are still somewhat at risk.

AI can play a vital role in an overall metaverse network or platform, to detect and prevent privacy or data leaks. For instance, a child should automatically be prevented from creating its avatar based on

a picture. So, the picture will not travel from the device to the network at all. Similarly, the AI should not allow a child to open their camera or chat mode without parental supervision. These precautionary AI techniques can help minimize the storage of children's data, thereby providing a level of protection against potential damage during a cyber-attack.

9 Global policies and initiatives for protecting child online

9.1 The child online protection (COP) initiative

The child online protection (COP) is an initiative launched by ITU in November 2008 aiming to bring forth stakeholders from communities across the globe. The mandate was to develop an empowering and a safe digital experience for children around the world. This initiative offers a holistic approach in promoting child safety in cyberspace and creates strategies in five key areas: legal measures, technical and procedural measures, organizational structures, capacity building, and international cooperation [b-ITU COP].

9.2 The White House guidance for the regulation of artificial intelligence

The blueprint for an AI bill of rights is a guide issued by the White House office of science and technology policy. It has five principles which guide the design, use, and deployment of AI systems to protect US citizens in the age of artificial intelligence. This is a guide for people that protects them from these threats and uses technology ensuring the highest values [b-WH].

9.3 The EU Artificial Intelligence Act

The EU artificial intelligence act is the first law on AI in the world that aims in facilitating a single market for AI based applications. It sets out general guidelines to apply on AI-driven applications, products, systems and services within the EU territory with the focus to protect the fundamental rights of its citizens [b-EU].

9.4 The UK ICO's age appropriate design code

The UK ICO's age appropriate design code identifies 15 standards which online services need to follow and comply with under the data protection law to protect children's data online. The online products and services which are covered by the code are wide-ranged including games, applications, news services, connected devices and toys for children [b-ICO].

9.5 The UNICEF's AI policy for children

UNICEF's policy describes the significance of promoting children's development in AI practices and suggests recommendations for industry and governments. The policy aims in bringing a balanced perspective with implementable principles that support children's rights [b-UNICEF a].

This policy offers a human-centric AI approach by introducing a child rights lens. This is an important document in terms of the empowerment of children once they are interacting with AI based systems. It contains a range of AI systems that actually impact children and discusses in what ways AI systems should promote child safety in the digital world and empower them to participate in the AI world by contributing to the use of AI.

9.6 The OECD recommendation on children in the digital environment

The organisation for economic co-operation and development (OECD) recommendation on children in the digital environment was adopted in May 2021. It focuses on assisting countries to find a balance between protecting children from digital risks and promoting the benefits offered by cyberspace [b-OECD d].

10 Policy proposals

This Technical Report presents various policy-making roadmaps on how to protect children in the metaverse. We have categorized the recommendations at two levels: strategic and technical.

10.1 Strategic level proposals

10.1.1 Role of education in protecting children in the metaverse

Young people have become increasingly dependent on the Internet as a means of accessing information. It is convenient and accessible. It can provide access to education for hard-to-reach groups. However, web-based information can also by-pass the established protocols defining expertise and authenticity. Funded by sizeable marketing budgets or organised crime, fake news is increasingly credible and pervasive. So, the curriculum must include guidance on how to question and test authenticity, how to spot fakes, and resist peer group chit-chat. Bad actors with a manipulative or exploitative disposition will always discover new scams and routes to exploitation. As things stand, the Internet enhances opportunities for anonymity, and with anonymity comes the opportunity for objectionable behaviours. It is important that we look at how learning to discriminate between fact and fiction might be incorporated more widely into the curriculum.

Such a curriculum should include:

1) A clear identification of the developing range of threats to children's understanding, and the gaps in protective provision.
2) An educational syllabus for each stage of education that empowers students to evaluate, investigate and challenge fictional or manipulative material.
3) Broadening critical thinking skills to identify fake online content, including disinformation, misinformation, and online deceit.

As children mature, they need to have developed their own self-protection in a world that values free speech. They cannot, and should not, be sheltered forever. Snower and Twomey (2020) in their work on humanistic digital governance take a realistic approach to the wider issue of cyber-security. They make the point that communications on the Internet lack the social norms that constrain interactions in any normal society. Therefore, it is important that teachers help young people to develop new norms appropriate to the metaverse.

Increasingly the term 'media literacy' is being used McDougal et al., (2018) note that *'policy initiatives on media literacy and media education have been growing across Europe and the English-speaking world for a few decades. Recent research at EU level has provided useful evidence on the role of informal media education and formal media education to acquire media literacy competencies*. They found that *teaching and learning practices for media literacy education can involve various classroom-based methods most of which are based on active learning*. They describe effective practice as:

- Analysis and evaluation: the capacity to comprehend messages and use critical thinking and understanding to analyse their quality, veracity, credibility and point of view, while considering their potential effects or consequences.
- Creation: the capacity to create media content and confidently express oneself with an awareness of purpose, audience, and composition techniques.
- Reflection: the capacity to apply social responsibility and ethical principles to one's own identity, communication, and conduct, to develop an awareness of and to manage one's media life.
- Action/agency: the capacity to act and engage in citizenship through media, to become political agents in a democratic society.

These are skills that go beyond the boundaries of cyber-space. They are fundamental capacities for survival and prosperity in a world of artificial intelligence.

10.2 Technical level proposals

10.2.1 Responsible use of AI in protecting children in the metaverse

Several types of metaverse exist in the market and if they follow a certain protocol or framework then it will drastically help in maintaining security for children. To streamline these protocols artificial intelligence can play its role in regulating children's data privacy and protection.

Here are some of the ways that AI can be used to protect children in the metaverse:

10.2.1.1 Moderating visual and text contents

AI can be used to scan for and remove harmful content from the metaverse, such as child sexual abuse material, hate speech, and cyberbullying [b-Wired]. For this kind of technique, a system can be built using large language model (LLM) based AI model specifically looking for "children abusing content". This model should be trained on abusive words as well as on visual content scanning. The visual content may include videos, pictures, textures and 3D simulations.

10.2.1.2 Age verification process

There are certain places where the metaverse only allows age-restricted content. These contents may include adult concerts or events or places where the age restriction should be above 18. Technically such spaces in 3D simulations are built without any proper protocol or age verification. This means a child can easily fake their date of birth and enter the space without any proper verification protocols.

AI can be used to verify the age of users, which can help to prevent children from accessing inappropriate content or being exposed to harmful experiences [b-Wheeler]. To control such age restriction may require a tremendous amount of debate and control system on whether a decentralized or centralized KYC system can be implemented inside the metaverse framework. In addition to this, how can one maintain the same procedure on several devices for instance headset, mobile and Web in unified protocols? The AI component can restrict and maintain discipline between cross platforms much better than any other technology in terms of verifying age restrictions.

10.2.1.3 Personalized behavioural monitoring

The user experience and behaviour pattern are the most important data which can be utilized to monitor child safety [b-Convizit]. These data can be maintained and collected while allowing children to browse the metaverse regardless of the type of metaverse they are using. AI can be used to generate personalized safety recommendations for users, based on their individual needs and risk factors.

For example, AI could recommend that a child who has been bullied in the past should avoid certain areas of the metaverse or only interact with friends they know in real life. Additionally, if such a child is surfing and browsing the metaverse area which could be harmful to them AI can automatically block them from accessing it.

10.2.1.4 Safety proposals alerts and resources

AI can be used to create educational resources about online safety for children and parents. These resources could teach children about the risks of the metaverse and how to stay safe, and they could also help parents talk to their children about online safety.

10.2.1.5 AI based child parental controls and gaming platforms

AI based parental control systems specifically written for gaming and Web 3 metaverses can be very helpful for children's privacy and protection. Registration on any metaverse can provide parent and child access separately. Here parents can fix the threshold of security and privacy of their children.

10.2.1.6 Censoring content filtering

AI can be used to filter out inappropriate content from gaming platforms, such as violence, gore, and explicit language. A child player will not face these words if a filter is explicitly set by the parent.

10.2.1.7 Location tracking

AI can be used to track the location of children while they are gaming so that parents can know where they are. This can help parents in knowing that a child is not changing their location while playing the metaverse. Otherwise, service can be halted based on the filter adjusted by any parent.

10.2.1.8 Purchase monitoring

AI can be used to monitor children's purchases on gaming platforms so that parents can prevent them from making unauthorized purchases. For instance, Roblox allows children to spend money in buying digital assets such as armour, etc. However, parents are not notified explicitly about what type of digital asset a child is purchasing. These alerts and notifications can help parents be vigilant about any harmful item. There is a case when a child bought something for 2 500 pounds without the parent's knowledge inside a Roblox metaverse [b-Standard]. AI can tackle such things much faster and easier if it is properly set up by the metaverse company.

10.2.1.9 Chat filtering and behavioural analysis

AI can be used to filter out inappropriate content from chats in gaming platforms such as bullying and hate speech. It can also analyse children's behaviour in gaming platforms, so that parents can identify any potential problems.

10.2.1.10 Parental education

AI can be used to provide parents with educational resources about online safety so that they can learn how to keep their children safe. Security awareness is a topic that is supposed to be enforced by these platforms so they should always be aware of the best practices [b-Aura].

11 Conclusion

The metaverse is full of opportunities and challenges at the same time. This Technical Report proposes that regulators and policymakers adjust and asses the regulatory framework and regulations to make sure that the metaverse-related technologies do not violate children's rights. The AI-based systems should be adopted to foster enabling environments for children through promoting inclusion, child rights by design and research. All stakeholders should sit together and work to protect children in the metaverse and address harmful content and behaviour. This is the need of the hour that technology companies should take proactive steps to understand the privacy challenges in the metaverse and to adopt the best practices in order to ensure responsible use of AI for children in the metaverse.

Bibliography

[b-ITU-T M.3080] Recommendation ITU-T M.3030 (2021), *Framework of artificial intelligence enhanced telecom operation and management (AITOM)*.

[b-ITU-T Y.4000] Recommendation ITU-T Y.4000/Y.2060 (2012), *Overview of the Internet of things*.

[b-ITU COP] ITU (2008), *Child Online Protection*.

[b-ITU-Webinar] ITU-Webinar (2022), *Keeping our children safe with AI,* AI for Good.

[b-ISO/IEC 2382] ISO/IEC 2382:2015, *Information technology – Vocabulary*.

<https://www.iso.org/standard/63598.html>

[b-IEEE 2089] IEEE 2089-2021, *IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children.* <https://standards.ieee.org/ieee/2089/7633/>

[b-101 Blockchains] 101 Blockchains. (2021), *AWS vs AZURE vs Oracle Blockchain Solution Offering: The BaaS Comparison*.

<https://101blockchains.com/aws-vs-azure-vs-oracle-blockchain/>

[b-Arab News] Arab News (2023), *EU hits TikTok with big fine over child data.* <https://www.arabnews.com/node/2374266/media>

[b-AJET] Australasian Journal of Educational Technology. (2022), *Special Issue: Achieving lasting education in the new digital learning world*. Volume 38 No. 4.

<https://ajet.org.au/index.php/AJET/issue/view/156>

[b-Aura] Aura. (2023), *Online Gaming Safety for Kids: What Parents Need To Know*.

<https://www.aura.com/learn/online-gaming-safety>

[b-Ball] Ball, M. (2022), *The Metaverse: And How it will Revolutionize Everything.*

<https://www.amazon.in/Metaverse-How-Will-Revolutionize-Everything-ebook/dp/B09KMWYHX8>

[b-Barrera] Barrera, O., Guriev, S., Henry, E., and Zhuravskaya, E. (2020), *Facts, alternative facts, and fact checking in times of post-truth politics*, Journal of Public Economics. Elsevier Volume 182(C).

<https://ideas.repec.org/a/eee/pubeco/v182y2020ics0047272719301859.html>

[b-Bozzola] Bozzola, E., Spina, G., Agostiniani, R., Barni, S., Russo, R., Scarpato, E., Di Mauro, A., Di Stefano, A.V., Caruso, C., Corsello, G., and Staiano, A. (2022), *The Use of Social Media in Children and Adolescents: Scoping Review on the Potential Risks*. International Journal of Environmental Research and Public Health. 19(16).

<https://doi.org/10.3390%2Fijerph19169960>

[b-Brodsky] Brodsky, B. (2022), *Amazon's New AI Art Tool Can Help Kids Make Art, or Stifle Creativity.* Lifewire.

<https://www.lifewire.com/amazons-new-ai-art-tool-can-help-kids-make-art-or-stifle-creativity-6833604>

[b-California Act] The Californian Act (2021), *The California Age-Appropriate Design Code Act*. Assembly Bill No. 2273.
<https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202120220AB2273>

[b-Chalmers] Chalmers, D. (2022), *What Should be Considered a Crime in the Metaverse*?
<https://www.wired.com/story/crime-metaverse-virtual-reality/#:~:text=As%20the%20experience%20of%20virtual,counterparts%20in%20the%20physical%20world.&text=This%20story%20is%20adapted%20from,Chalmers.>

[b-Combs] Combs, V. (2021). *The Metaverse: What is it*?
<https://www.techrepublic.com/article/metaverse-what-is-it/>

[b-Convizit] Convizit, *AI-driven Behavioral Data Capture*.
<https://convizit.com/product/automatic-behavioral-data/>

[b-Decentraland a] Decentraland, *Introduction*.
<https://docs.decentraland.org/player/general/introduction/>

[b-Decentraland b] Decentraland, *Community Grants*.
<https://docs.decentraland.org/player/general/dao/grants-v1/community-grants/>

[b-Druga] Druga, S., and Williams, R. (2017), *Kids, AI devices, and intelligent toys*, MIT Media Lab.
<https://www.media.mit.edu/posts/kids-ai-devices/>

[b-Dumitru] Dumitru, E-A. (2020), *Testing Children and Adolescents' Ability to Identify Fake News: A Combined Design of Quasi-Experiment and Group Discussions*.
<https://www.mdpi.com/2075-4698/10/3/71>

[b-Ekin] Ekin, C.Ç., Çağiltay, K., and Karasu, N. (2018), *Usability study of a smart toy on students with intellectual disabilities*. Journal of Systems Architecture. Volume 89, pp 95–102.
<https://scholar.google.com/citations?view_op=view_citation&hl=en&user=CQ_GvSwAAAAJ&citation_for_view=CQ_GvSwAAAAJ:hqOjcs7Dif8C>

[b-EU] European Commission. (2021), *A European approach to Artificial Intelligence*.
<https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>

[b-ECPAT] ECPAT (2021), *The role of Artificial Intelligence in protecting children in the digital space*.
<https://ecpat.org/ai-digitalspace/>

[b-Herrero-Diz] Herrero-Diz, P., Conde-Jiménez, J., and Reyes de Cózar, S. (2020), *Teens' Motivations to Spread Fake News on WhatsApp*.
<https://journals.sagepub.com/doi/pdf/10.1177/2056305120942879>

[b-ICO] ICO. (2020), *Age appropriate design: a code of practice for online services*. 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>

[b-Juego] Juego (2022), *The Various Types of Metaverse*.
<https://www.juegostudio.com/blog/the-various-types-of-metaverse>

[b-Kaspersky] Kaspersky (n.d), *Internet safety for kids: How to protect your child from the top 7 dangers they face online.*

<https://usa.kaspersky.com/resource-center/threats/top-seven-dangers-children-face-online>

[b-Koduri] Koduri, R. (2021), *Powering the Metaverse.*

<https://www.intel.com/content/www/us/en/newsroom/opinion/powering-metaverse.html#gs.94xyoo>

[b-Lancet] The Lancet Child & Adolescent Health (2018), *Growing up in a digital world: benefits and risks*. Editorial, Volume 2, Issue 2, p 79.

<https://doi.org/10.1016/S2352-4642(18)30002-6>

[b-Livingstone] Livingstone, S., Haddon, L., and Görzig, A. (2012), *Children, risk and safety on the internet: research and policy challenges in comparative perspective*. LSE Research Online.

<https://eprints.lse.ac.uk/44761/1/EUKidsOnlinebookExecSummary.pdf>

[b-McDougall] McDougall, J., Zezulkova, M., van Driel, B., Sternadel, D. (2018), *Teaching media literacy in Europe: evidence of effective school practices in primary and secondary education.*

<https://nesetweb.eu/wp-content/uploads/2019/06/AR2_Full_Report_With_identifiers_Teaching-Media-Literacy.pdf>

[b-Medium] Medium. (2022), *The Sandbox Alpha Season 2 Security Update*.
<https://medium.com/sandbox-game/the-sandbox-alpha-season-2-security-update-136689c19d2d>

[b-Müller] Mueller, K., and Schwarz, C. (2020), *Fanning the Flames of Hate: Social Media and Hate Crime.*

<https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3082972>

[b-Mullin] Mullin, E. (2017), *Voice analysis tech could diagnose disease.* MIT Technology Review.

<https://www.technologyreview.com/2017/01/19/154498/voice-analysis-tech-could-diagnose-disease/>

[b-OECD a] OECD (2021), *21st Century Children.*

<https://www.oecd.org/education/ceri/21st-century-children.htm>

[b-OECD b] OECD (2012), *OECD Recommendation on Children in the Digital Environment.*

<https://www.oecd.org/digital/children-digital-environment/#:~:text=OECD%20Recommendation%20on%20Children%20in%20the%20Digital%20Environment&text=The%20Recommendation%20sets%20out%20principles,importance%20of%20international%20co%2Doperation.>

[b-OECD c] OECD (2018), *The Future of Education and Skills - Education 2030*.
<https://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf]>

[b-OECD d] OECD (2020), *Protecting children online*.
<https://www.oecd.org/education/protecting-children-online-9e0e49a9-en.htm#:~:text=Protecting%20children%20online-,An%20overview%20of%20recent%20developments%20in%20legal%20frameworks%20and%20policies,and%20inappropriate%20contact%20with%20strangers.>

[b-Ogundare] Ogundare, I. (2023), *Centralized vs Decentralized Metaverse: Complete Guide*.

<https://www.coinspeaker.com/guides/centralized-vs-decentralized-metaverse-complete-guide/>

[b-Pennycook] Pennycook, G., McPhetres, J., Zhang, Y., Lu, J.G., and Rand, D.G. (2020), *Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention*, Psychological Science, 31(7):770–780.

<https://pubmed.ncbi.nlm.nih.gov/32603243/>

[b-Ritterbusch] Ritterbusch, G.David., and Teichmann, M.R. (2023), *Defining the Metaverse: A Systematic Literature Review*.

<https://ieeexplore.ieee.org/document/10035386>

[b-Standard] Standard. (2023), *Essential Roblox parental-control settings, after 10-year-old spends £2,500-plus on game*.

<https://www.standard.co.uk/tech/gaming/roblox-app-parental-controls- settings-purchases-b1082812.html>

[b-Sandbox] The Sandbox.

<https://www.sandbox.game/en/>

[b-Staksrud] Staksrud, E., Livingstone, S., Haddon, L., and Ólafsson, K. (2009), *What do we know about children's use of online technologies?:* a report on data availability and research gaps in Europe [2nd edition].
<http://eprints.lse.ac.uk/24367/1/What%20do%20we%20know%20about%20children's%20use%20of%20online%20technologies(lsero).pdf>

[b-Tucker] Tucker, J.A., Andrew, G., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., and Nyhan, B. (2018), *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*.

<https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3144139>

[b-UNESCO] UNESCO (2021), *Reimaging our futures together: a new social contract for education*. International Commission on the Futures of Education.
<https://unesdoc.unesco.org/ark:/48223/pf0000379707>

[b-UNICEF a] UNICEF Version 2.0. (2021), *Policy guidance on AI for children*.
<https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdfs>

[b-UNICEF b] UNICEF (2015), *Guidelines for Industry on Child Online Protection*.

<https://www.unicef.org/media/66616/file/Industry-Guidelines-for-Online-ChildProtection.pdf>

[b-WH] The White House (2022), *Blueprint for an AI bill of rights*.
<https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf>

[b-Wheeler] Wheeler, J. (2024), *Age Verification vs. Age Gating: How AI Aids Online Minor Safety.*
<https://www.jumio.com/age-gating-age-verification>

[b-Wired] Wired (2023), *Apple Expands Its On-Device Nudity Detection to Combat CSAM*.

<https://www.wired.com/story/apple-communication-safety-nude-detection/>

[b-Wood] Wood, R. (2021), *This Oculus VR headset could feature lifelike resolution – here's why that matters*.

<https://www.techradar.com/news/this-oculus-vr-headset-could-feature-lifelike-resolution-heres-why-that-matters>

[b-Zhuravskaya] Zhuravskaya, E., Petrova, M., and Enikolopov, R. (2020), *Political Effects of the Internet and Social Media*. Annual Review of Economics.

<https://www.annualreviews.org/content/journals/10.1146/annurev-economics-081919-050239>