METAVERSE SAFETY WEEK

4th annual edition | DEC 10-15, 2023

CHILD SAFETY AND CHILDREN'S RIGHTS

SAFEGUARDING THE YOUNG GENERATION IN THE AI-AUGMENTED WORLDS

DECEMBER 12, 2023
ROUNDTABLE REPORT

CO--ORGANIZED BY:

unicef

CHILD SAFETY INITIATIVE

SUPPORTED BY:

∞ Meta

www.xrsi.org   www.MetaverseSafetyWeek.org

TABLE OF CONTENTS

EXECUTIVE SUMMARY

In an era where AI-augmented realities are rapidly transforming the digital landscape, the protection and empowerment of our young generation emerge as a crucial imperative. The duality of these evolving environments, blending unparalleled opportunities with profound challenges, necessitates a vigilant and proactive approach to child safety. As immersive technologies increasingly captivate younger audiences, the responsibility to safeguard their digital well-being intensifies. It underscores the complexities of nurturing and protecting young minds in AI-augmented worlds. It emphasizes the need for proactive, empowering, and ethics-driven efforts across various sectors, including technology, education, policy, and child advocacy, to ensure that these burgeoning virtual spaces serve as platforms for positive growth and learning rather than avenues of risk and harm.

Emerging Concerns and Collaborative Solutions:

In our rapidly changing digital world, children's increasing immersion in virtual environments brings forth a spectrum of educational possibilities and substantial risks, including psychological effects and exposure to harmful content.[1] The interaction between young users and AI technologies calls for vigilant measures to prevent dependency and manipulation, highlighting the critical importance of integrating digital literacy and AI education into school curricula. Ensuring the safety of children's online experiences requires embedding robust AI safety features in technologies, fostering diverse input in AI design for equitable development, and enacting strong legislative and privacy protections. Additionally, leveraging the influence of digital influencers for educational outreach is critical in confronting these challenges collaboratively. We must commit to immediate, practical actions and strategies to protect the well-being and safety of children in the digital realm.

1. Kuhne, C., Kecelioglu, E. D., Maltby, S., Hood, R. J., Knott, B., Ditton, E., Walker, F. R., & Kluge, M. G. (2023). Direct comparison of virtual reality and 2D delivery on sense of presence, emotional and physiological outcome measures. Frontiers in Virtual Reality, 4. https://www.frontiersin.org/articles/10.3389/frvir.2023.1211001/full#:~:text=Directly%20comparing%20VR%20and%202D, a%20first%2Dperson%20shooter%20game

Actionable Strategies and Collective Responsibility:

In the realm of AI and digital safety, accountability is not the domain of a select few; it is a collective imperative, cutting across various roles from technologists and educators to policymakers and child rights advocates. This diverse community of interest is pivotal in molding AI into a tool for empowerment rather than a risk. The path to ensuring this lies in robust dialogues among stakeholders, focusing on pragmatic solutions. It includes advocating for mandatory AI safety features, fostering transparency in AI interactions, and involving children in decision-making processes to safeguard their digital experiences. The breadth of challenges necessitates a united approach, with tech companies, policymakers, and educators collaborating to forge a safer online world for children.[2] This collaborative framework must prioritize actionable strategies, underscoring our shared responsibility in cultivating a secure and enjoyable digital landscape for the young generation.

Strategic Intelligence Gathering by UNICEF and Child Safety Initiative:

The Child Safety and Children's Rights roundtable, jointly organized by XRSI Child Safety Initiative and UNICEF, served as a pivotal platform for strategic discussions among global stakeholders, including technologists, educators, and policymakers. Aiming to address children's intricate challenges in AI-augmented environments, the roundtable focused on advancing digital literacy, promoting safe online practices through influencer outreach, and embedding robust AI safety features in technologies. Emphasizing the need for diverse perspectives in AI development, the discussions fostered a collaborative framework for creating safer online spaces for children. This event marked a significant stride in establishing comprehensive, child-centric strategies for the responsible evolution of digital technologies and AI-powered environments.

2. eSafety Commissioner. (2023, December). eSafety Metaverse Report. eSafety.
https://www.esafety.gov.au/sites/default/files/2023-12/eSafety-Metaverse-Report.pdf

The Child Safety & Children's Rights Roundtable, a key component of Metaverse Safety Week, brought together a diverse group of experts to address this crucial issue. The report encapsulates the discussions and insights from the roundtable, emphasizing the collective effort needed to protect and empower our digital youth in AI-augmented immersive worlds.

The multi-stakeholder experts and professionals sought to identify challenges and delineate real-world short-term solutions and actions.

Participants collaborated to explore practical strategies to manage the previously described challenges. Emphasizing the urgency of proactive measures, the objective was to propose actionable steps that could be swiftly implemented to mitigate risks and enhance the safety of the young generation within the metaverse. This emphasis on real-world applicability underscored a commitment to translating discussions into tangible outcomes, ensuring that the Metaverse evolves as a space where the well-being of its youngest denizens is actively safeguarded.

Valentino Megale
CSI Program Lead, XRSI
Chair - Child Safety and
Children's Rights, MSW2023

Hlekiwe Kachali
Advisor, UNICEF
Co-Chair - Child Safety and
Children's Rights, MSW2023

INTRODUCTION

On **December 12**, a landmark event unfolded: **The Global Roundtable Discussion on Child Safety & Children's Rights**. This pivotal gathering, co-organized by **UNICEF** and **XRSI Child Safety Initiative**, brought together various stakeholders, including parents, educators, child psychologists, technology companies, game developers, and children's rights organizations. The focus of this roundtable was to navigate the **complexities of immersive digital environments** that are increasingly captivating a young audience.

In an era where virtual worlds and emerging technologies are becoming a staple in the lives of the younger generation, this roundtable shed light on the various issues, concerns, and risks that arise from their interaction with these digital realms.

The discussions aimed to go beyond theoretical discourse, focusing on tangible actions and strategies that could be immediately applied to safeguard the well-being of children in digital environments. This global roundtable discussion represented a significant step forward in the collective effort to protect and empower children in the digital age. By combining diverse perspectives, expert knowledge, and the innovative use of Swarm AI® technology, the session marked a crucial stride towards ensuring that children's rights and safety remain at the forefront of technological advancement in our increasingly digital world.

KEY OBJECTIVES INCLUDED

- Provide an overview of the expanding young audience of **Immersive digital environments**.

- Share and discuss practical examples of issues, concerns, and risks for the young audience from the **access and use of virtual worlds** and emerging technology.

- Describe **real-world short-term solutions and actions** to manage the previously described challenges.

INTRODUCING SWARM AI

Metaverse Safety Week 2023 elevated the roundtable experience by integrating **Swarm AI®** technology from **Unanimous AI**. The innovative approach combined real-time human insights with AI algorithms, inspired by nature's swarm intelligence, to amplify collective decision-making. Participants engaged in a dynamic voting process, contributing to decisions that reflect a more profound collective wisdom for safeguarding the interests of AI and Emerging Technologies.

SWARM INTELLIGENCE

via **UNANIMOUS AI**

Swarm AI® technology, developed by <u>Unanimous AI</u>, employs a unique combination of real-time human input and AI algorithms that are modeled after swarms in nature. Swarm Intelligence is the reason why birds flock, bees swarm, and fish school – they are smarter together than alone. Nature shows us that by forming closed-loop systems, groups can produce insights that greatly exceed the abilities of any individual member. While humans have not evolved this ability naturally, Swarm AI technology enables this artificially, allowing groups to amplify their intelligence by forming real-time swarms.

> *"Now that the metaverse is no longer science fiction, it's time to get serious about making it safer, especially for children. eSafety commends the ongoing work of XRSI to raise awareness and foster global collaboration and looks forward to continuing this important conversation and working collectively in 2023 and beyond."*
>
> **- Julie Inman-Grant, eSafety Commissioner, Australia**

THE ROUNDTABLE OVERVIEW

On December 12, 2023, UNICEF and XRSI Child Safety Initiative organized a strategic intelligence roundtable focused on **child safety and children's rights** as a part of the 2023 Metaverse Safety Week, an annual awareness campaign to promote a safe and positive experience within immersive environments. The primary goal of this year's campaign was to explore the intersections of AI and emerging technologies and raise awareness about the importance of building safe experiences and promoting responsible behavior within the Metaverse. The Roundtable was three hours of exciting and engaging discussions from experts in various areas who brought their experience about "Safeguarding the young generation in the AI-augmented worlds" to the audience.

The hosts, **Valentino Megale** and **Hlekiwe Kachali**, were joined by the following individuals for the discussion. **Kavya Pearlman** also supported the discussion, with support from several XRSI Team Members and Advisors in the background.

The conversation started with distinct opening remarks from **Catherine Knibbs**, Online Harms Consultant, Psychotherapist, and FRSA Human Behaviour Technologist. Catherine Knibbs opened the event by focusing on safeguarding children, particularly the vulnerable, in a rapidly evolving digital space. The current emphasis is on biopsychosocial and psychological concerns, aiming to equip children with tools and resilience to navigate this dynamic environment. Knibbs acknowledged the lack of robust literature on child development in the digital age and stressed the need for professionals, including therapists and youth services, to collaborate globally using accessible and unifying language. The discussion addresses how children can safely engage in online communities, prevent harm, and establish supportive systems following incidents of harm. Knibbs, an advocate for children who have experienced harm, highlighted the importance of a cohesive understanding among professionals in various settings.

The following distinguished panelists brought forth their knowledge and explored actionable strategies and collaborative initiatives to ensure that our AI-augmented realities prioritize every child's safety, dignity, and rights.

Catherine Knibbs
FRSA Human Behaviour Technologist

Dona J. Fraser
BBB National Programs

Lisa Robinson
OECD

Steven Vosloo
UNICEF

Jessica Stone
Virtual Sandtray, LLC

Girard Kelly
Common Sense Media

Janaé N. McPhaul
XRSI | Cyber XR Coalition

Anne Hobson
Meta

Larry Magid
ConnectSafely

Hlekiwe Kachali, Advisor UNICEF, started the conversation with a quote that directed the roundtable and focused on shared values before shared responsibilities for charting a safe path forward for the young generation in immersive digital environments.

> *"In the dynamic world of the Metaverse and Artificial Intelligence, UNICEF champions the safety and well-being of children as a paramount concern. Our collaborative efforts with technology developers, policymakers, and guardians are essential in shaping these digital spaces to be inclusive, secure, and enriching for children. Our priority is to ensure that as these technologies evolve, they do so with children's best interests at their core, cultivating shared values, encompassing safety, privacy, and inclusivity. This collective responsibility is the cornerstone of our mission to create an innovative, safe, and empowering digital future for every child."*
>
> **- Hlekiwe Kachali, Advisor UNICEF**

3. Common Sense, Privacy of Virtual Reality: Our Future in the Metaverse and Beyond (November, 2022) https://www.commonsensemedia.org/sites/default/files/research/report/privacy-of-virtual-reality-our-future-in-the-metaverse-and-beyond.pdf

In a thought-provoking opening, the importance of age-appropriate content and a nuanced understanding of data collection and usage in decision-making took center stage. **Girard Kelly,** the Head of Privacy at Common Sense Media, underscored the pressing need for transparency in VR privacy policies, expressing concern and advocating for industry initiatives to ensure safe experiences for children. While discouraging the indiscriminate use of VR headsets for children, a note of optimism permeated the statement, anticipating positive changes that could redefine safety measures in immersive digital environments.

Embarking on Meta's early journey in constructing the Metaverse, **Anne Hobson,** Policy Manager, Reality Labs at Meta, articulated three fundamental characteristics of Metaverse interactions: ephemerality, embodiment, and immersion. Addressing the paramount safety issue, recent progress made by Meta in safeguarding children was outlined, including the introduction of parental supervision tools and norm development. While acknowledging the strides taken, the statement emphasized the ongoing need for collaborative efforts, recognizing the evolving nature of safety challenges within the ever-expanding Metaverse.

The challenges inherent in creating secure Metaverse environments for children and teenagers were candidly shared. **Dona Fraser,** Senior V.P., Privacy Initiatives at BBB National Programs, delved into the risks, ranging from bullying and harassment[4] to exposure to harmful content. Introducing the Metaverse working group, a collaborative initiative to guide companies in crafting child-friendly experiences, the statement emphasized the pivotal role of digital citizenship education in fostering a safer digital landscape for the younger generation through a collaborative and preventative approach.

Presenting a strategic perspective, the discussion revolved around safeguarding children within the context of technological convergences and the dearth of data on generative AI. **Steven Vosloo**, policy and innovation specialist at UNICEF, highlighted the imperative to consider the "global self" and questioned prevailing parental and teacher

4. Ramirez, E. J., Jennett, S., Tan, J., Campbell, S., & Gupta, R. (2023). XR Embodiment and the Changing Nature of Sexual Harassment. Societies, 13(2), 36. https://www.mdpi.com/2075-4698/13/2/36

supervision assumptions. This statement underscored the need for a comprehensive approach that extends beyond traditional paradigms,[5] urging a reevaluation of existing norms in the face of evolving technologies.

The discussion then pivoted to the OECD's recommendations for child safety in the digital environment, advocating for a common language rooted in digital safety by design. Lisa Robinson, Online Safety Policy Analyst at OECD, emphasized reducing reliance on parental responsibility, calling for a collective commitment to prioritize the well-being of children in the rapidly

evolving digital landscape.[6,7] This statement set the stage for a broader conversation on industry-wide standards and shared responsibilities in ensuring a safe online environment for the younger demographic.

Exploring the concept of extending reality, the discussion delved into the delicate balance between child protection and child rights. **Larry Magid**, form ConnectSafely, stressed the importance of agency, freedom of expression, and privacy for teenagers in the evolving landscape of XR, AR, and VR. While encouraging the development of immersive technologies, the statement emphasized the need to do so responsibly, with due consideration for safety measures and individual control, reflecting a nuanced approach to navigating the complexities of the extended reality space.

The contributors collectively addressed issues such as age-appropriate content, privacy, safety policies, trauma-informed content creation, and the role of organizations and parents in ensuring a secure digital environment for children and teens. The discussions also emphasized the need for ongoing collaboration, education, and awareness.

This roundtable served as a nexus for dialogue, aiming to explore actionable strategies and collaborative initiatives to ensure that our AI-augmented realities prioritize every child's safety, dignity, and rights,

---

5. UNICEF Office of Global Insight & Policy. (2023, May). The Metaverse, Extended Reality and Children. UNICEF. https://www.unicef.org/globalinsight/reports/metaverse-extended-reality-and-children
6. OECD. (2023, December 21). Recommendation of the Council on Children in the Digital Environment. OECD Legal Instruments. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0389
7. OECD. Children in the digital environment. OECD. https://www.oecd.org/digital/children-digital-environment/

fostering an environment where innovation and protection coexist harmoniously.

KEY MESSAGES FROM THE DISCUSSION:

- One prominent focus was the in-depth exploration of risks children may encounter in AI-augmented environments, delving into the intricacies of potential harm. This comprehensive analysis laid the foundation for a more informed and targeted approach to ensuring the safety of the young generation within these immersive digital realms.

- Real-world case studies highlighted the significant impact of AI-augmented environments on child safety, offering critical insights for future strategies in the Metaverse. These practical examples shed light on the complexities of digital spaces, emphasizing the need to learn from lived experiences to enhance protection for children in these evolving environments.

- The dialogue centered on devising practical, actionable strategies to enhance child safety in the metaverse, highlighting the critical need for collaboration between technology firms, regulators, and stakeholders. This focus on implementable measures reflects a commitment to ensuring the metaverse incorporates safety controls by design, actively protecting children's well-being in this evolving digital landscape.

- Collaborative initiatives between technology companies, philanthropic foundations, and advocacy groups are essential in shaping child-friendly AI technologies in the Metaverse. This united effort, focusing on every child's safety, dignity, and rights, reflects a commitment to ensuring the Metaverse evolves as a secure and empowering environment for young users.

The discussions during the roundtable session were enriched by diverse perspectives, with contributions spanning immersive technology stakeholders, technologists, human rights experts, and civil liberties organizations. Through encouraging active dialogue, participants collectively shaped and influenced child safety measures, recognizing the need for a unified approach to empower these influential voices to contribute to the ongoing discourse and development of robust protective measures within AI-augmented environments for children and young audiences.

THEMATIC SESSION 1:
EMERGING PLAYGROUNDS: PREVENTING HARM TO ENABLE OPPORTUNITIES FOR YOUNGSTERS

The thematic discussion on Emerging Playgrounds - Preventing Harm to Enable Opportunities for Youngsters focused on the dual nature of AI-augmented environments as both a fertile ground for learning and creativity and a potential source of risks for children. These risks include psychological impacts and exposure to inappropriate content. Key areas of consideration were the evolving interactions between youngsters and AI agents, highlighting concerns over dependency and manipulation and the role of young creators in potentially influencing or being influenced by harmful content.

The primary goal was to assess risks and harms associated with AI-augmented technologies for children, examine the role of young creators in these digital spaces, and devise strategies to foster safe, healthy, and creative environments. The discussion aimed to balance the need for innovation with the imperative of child safety, addressing the complexity of these emerging playgrounds.

Jessica Stone set the stage for the thematic discussion by highlighting the critical role of AI in therapeutic settings for child safety, highlighting several key points. She emphasized the importance of maintaining confidentiality, safety, and privacy of personal health information, underscoring the need for informed consent when AI is involved. Stone also stressed the necessity for caregivers and clinicians to be fully aware of the potential risks and benefits of using AI with children. This includes engaging in meaningful conversations with children about using AI safely and responsibly. Furthermore, she pointed out the urgent need for further research on the effects of AI on children's development and well-being. Stone advocated for developing comprehensive guidelines to ensure AI's safe and ethical use in therapeutic contexts, ensuring that children's best interests are always at the forefront.

"I emphasize the importance of protecting children's privacy and safety when using AI in therapeutic settings. It is essential to ensure informed consent and a balanced understanding of both risks and benefits. Ongoing research and developing ethical guidelines are crucial to safeguard children's well-being in these emerging digital environments."

- Jessica Stone, Virtual Sandtray, LLC

The session brought about several key recommendations, as follows:

- **Integration of Protective Components**
  - Advocate for child-safe features in AI-augmented environments.
  - Encourage proactive risk management by industry stakeholders.
  - Emphasize the importance of play, socialization, and cognitive impacts in digital realms.
  - Support initiatives promoting responsible online behavior among young creators.

- **Addressing Dependency and Manipulation Risks**
  - Develop technologies that empower young creators while minimizing undue influence.
  - Implement robust age verification methods.
  - Establish clear accountability measures within the tech industry to address misrepresentation and age falsification.

- **Holistic Approach for Safe Digital Playgrounds**
  - Incorporate human oversight and ethical considerations in technology design.
  - Ensure transparent age determination and adherence to established frameworks and guidelines.
  - Engage in ongoing research for continuous evaluation and adaptation.

The session advocated for a comprehensive approach that combines innovation with safety measures. This approach aims to create digital playgrounds that stimulate individual growth and minimize potential risks for young creators in the rapidly evolving field of emerging technologies. The discussion underscored stakeholders' shared responsibility in shaping a digital future that prioritizes the well-being and development of young users in AI-augmented environments. The recommendations aimed to strike a balance between innovation and safety, contributing to creating digital playgrounds that foster individual growth while minimizing potential risks for young creators in the evolving landscape of emerging technologies.

THEMATIC SESSION 2: SAFEGUARDING THE YOUNGER GENERATION IN THE AI WORLD

The thematic discussion on embracing shared responsibility: Uniting parents, guardians, big tech, and policymakers to safeguard children in AI-augmented worlds, focused on the shared responsibility of safeguarding children in AI-augmented environments. It highlighted the increasing engagement of children with AI-driven platforms and the urgent need for collaboration among parents, guardians, big tech, and policymakers to ensure their digital safety.

The primary objective was to develop a unified approach to protect and empower children in AI-augmented worlds, ensuring their experiences are safe, enriching, and respectful of their rights and needs. The goal was to establish a cohesive strategy emphasizing collective responsibility for a secure digital future for younger generations.

Janae McPhaul laid the foundation for the discussion and shared the vital role of engaging young people in policy discussions, especially those policies that impact their experiences in digital spaces. She emphasized the importance of creating safe digital environments for youth. She advocated for a collaborative approach among stakeholders, including policymakers, educators, and technology developers, to ensure online safety and well-being. McPhaul also highlighted the need for awareness and understanding of emerging technologies and their potential effects on young people. She stressed the importance of empowering youth with the knowledge and skills necessary to navigate these spaces safely and responsibly, underscoring the collective responsibility in shaping a positive and secure digital future for the younger generation.

"As an advocate for youth safety in digital spaces, I stress the importance of involving young people in policy discussions that affect them. We must collaborate across sectors to create and maintain safe digital environments for our youth, equipping them with the knowledge and skills to navigate these spaces safely. It is crucial to understand the impact of emerging technologies and foster a culture of informed and responsible use among young people."

- Janaé N. McPhaul, XRSI | Cyber XR Coalition

The session brought about several key recommendations, as follows:

- **Collaborative Efforts**
  - Emphasize shared responsibility among parents, guardians, educators, and technology creators.
  - Foster partnerships and joint initiatives to enhance child safety in digital spaces.

- **Importance of Digital Literacy**
  - Highlight the need for digital literacy programs for children and educators.
  - Develop and implement comprehensive digital literacy curricula to help users navigate online environments responsibly.

- **Platform Accountability**
  - Discuss the critical role of online platforms in content design, filtering, and moderation.
  - Implement effective content moderation strategies to ensure a safe online environment for children.

- **Transparency in AI Interactions**
  - Stress the importance of educating children about AI interactions.
  - Foster transparent AI interactions and education to enhance user control and awareness.

- **Government Support**
  - Recognize the necessity for government funding in educational resource development.
  - Secure government support for creating curriculum materials focusing on online behavior and digital safety.

- **Holistic Education**
  - Address the need for comprehensive education covering digital safety and mental health.
  - Promote holistic educational approaches encompassing digital safety and mental health in virtual worlds.

- **Proactive Approach to Digital Safety**
  - Anticipate and address potential risks in AI environments for children.
  - Develop strategies and resources for digital safety awareness and prevention.

The session underscored the complexity of creating a safe and responsible AI environment for children. It called for a united effort among various stakeholders to implement proactive measures, prioritizing the well-being of children in the rapidly evolving AI landscape. The collective commitment to this mission is crucial for fostering a digitally secure future for the younger generation.

STRATEGIC INTELLIGENCE GATHERING SESSION VIA SWARM AI

In the thematic session focused on safeguarding children in AI-augmented environments, the innovative Swarm AI® technology developed by Unanimous AI played a crucial role. This technology uniquely blends real-time human insights with advanced AI algorithms inspired by the natural phenomenon of swarm intelligence observed in birds, bees, and fish. In these natural contexts, collective groups demonstrate enhanced decision-making capabilities, surpassing the abilities of individual members. Similarly, Swarm AI enables human groups to artificially amplify their collective intelligence by forming real-time swarms, a capability not inherently present in humans.

This technology was instrumental in guiding our discussions and decision-making. It allowed us to approach complex questions with collective wisdom, providing more nuanced and comprehensive insights. The key questions we addressed included determining the most effective ways to create a unified approach for safeguarding children, preparing children in educational systems, keeping Big Tech accountable for child safety, defining the role of parents in AI spaces, ensuring AI spaces are inclusive and unbiased for youth, and guiding young AI creators against harmful content.

Question for deliberation: **Best method to teach kids about AI manipulation**

Response: **Mandatory school AI literacy programs**

After thoughtful deliberation and using Swarm AI® technology to seek consensus, the unequivocal recommendation is to introduce mandatory AI literacy education in schools (Fig -1). This approach integrates digital and data education into various subjects, fostering a holistic understanding of AI's impact from an early age, akin to math and reading. It ensures every child, regardless of socio-economic background, receives foundational knowledge of AI manipulation. While diverse methods like influencers and online courses are acknowledged,

mandatory AI literacy in schools establishes a standardized and accessible platform. Collaboration among federal and state governments, educational institutions, and influencers enriches and reinforces this vital educational mandate, preparing the next generation for the digital age.
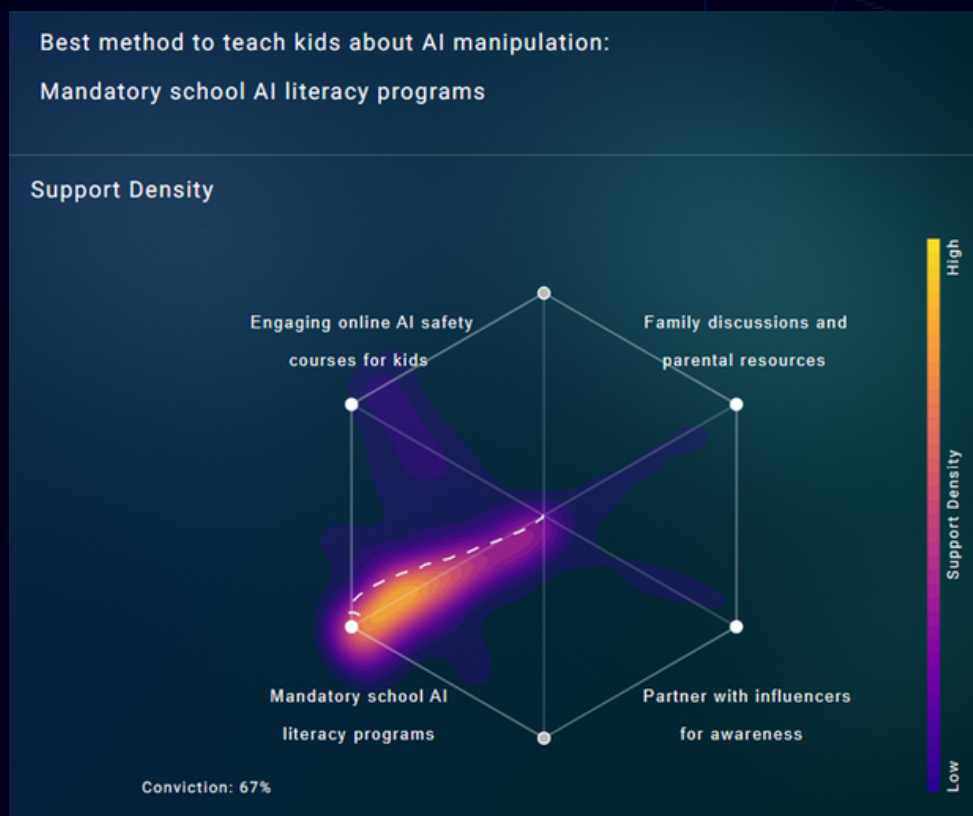

Best method to teach kids about AI manipulation:
Mandatory school AI literacy programs

FIG -1

Question for deliberation: What is most effective to create a unified approach to safeguarding children
Response: Shared digital safety framework

In response to safeguarding children in AI-augmented environments, the group presents a resounding recommendation for adopting a robust shared digital safety framework(Fig -2). This collective recommendation advocates for a proactive and collaborative approach among all stakeholders, including the participation and voice of children and youth, emphasizing the critical need to implement a unified framework that prioritizes digital safety. This solution encourages a comprehensive strategy encompassing digital literacy and AI integration into educational curricula while recognizing the pivotal role of responsible

stakeholders in constructing and regulating digital environments. Emphasizing the urgency of this recommendation, it underscores the importance of collective efforts in ensuring a secure and protective environment for children navigating AI-augmented landscapes.
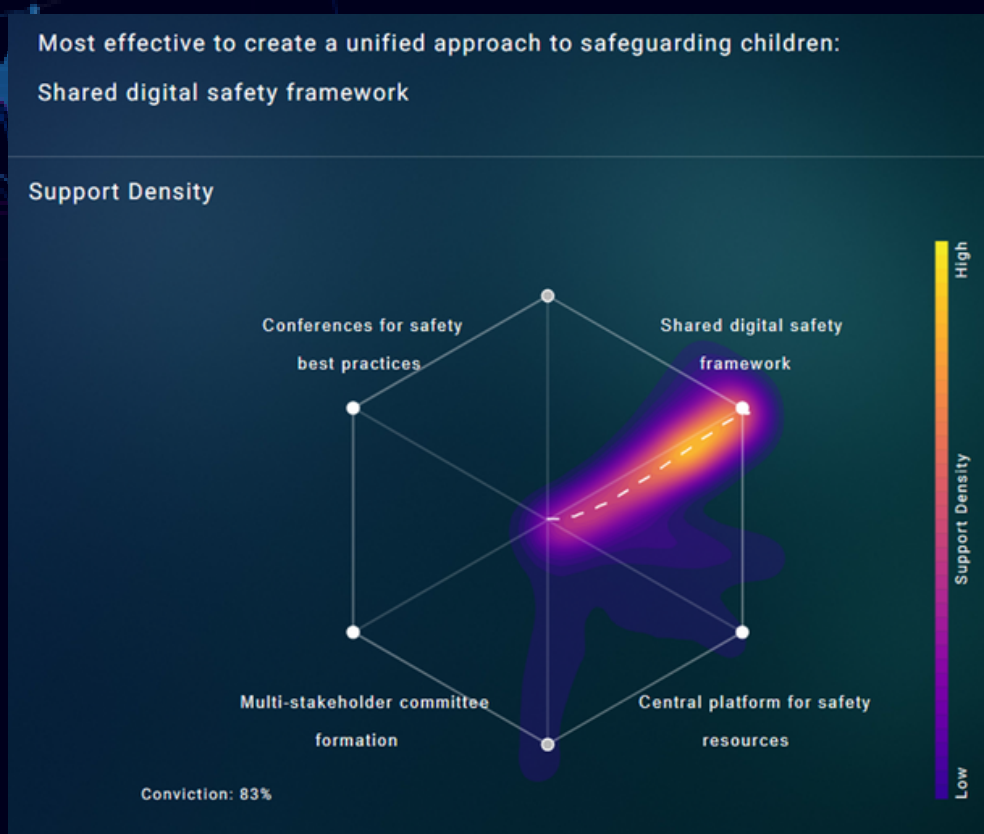

FIG -2

By leveraging Swarm AI, our session benefited from a synergy of diverse perspectives, enabling us to reach conclusions deeply rooted in a shared understanding. This approach was critical in formulating strategies and recommendations that are innovative, effective, and inclusive, ensuring that our solutions are well-suited to address the multifaceted challenges of child safety in AI spaces.

SHAPING AI-DRIVEN IMMERSIVE REALITIES FOR THE YOUNG GENERATION

Children's privacy, health, and well-being emerged as the primary concern, with participants highlighting the need for responsible data practices, including age verification methods[8] and ensuring the security and protection of sensitive information. Based on the multistakeholder domain experts and researchers, some of the key risks to consider for safeguarding children in immersive spaces are as follows:

1. Extensive Data Collection and Use in Immersive Spaces:

- One significant risk in the immersive technologies and use of AI is that organizations can now extensively collect and use far more precise and highly sensitive data of users, including that of children. In order to realize its full potential, organizations need to have greater contextual understanding when deploying immersive technologies, as compared to digital non-immersive technologies. This is only possible through collecting more situational, behavioral, device, and usage data of users in real-time.

- Such data and inferences from them are combined with data from other sources, internal or external to the organization. They are used to train various AI models proliferating throughout the organization. As a result, there is immense opportunity to use this trove of significant data for personalization, profiling, recommendation, and so on.

---

8. Yoti. (2023, December 15). Facial Age Estimation white paper. Yoti. https://www.yoti.com/blog/yoti-age-estimation-white-paper/

CHILD SAFETY AND CHILDREN'S RIGHTS

2. Trust and Manipulation in the Immersive Space:

- The risk of blurring between advertisement and children's content is reported to have intensified in the metaverse, AR, and VR space.[9] Interweaving advertisements into the children's content makes it easier to manipulate children to make purchases, to make them view advertisements, and to engage them more effectively.

- This is mainly because children, through their interactions with avatars and influencers, may have a false sense of personal connection and trust within the immersive space.

3. Bio-psycho-social Risk in Immersive Environment:

- For children in particular, prolonged exposure to AI-augmented environments is concerning since it can harm their health, development, and overall well-being. It can impact their eyesight and sleep patterns and even result in addiction-like behaviors. Moreover, the instances of violation, exposure to inappropriate content, or inappropriate social interactions in their online experiences can all have a profound psychological impact on children.

- The risk of cyberbullying and even social isolation resulting from excessive engagement with AI-augmented platforms all highlight the negative impacts on the cognitive and emotional well-being of children. The need for new guidelines and content filters, as well as the need for robust AI moderation to address these issues, are highly critical and cannot be ascertained further.

> *"In tech companies we currently drive that 'Privacy is everyone's responsibility', not just for Privacy team, we need to ensure the same for children safety in digital world."*
>
> **- Vandana Ujjual, Data Privacy Expert**

9. BBB National Programs. (2023, October 3). Guardrails for Child-Directed Advertising & Privacy in the Metaverse - 116th Congress (2021-2022): Metaverse Guardrails. BBB National Programs. https://resources.bbbprograms.org/metaverse-guardrails-for-children

4. Mis-Aligned Privacy and Product Strategy:

○ It is seen that the vast majority of organizations using or deploying these systems need a centralized system to capture and document the extensive data they process or the sensitive biometric data they collect, including that of children. It is reported that only 39% of organizations systematically track the AI systems used in their organizations and conduct assessments of potential risks, operational failures, and bias.[10]

○ To ensure accountability for privacy, any data collected and used, including their downstream uses, need to be tracked. Further, ensuring accountability for user data not only helps organizations strategize for responsible data use but also helps organizations monitor for privacy compliance.

5. Ensuring Transparency in Immersive Context:

○ Organizations face many challenges in providing transparency in the context of AR and VR. Due to the unique nature of immersive technologies and the use of newer data, such as eye-tracking, the current practices, like layered privacy policies, just-in-time notices, and in-context notifications, may need to be revised to provide meaningful notice to users. Meta has highlighted the inherent tensions with communicating about privacy,[11] since the collection and use of data in these contexts may differ from traditional digital technologies.

○ Furthermore, when users are children, it is essential to consider their age and cognitive ability, while communicating to them and providing transparency.

10. Eversheds Sutherland. (2022, March 10). Eversheds Sutherland report reveals global focus on ethics in technology, responsible data use and new regulations set to spark investment in CDR. Eversheds Sutherland. https://www.eversheds-sutherland.com/en/united-states/news/eversheds-sutherland-report-reveals-global-focus-on-ethics-in-technology-responsible-data-use-and-ne

11. Facebook. (2020, July). Charting a Way Forward: Communicating About Privacy. Facebook – About Facebook. https://about.fb.com/wp-content/uploads/2020/07/Privacy-Transparency-White-Paper.pdf

6. Code of Conduct in Virtual Experiences and Governance:

- The code of conduct in the virtual environment and how it is governed, demands greater attention because of the bio-physio-social impacts on children.

- It is critical to know whether it is more effective if organizations enforce them by means of sanctions (for negative behavior or for violating organization's Community policy on their platform or app), as opposed to allowing the user community to enforce them as a collective, through voting, etc.

- Accordingly, a community-driven approach means that "VR etiquette can be nuanced, contextual, and evolve naturally over time instead of being written in stone by a corporation which has its own interests as a top priority".[12]

- With regards to legal outcomes, it is not clear as to what level of severity of mis-conduct in an immersive environment is needed for any legal action to commence. Further, it is not clear about how the severity of mis-conduct is determined, when law enforcement typically takes over, and which of the compliance applies.

- Setting industry standards for code of conduct and governance, can help find answers to these questions.

---

12. Road to VR. (2023, December 15). Facebook's Expanded VR Policies Disallow "invading personal space" and "sexual gestures". Road to VR. https://www.roadtovr.com/facebook-expanded-vr-policies-oculus-quest-2-privacy-policy-terms-of-service/

7. Digital Literacy, Children's Rights and Online Protection:

- One main concern was the tension between child protection and child rights and whether our existing regulations for protecting children are adequate. In the United States, for instance, we do not have a federal privacy bill, and the Children's Online Privacy Protection Act (COPPA),[13] only applies to children under thirteen.

- On one hand, it is important that children are entitled to freedom of expression and access. On the other hand, as caregivers and guardians, not all parents or educators are equipped to guide the children and teens who engage in immersive space. In general, parents and caregivers are seldom able to fully comprehend the organization's privacy policies that apply to their XR platform or app.

- The consensus during the intelligence gathering session underscored the importance of incorporating digital literacy and AI education into the formal curricula. It was noted that children (and parents) may lack awareness of existing protections, and may not fully understand how to navigate and control the digital tools available to them.

"The learning experience is so much more immersive than anything I would have ever experienced. However, we know that these spaces also create potential risks and harms. These risks can include everything from bullying, body image issues, sexual exploitation, harassment, allowing children to reveal too much personal information, making unintentional purchases, exposure to self-harm, violence, and gambling."

- Dona Fraser, BBB Nationals Program

13. Federal Trade Commission. (2023, December 21). Children's Online Privacy Protection Rule ("COPPA"). Federal Trade Commission https://www.ftc.gov/legal-library/browse/rules/childrens-online-privacy-protection-rule-coppa

THE INTERSECTION OF YOUTH, TECHNOLOGY, AND SAFETY: OPPORTUNITIES IN THE METAVERSE

In an era defined by the rapid evolution of technology, where the boundaries between the physical and digital worlds continue to blur, the safety and well-being of our youth have never been more paramount. As we stand at the crossroads of the metaverse, a digital realm brimming with immersive experiences and limitless possibilities, it is imperative that we navigate this landscape with a vigilant eye toward safeguarding our children. Several opportunities also highlighted during the roundtable discussion offer a multifaceted approach to metaverse safety. These opportunities, grounded in empirical insights and expert analysis, unveil a holistic strategy that integrates innovation with responsibility, underscoring the pivotal role of technology in creating a digital space that prioritizes the welfare of its youngest users.

- **Balancing Immersive Technology and Safety:** One key opportunity also identified during the roundtable discussions was promoting immersive technologies while ensuring safety measures are in place. Participants recognized the transformative potential of immersive digital experiences but emphasized the need to protect users, especially children, from potential risks.

- **Multi-Stakeholder Collaboration and Regulatory Sandboxes:** The discussions highlighted the potential of multi-stakeholder collaborations and regulatory sandboxes to design effective privacy notices and controls in immersive technologies, aligning with a human-centric and iterative design approach.

- **Empowering Data Privacy Control:** Empowering individuals, particularly children, with control over their data privacy emerged as a significant opportunity. The focus was on providing users with tools and knowledge to manage their digital footprint, promoting autonomy and informed decision-making.

- **Integration of Digital Literacy in Education:** Digital literacy was recognized as a potent tool for empowerment, with a call to integrate it into educational curricula. The goal was to equip younger generations with the skills needed to navigate the digital world responsibly.

- **Age-Appropriate Content and Verification:** The importance of defining age-appropriate content, estimating its impact on different age groups, and implementing verification mechanisms was emphasized. This comprehensive approach aimed to create a safer online environment tailored to different developmental needs.

- **Safety-by-Design Approach:** Participants stressed the principle of safety-by-design,[14] advocating for safety measures to be integrated into the core of digital experiences. This proactive approach aimed to mitigate risks from the outset.

- **Scenario Planning for Governance:** Scenario planning was seen as a proactive governance opportunity, enabling stakeholders to address potential challenges and adapt to evolving circumstances within the metaverse.

- **Campfire Chats and Youth Engagement:** Creating spaces for open dialogue and involving youth in safety discussions were identified as opportunities to gather diverse perspectives and co-create solutions that resonate with the younger generation's experiences and needs.

---

14. eSafety Commissioner. Safety by Design. eSafety Commissioner. https://www.esafety.gov.au/industry/safety-by-design

- **Inclusive Design and Influencer Collaboration:** Inclusive design and collaboration with influencers were highlighted as key opportunities to promote diversity and safety in the metaverse. Designing accessible digital spaces and leveraging influencers to convey safety messages to diverse audiences were seen as strategic approaches.

In summary, these opportunities collectively form a forward-thinking approach to metaverse safety, emphasizing the integration of innovation and responsibility to prioritize young generations' well-being in the digital landscape.

"Anybody who has talked about extending COPPA beyond 18 needs to realize that teenagers need and deserve and are entitled to agency, to free speech, to participation. In fact, the UN commands that the rights of the child specifically lay out that children are entitled to freedom of expression, freedom of consumption of all media in all forms real and imagined, present, and future…We need to think about the rights of young people."

- Larry Magid, Connect Safely

EMPOWERING THE YOUNG GENERATION: STRATEGIES AND RECOMMENDATIONS

Several measures and recommendations have been identified to create a safer and more positive online environment for children. These strategies are grounded in factual considerations and aim to address the challenges posed by AI manipulation and online safety for children. Here are the key points:

- **Incorporate Digital Literacy into Curriculum:** To ensure children are well-prepared to navigate online spaces safely, there is a call to add digital literacy and AI education to formal curricula. This proactive approach emphasizes the importance of equipping children with essential skills.

- **Utilize Influencers for Education:** Collaboration with influencers effectively disseminates educational content about AI manipulation and online safety. Leveraging their influence can positively impact children's understanding of the metaverse.

- **Implement Robust AI Safety Features:** Prioritizing AI safety features, including age verification, content filters, and moderation tools, is crucial to detect and prevent harmful content on technology platforms.

- **Promote Diversity in Design Teams:** Advocating for diversity within AI design and testing teams is essential for ensuring inclusive and unbiased technology development. It will help minimize potential biases in AI systems.

- **Government Legislation and Regulation:** Encouraging government involvement in legislating, regulating, and governing online spaces is necessary to protect children. Policymakers are urged to engage with technology to make informed decisions actively.

- **Strengthen Privacy and Security Measures:** Stringent privacy and security measures, including age verification, are proposed to safeguard sensitive data and protect children from potential risks associated with the misuse of personal information.

- **Facilitate Parental Engagement:** Developing resources and campaigns to encourage open dialogues between parents and children about online experiences is vital. Educating parents about their role in overseeing their children's online activities is emphasized.

- **Include Children in Decision-Making:** It is recommended to create an inclusive environment where children are included in discussions about their online experiences. Their perspectives should be considered in developing online spaces to ensure relevance and safety.

- **Establish Collaborative Frameworks:** Creating shared digital safety frameworks involving collaboration among various stakeholders, including tech companies, policymakers, educators, parents, and children, is highlighted. These frameworks prioritize the safety and well-being of children in digital environments.

- **Campfire Chats and Youth Engagement:** Informal opportunities, such as campfire chats, are suggested for direct engagement with youth. This approach allows a deeper understanding of their experiences, concerns, and preferences, shaping interventions and policies accordingly.

In summary, these evidence-based measures and recommendations aim to create a safer online environment for children by focusing on education, inclusivity, collaboration, and responsible technology use.

GUARDIANS AND SHIELD FRAMEWORK BY CHILD SAFETY INITIATIVE

XRSI's Child Safety Initiative team has already embarked on the research and development of a shared digital safety framework: the Guardians and Shield Framework.[15] As a call to action emerging from the Swarm AI Strategics intelligence gathering session, we recommend that those wanting to take action immediately join this global and multidisciplinary team. The Guardians and Shield Framework will address children's significant risks and harms when interacting and using tools, platforms, or services created through immersive and emerging technologies. The project focuses on creating a comprehensive and actionable tool that guides different stakeholders in creating, developing, and delivering these experiences for children. The project uses a mixed-methods research approach involving stakeholders from different sectors, including children, parents, educators, policymakers, public safety personnel, medical staff, and industry experts.

The framework will guide emotional, behavioral, physical, social, and cognitive safety. Additionally, the framework will help raise awareness about the opportunities, risks, and harms stemming from the metaverse and promote children's rights to safety, well-being, and participation in digital spaces. To learn more and join this team, go to XRSI Child Safety Initiative Page.[16]

15. XRSI-CSI, Guardians and Shield Framework, https://silicon-drum-7c2.notion.site/Guardians-and-Shield-Framework-1abc950f47c64b518368ece24426591e
16. X Reality Safety Intelligence (XRSI), The Child Safety Initiative. XRSI. Retrieved from https://xrsi.org/programs/child-safety

CHILD SAFETY AND CHILDREN'S RIGHTS

CALL TO ACTION : ADOPT MSW

Since its inception in 2020, Metaverse Safety Week (MSW) has illuminated the intricate fusion of Immersive and emerging technologies. MSW 2023, in particular, highlighted the symbiotic relationship between AI and these environments, unveiling their immense potential alongside inherent risks. As the Founder and CEO of XRSI, I urge stakeholders to <u>adopt MSW as an annual awareness campaign</u>, galvanizing action to shape a secure future for the Metaverse.

- **The Imperative for 2023 and Beyond**
  MSW 2023 emphasized the urgency to safeguard the Metaverse, calling for action beyond risk acknowledgments. It's about actively shaping a secure future within this evolving landscape.

- **Empowering through Education**
  Engage with us and partake in a wide range of activities aimed at raising awareness, educating stakeholders, and promoting a safer and healthier Metaverse for global citizens. Prioritize educating teams about this evolving landscape, issuing transparency reports, and making commitments to drive collective action.

- **Shared Responsibility, Collective Action**
  Join us in adopting MSW as a yearly initiative to fortify alliances, develop best practices, and craft protective policies for these immersive landscapes. It's a shared responsibility across individuals, organizations, policymakers, and creators to ensure a secure Metaverse.

This isn't just a campaign; it's a commitment—a shared responsibility to safeguard the future of the Metaverse. Whether you're a government, a big technology organization, a creator, an educator, or a policymaker, your role is pivotal in promoting a culture of safety and trust within these emerging realities.

> *"Let's unite in this endeavor to fortify our shared vision of a secure and transformative Metaverse. **I implore you to join hands and hearts in adopting the Metaverse Safety Week (MSW) campaign by signing the MSW charter** and standing with us as we shape a future that's safe, ethical, and full of boundless possibilities."*
>
> **- Kavya Pearlman, Founder & CEO - XRSI**

CO-ORGANIZER DETAILS

ABOUT CHILD SAFETY INITIATIVE

XRSI Child Safety Initiative is collective of leaders, subject matter experts, and visionaries across industries and continents working to help build safe and inclusive experiences for children and adolescents using Extended Reality (XR) and the incoming Metaverse paradigm. Children start engaging with the internet and technology at a very young age without necessarily being aware of the risks associated with it. In the case of immersive experiences, young individuals are exposed to even additional challenges that should be addressed through a multidisciplinary approach. We are committed to fostering the positive elements associated with the use of Extended Reality (XR) and protecting kids from the risks associated with it.

We believe children can explore immersive worlds if they have consistent safety, security, and privacy measures to safeguard their experiences. By supporting the XR ecosystem with standards, awareness programs, research, and innovation specially designed to safeguard children, we can prevent risks and harm in such experiences. With that, we fulfill our mission of creating safe environments in immersive worlds. We also provide guidance to empower families, teachers, and policymakers around emerging technologies to be more active and decisive in their roles.

CO-ORGANIZER DETAILS

ABOUT UNICEF

UNICEF works in the world's toughest places to reach the most disadvantaged children and adolescents – and to protect the rights of every child, everywhere. Across more than 190 countries and territories, we do whatever it takes to help children survive, thrive and fulfill their potential, from early childhood through adolescence. The world's largest provider of vaccines, we support child health and nutrition, safe water and sanitation, quality education and skill building, HIV prevention and treatment for mothers and babies, and the protection of children and adolescents from violence and exploitation.

Before, during and after humanitarian emergencies, UNICEF is on the ground, bringing lifesaving help and hope to children and families. Non-political and impartial, we are never neutral when it comes to defending children's rights and safeguarding their lives and futures.

APPENDIX 1: OUTPUT FROM SWARM AI SESSIONS

Question & Answer	Conviction
Most effective to create a unified approach to safeguarding children **Answer: Shared digital safety framework**	**83%**
Most effective way educational systems can prepare children **Answer: Add digital literacy and AI to the curriculum**	**87%**
Highest priority child-safety step policymakers should take: **Answer: BRAIN FREEZE!**	**0%**
Most effective in keeping Big Tech accountable for child safety in AI **Answer: Regular audits and transparency reports**	**58%**
Most appropriate role for parents in kids' safety in AI spaces **Answer: Engage with kids in online activities**	**69%**
Best approach to make AI spaces inclusive and unbiased for youth **Answer: Diverse design and testing teams**	**83%**
Best strategy to guide young AI creators against harmful content: **Answer: Mentorship in ethical content creation**	**39%**
Best approach to reduce youth AI dependency **Answer: BRAIN FREEZE!**	**0%**
Best method to teach kids about AI manipulation **Answer: Mandatory school AI literacy programs**	**67%**
Most effective to identify & mitigate psychological risks to children **Answer: Child-specific guidelines and filters**	**40%**

LEAD AUTHORS

Vandana Ujjual
Data Privacy Expert

Desmond Israel
Center for AI and
Digital Policy | XRSI

Jessica Ray
Program Manager,
Child Safety Initiative,
XRSI

PROGRAMMING COMMITTEE MEMBERS: CHILD SAFETY AND CHILDREN'S RIGHTS

Valentino Megale
XRSI | Softcare Studios

Hlekiwe Kachali
UNICEF

Kavya Pearlman
X Reality Safety Intelligence
(XRSI)

Jessica Ray
XRSI Child Safety Initiative

Desmond Israel
Center for AI and Digital Policy
| XRSI

Vandana Ujjual
Data Privacy Expert

Janaé N. McPhaul
XRSI | CyberXR Coalition

**Bhanujeet
Choudhary**
X Reality Safety Intelligence
(XRSI)

Aashka Patel
Wellpack Pharmacy | XRSI

Francesco Pagano
JAKALA | XRSI

ACKNOWLEDGEMENTS

PARTICIPATING ENTITIES AND PARTNERS

PARTICIPATING ENTITIES AND PARTNERS

MSW 2023 Key Supporters and Partners

www.metaversesafetyweek.org

CHILD SAFETY AND CHILDREN'S RIGHTS

ORGANIZED BY

XRSI
Human Intelligence In The Loop

www.xrsi.org