METAVERSE SAFETY WEEK

4th annual edition | DEC 10-15, 2023

# CYBERSECURITY AND DATA PROTECTION

## NAVIGATING THE CYBER FRONTIER IN THE AI-POWERED METAVERSE

DECEMBER 13, 2023

THE ROUNDTABLE REPORT

CO-ORGANIZED BY:

XRSI™ Human Intelligence In The Loop

GPA Global Privacy Assembly

SUPPORTED BY:

Meta

www.xrsi.org

www.MetaverseSafetyWeek.org

# TABLE OF CONTENTS

# EXECUTIVE SUMMARY

In the dynamic landscape of cybersecurity, AI, and data protection, a profound challenge emerges the need for a deeper understanding and focused navigation of the cyber frontier in the AI-powered Metaverse. Current dialogues on AI safety, security, and data protection often need to be anchored to the traditional Internet framework, thereby missing the intricate nuances of the immersive Internet. This immersive internet represents a confluence of technologies characterized by presence, persistence, immersion, and interoperability, presenting a new paradigm in digital interaction and data stewardship. However, only a select few groups are actively addressing the unique challenges of this emerging realm, especially in cybersecurity and data protection. These challenges are particularly pronounced in AI-powered communication infrastructure, which introduces unprecedented freedoms regarding identity, currency, community choice, cross-jurisdictional networking, and complex data protection concerns.

Amidst this backdrop, most discussions about AI safety, security, and data protection still focus on the use, development, and impact of general AI and large language models (LLMs) within a more conventional Internet-based societal information infrastructure. This approach overlooks the rapidly evolving domain of the Metaverse, highlighting a critical need to comprehend and address the complexities of cybersecurity and data protection in this ever-changing landscape.

**Emerging Regulatory Landscape and International Efforts:**
Discussions on global policy commenced by examining critical international efforts to regulate AI. The European Union has made significant strides with its proposed Artificial Intelligence Act, aiming to set legal precedents for the responsible use of AI, with a keen focus on high-risk applications and transparency. In the United States, legislative developments such as the Algorithmic Accountability Act of 2023 and executive orders are shaping the landscape to foster innovation and

collaboration in AI. In the East, nations like Japan and South Korea are seamlessly weaving AI governance into their digital strategies.

**Emerging Threats and the Lack of Preparedness:** As we venture into this new frontier, malicious actors will inevitably exploit these novel environments to target humans. The threats in the Metaverse are amplified versions of existing cybersecurity concerns and include new challenges, particularly given the vast amount of personal and biometrically inferred data flowing through Metaverse data pipelines. There is a critical gap in understanding and preparedness for these threats regarding the evolving spectrum of challenges in the Metaverse, increased by AI-driven vulnerabilities.

**Identifying the Gap and Addressing the Challenges:** The roundtable addressed these challenges by understanding the current landscape of cybersecurity threats and data protection challenges. It highlighted how these threats will be amplified in the AI-powered Metaverse and identified novel challenges emerging from this convergence. Discussions covered the evolving spectrum of cyber threats, the need for dynamic data protection frameworks, and collaborative strategies for securing emerging technologies. The roundtable underscored the importance of developing collaborative resilience strategies for securing emerging technologies, focusing on safeguarding the integrity and privacy of Metaverse citizens.

**Strategic Intelligence Gathering by XRSI and GPA:** Hosted by X Reality Safety Intelligence (XRSI) and the Global Privacy Assembly (GPA), the roundtable served as a strategic intelligence gathering event involving global multi-stakeholder participants. The overall focus was not just on current concerns but forward-looking measures, specifically aimed to address the gap in overall understanding and thought narrative regarding the cybersecurity and data protection challenges and threats in the AI-powered Metaverse. It focused on creating strategies to build resilience and developing fact-based frameworks and regulations. The discussions included what actionable steps to create resilience and develop fact-based frameworks and laws, recognizing that the risks associated with the AI-powered Metaverse are real and could lead to

human harm and rights violations if unaddressed. The event marked a significant step in developing multi-dimensional strategies for a safer, more equitable, and innovative digital future.

This report acknowledges that while the Metaverse may be an interconnected network of virtual worlds powered by AI and human interaction, its risks are genuine. These risks extend beyond digital harm, potentially leading to human harm, loss of lives, and violations of human rights if not proactively addressed.

Protection of a community's data can only succeed collaboratively, with society investing in the necessary security infrastructure, with organizations accepting responsibility for the data they collect, with each individual taking steps to protect their data, and with oversight bodies working to provide expertise on best practices. This need for collaborative solutions implies a new guiding philosophy to our global society, a Data Protection Social Contract.

The roundtable represents a pivotal step in bridging the gap between current understanding and the urgent need for informed, collaborative, and strategic approaches to cybersecurity and data protection in the AI-powered Metaverse.

**Kavya Pearlman**
Founder & CEO, XRSI
Chair - Cybersecurity and Data Protection, MSW2023

**Nandita Narla**
Advisor, XRSI
Co-Chair - Cybersecurity and Data Protection, MSW2023

# INTRODUCTION

On **December 13, 2023**, a significant event unfolded: the Global Roundtable Discussion on Cybersecurity and Data Protection. This pivotal gathering, co-organized by the **Global Privacy Assembly (GPA)** and **X Reality Safety Intelligence (XRSI)**, brought together global experts, industry leaders, academics, policymakers, and multidisciplinary stakeholders. This roundtable focused on exploring the balance between technological advancement and data protection, with the primary theme of "Navigating the Cyber Frontier in the AI-Powered Metaverse." The goal was to address rising cybersecurity threats that influence global stability and individual safety, focusing on threat intelligence, data protection mechanisms, and strategic responses.


The rise of AI-based cyberattacks blurred the lines between digital and physical threats, presenting risks that rippled across societies, impacting human lives, economic stability, and national peace. In this ever-shifting digital landscape, threat intelligence emerged as a linchpin, bridging the divide between the virtual and the real and offering solutions to protect both.

This roundtable report encapsulates the essence of a convergence of diverse insights; all focused on the multifaceted challenges and possibilities of cybersecurity in the AI-powered Metaverse. The multistakeholder groups' collective aspiration was to define strategies and blueprints that empower every individual to thrive in the digital realm, with their rights, safety, and data integrity paramount.

> *"Traditionally, the cyber attack surface had remained limited to nodes, networks , and serevers. But with the Metaverse, the attack surface has now expanded to human brains and our societies."*
>
> **- Kavya Pearlman, Founder & CEO - XRSI**

## KEY OBJECTIVES INCLUDED

- **Identify and understand Emerging Cyber Threats:** The foremost objective was to delve deep into the evolving spectrum of cyber threats in the AI-powered Metaverse. By comprehending these threats, the group aimed to develop proactive strategies to counter them effectively.
- **Dynamic Data Protection Frameworks:** The group sought to formulate approaches to develop dynamic, responsive data protection and privacy frameworks tailored specifically for immersive digital environments. These frameworks empowered individuals to control their personal data in the Metaverse.
- **Collaborative Resilience Strategies:** Recognizing the interconnectedness of the digital ecosystem, the group endeavored to develop collaborative resilience strategies for securing emerging technologies. This approach entailed safeguarding the integrity and privacy of Metaverse citizens, enabling them to thrive securely in this evolving digital landscape.

Participants from various fields brought their expertise to the table, engaging in a multi-faceted dialogue about safeguarding data and privacy. The presence of cybersecurity experts, data protection officers, legal professionals, policymakers, and technology leaders underscored the roundtable's commitment to a holistic understanding of cybersecurity and data protection in digital spaces. The discussions aimed to go beyond theoretical discourse, focusing on tangible actions and strategies that could be immediately applied to protect sensitive data and ensure privacy in digital environments.

This global roundtable discussion represented a significant step forward in the collective effort to strengthen cybersecurity and data protection 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 cybersecurity and data protection remain at the forefront of technological advancement in our increasingly digital world.

## 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 AI® technology, developed by Unanimous AI, 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.

# THE ROUNDTABLE OVERVIEW:

On December 13, 2023, GPA and XRSI organized a strategic intelligence roundtable focused on cybersecurity and data protection as part of the 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 featured three hours of dynamic and engaging discussions with experts in various fields who shared their experiences and insights on "Navigating the Cyber Frontier in the AI-Powered Metaverse" with the audience.

The hosts, **Kavya Pearlman** and **Nandita Rao Narla** were joined by the following individuals for the discussion. **Jordan Wiseman** also supported the discussion, with support from several XRSI Team Members and Advisors in the background. Co-Host GPA representative Alexander McD White, also the Privacy Commissioner for Bermuda, started the conversation with a quote that directed the roundtable and focused on the human aspect of Cybersecurity and the Metaverse. "This need for different groups to work together implies a new guiding philosophy to our global society: a data protection social contract."

The conversation followed with distinct opening remarks from **Madan Oberoi**, Executive Director for Technology and Innovation at Interpol, emphasizing the complex landscape at the intersection of technology, innovation, and security. He acknowledged the rapid evolution of technology and its significance for global security. As the Metaverse blurs the lines between the digital and physical worlds, new threats emerge, potentially harming societies, businesses, and individuals. Madan pointed out that the challenge does not solely lie in technology but in the disparities in its adoption rates, with criminals often being early adopters, presenting unique challenges.

Madan touched upon important questions regarding avatars, data management, interoperability, and the responsibility for safety within the Metaverse, emphasizing the need to define various stakeholders' intervention thresholds. Interpol's commitment to collaboration with stakeholders, including the Interpol Metaverse Expert Group,[1] was highlighted, with plans to publish a white paper on the Metaverse in early 2024. Madan concluded his remarks by emphasizing the importance of collaboration and innovation to ensure the safety and security of the virtual world.

The following distinguished panelists each brought forth their knowledge and how we should bring the navigation of the cyber frontier in the AI-powered Metaverse while protecting the humans involved.


The aspect of focus on how we define and guide the technology for the protection of the individuals is a common theme among the various speakers. **Josefina Román Vergara** provided examples of how the work of INAI is driving the alignment of regulations for using personal data. At the same time, **Dr. Mohammed Khamis** reminded us that even though there may be harm in the use of the personal/digital data of an individual, we "need to find the balance between regulation and making sure this data is used properly, but also, if we restrict it too much, we render these technologies useless."

1. Groenewald, A. (2023, February 15). Interpol Wants to Start Policing the Metaverse. CyberGhost Privacy Hub. https://www.cyberghostvpn.com/en_US/privacyhub/interpol-metaverse

**Alvin Wang Graylin** discussed his work both at HTC and in various security roles and his new book with Louis Rosenberg, "Our Next Reality." The key points Alvin brought forth were how it is "also important to maintain the privacy and identity of individuals in addition to the security of networks and the security of nations."

**Jameson Spivack**, FPF, discussed how organizations further incorporate immersive technology into products, blurring the boundaries between digital and physical worlds. The fact is that we as a society want ease of use and faster access and are willing to share our data, which creates a risk level we have not seen before. Thus, the Future of Privacy Forum has developed a Risk Assessment[2] to help organizations understand and mitigate the new and expanded risks in this globalized world.

**Peter Price**, Crimestoppers, reminded us that we, as people, are working to improve the use of immersing technology and data protection, "very importantly, we need to think about the positive side of what this (technology) can offer. As a security group, we often look at the negative side of things, and our job is to look at risk. However, when we look at risk, the flip side of risk is looking at community resilience." He continues to say that "community resilience is 100% our top priority." It is an excellent thought to focus on protection while helping our communities build resilience and awareness of the risks so they know what to watch for.

**Anahiby Becerril** brought forth another aspect: national peace and security, awareness, and training, which are some priority areas we must focus on. Her comments reinforce many other comments we heard that day on where to start, drive alignment, and bring improvements while remaining positive and open to seeing any risks.

Joining us via a prerecorded message, **Philipp Amann** provided his thoughts about the areas of risk that continue to expand in the Metaverse and how Cybercriminals will mold to the new opportunities to gather data and use it for the wrong reasons. Also, if there are no regulations and requirements for companies/organizations to protect the

2. Spivack, J., & Berrick, D. (2023, December 12). Risk Framework for Body-Related Data in Immersive Technologies. Future of Privacy Forum. https://fpf.org/blog/risk-framework-for-body-related-data-in-immersive-technologies/

data, we will see increased data breaches. We already see the deep fakes, voiceovers, and use of these platforms to harass and cyberbully rapidly increasing. Philipp discussed many risk areas, which we, as those focused on improving the protections, need to drive forward with.

**Philip Corona Fraga** discussed the aspect of AI security and the use of anonymous data for training; however, if not used correctly, the impact could be very high. His comment, "AI requires reconsiderations and decisions that should not only be made by technologies or programmers. This requires people."

For the last portion of the Roundtable, Jordan Wiseman hosted a Q&A Discussion. During this, there were several fantastic comments made by **Dr. Joshua Sipper**, UAF, "here are so many areas with massive amounts of data that we do not include in many our discussions about extended reality and safety that deal with the deep web, and that goes all the way from military and government offices down to corporate down to even private local governments." Another very insightful set of comments from **Andreea Ion CojoCaru**, VR Developer. She talks about how people are "starting to identify more with the virtual avatar or with a certain behavior that's only available to them in the virtual world than with the physical world."

As the session wrapped up, it was very apparent that all the speakers, XRSI Team members, and those attending the sessions were given a substantial amount to think about and determine how they can contribute to the protection of individuals, training, and awareness of organizations and be diligent in helping this growing technology improve our society and not damage it or the people involved. We have options, but partnership and alignment on focus areas allow us to take steps forward.

# THEMATIC SESSION 1: THREAT LANDSCAPE IN THE AI-ENABLED METAVERSE: IDENTIFYING RISKS AND VULNERABILITIES

This first thematic discussion focused on the AI-powered metaverse's evolving risk and threat landscape. The panel explored the complex interplay between digital advancements and their security and risk impacts on virtual and physical realms.

Central to the roundtable was exploring the convergence of AI and data within the metaverse. This convergence was identified as a catalyst for an expanded cyber threat landscape, encompassing traditional digital risks and new AI-driven vulnerabilities. The discussion highlighted the transformation of cyber risks in the increasingly intertwined virtual and physical worlds, including sophisticated identity theft, data breaches, and advanced cyber-attacks.

**Philip Amann**, former head of strategy at the European Cybercrime Center (EC3), Europol, set the stage for the thematic discussion by emphasizing the significance of comprehending and confronting the risks and challenges associated with Metaverse and artificial intelligence. He highlighted the historical context of these technologies, dating back to the 1960s and even earlier, underlining the continuity of virtual interactions over time. Philip emphasized the convergence of technology and data, making Metaverse accessible to a broader audience and stressing the importance of recognizing new and traditional forms of cyber threats, such as identity theft, impersonation, data breaches, and cyberbullying. His insight shed light on the need for a collaborative effort between various stakeholders to build a safe and secure Metaverse, enabling innovative business models while mitigating risks.

*"We need to be aware of what is happening already, how traditional crime will move and has moved into the metaverse, how new types of crime, abusing artificial intelligence to automate attacks, will be used by criminals, and then build systems that are secure, safe, and resilient."*

**- Philipp Amann, Former EC3, Europol**

The potential for AI to automate and enhance criminal activities was a topic of considerable concern. The discussion pointed to the increased ease of impersonation and the growing dependency on digital systems daily. The implications of AI in crafting sophisticated phishing schemes and creating false identities were recognized as significant threats.

Most of the conversation revolved around the psychological and cognitive implications of emerging threats in the metaverse. The shift in identity perception, with individuals increasingly identifying with their digital avatars, was a key change in how identity and safety are understood in virtual environments. This shift raised questions about the nature of digital rights, suggesting that metaverse violations could be tantamount to human rights infringements due to the digitization of human presence and data.

The borderless nature of cybercrime in the metaverse was another focal point. The absence of geographical boundaries in virtual spaces posed unique challenges in law enforcement and jurisdiction, leading to hyper-personalized and global cybercrime. The discussion underscored the need for cross-border cooperation and clear jurisdictional frameworks to address these emerging threats effectively.

The technological advancements and their impact on user experience in the metaverse were also examined. The conversation explored the evolution of interfaces, from headsets to potential holographic technologies, highlighting the varied levels of comfort and adaptation across generations. This aspect emphasized the need for inclusive and accessible technological advancements to ensure a broad adaptation to these new virtual experiences.

**Key Messages:**

- **Convergence of Technology and Data in the Metaverse:** The merging of AI and virtual environments creates a complex landscape where digital and physical realities blur, bringing new risks and opportunities.
- **Dual Nature of AI:** AI is highlighted as both a tool and a target in the metaverse, involved in creating immersive experiences and also in new forms of cyber threats.
- **Expanded Threat Landscape:** The discussion identifies a spectrum of cyber threats, including traditional risks like data breaches and novel AI-driven vulnerabilities such as sophisticated impersonation and AI-based attacks.
- **Psychological and Cognitive Threats:** A shift in identity perception is noted, with individuals increasingly identifying with their avatars, raising concerns about psychological and cognitive impacts in the metaverse.
- **Digital Rights as Human Rights:** The digitization of human presence and data in the metaverse calls for reconsidering digital rights and violations as human rights issues.
- **Borderless Nature of Cybercrime:** The absence of geographical boundaries in the metaverse presents unique challenges in law enforcement and jurisdiction, leading to hyper-personalized and global cybercrime.

- **Technology and User Experience:** The role of different technologies, like headsets and holographic systems, in shaping user experiences in the metaverse is debated, alongside the varied adaptation levels across generations.
- **Need for Collaborative Security Efforts:** There is a call for joint efforts among industries, academia, and user communities to develop secure and resilient virtual environments.
- **Rethinking Concepts of Identity and Safety:** The discussion urges reevaluating how identity and safety are defined in the virtual world in light of the unique challenges posed by the metaverse.

In conclusion, the roundtable provided an exhaustive overview of the current and foreseeable challenges in the AI-powered metaverse. The discussion called for a collaborative effort among various sectors to develop secure, resilient, and ethical virtual environments. It highlighted the importance of rethinking traditional concepts of identity, safety, and rights in this rapidly evolving digital landscape, advocating for a proactive and collaborative approach to navigating the new challenges of the cyber frontier.

# THEMATIC SESSION 2: FOSTERING RESILIENCE: DATA PROTECTION AND RESPONSE STRATEGIES IN THE METAVERSE

The thematic discussion focused on the imperative need for robust data protection and incident response strategies within the AI-powered metaverse. This session, led by Kavya Pearlman from XRSI and featuring various experts, delved into the challenges and opportunities for ensuring user data privacy and security in an environment increasingly characterized by the integration of AI and immersive technologies.

The primary objective of the thematic session was to formulate approaches for developing dynamic, responsive data protection and privacy frameworks specifically designed for immersive digital environments. Additionally, it aimed to develop collaborative resilience strategies for securing emerging technologies, ensuring a comprehensive and forward-looking approach to safeguarding the metaverse against evolving cyber threats.

**Pablo Corona's**, President of the Internet Association MX intervention in the thematic session set a foundational stage, highlighting the intricate challenges of integrating AI into the metaverse and the critical importance of data protection. By their intrinsic design, he emphasized that AI systems necessitate vast amounts of personal and sensitive data for their learning and evolution. This dependency on extensive data collection introduces significant risks, particularly in a digital landscape susceptible to misinformation and breaches of trust. Corona stressed the imperative need for dynamic and adaptable regulatory frameworks that can keep pace with the rapid technological advancements in the metaverse, ensuring that data protection is both resilient and flexible enough to accommodate the evolving nature of these immersive digital environments.

*"The integration of AI in the metaverse, while innovative, introduces complex changes in data protection. AI systems, by nature, require a vast amount of information and data to learn and evolve. This is often sensitive data, the lifeblood of the metaverse, and the risks in processing it cannot be overstated. It is not just about bits and bytes; it is about a breach of trust, a breach of the social fabric. Thus, developing comprehensive frameworks for data protection requires adaptive and dynamic nature rules and regulations for the metaverse."*

**- Pablo Corona, President of the Internet Association MX**

The session emphasized the necessity of developing data protection frameworks adaptable to the metaverse's dynamic nature. Kavya Pearlman pointed out that traditional data protection models are insufficient for this new digital realm's fluid and expansive nature. The discussion underscored the need for comprehensive frameworks that are not just defensive against current threats but also predictive of future challenges, ensuring a trustworthy and secure cyber environment.

**Dr. Joshua Sipper** introduced an innovative concept known as the 'glass box' model in contrast to the traditional 'black box' model of AI. This approach is designed to make AI operations more transparent and understandable, thus enhancing trust and security in AI-driven systems.

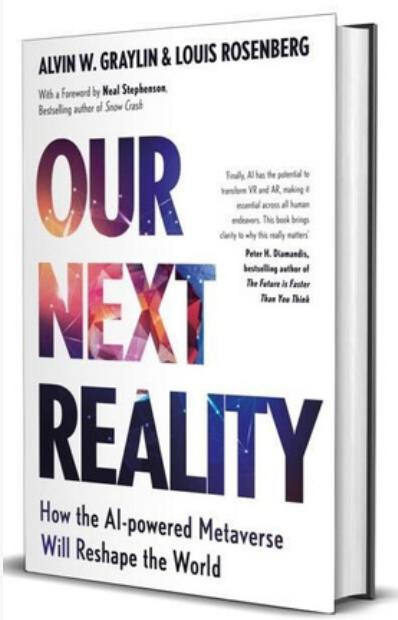
The role of training and awareness in cybersecurity was also acknowledged. The session highlighted the need to educate users on the risks of sharing personal information and the importance of implementing secure practices. Moreover, discussions highlighted the importance of the insurance industry in developing new products to mitigate risks like ransomware attacks.

**The session brought about several key recommendations, as follows:**

- Dynamic, adaptable data protection frameworks are crucial for the ever-evolving metaverse.
- AI integration in the metaverse requires a balanced approach to protect user data while maintaining privacy and anonymity.
- Transparency in AI operations, through concepts like the 'glass box' model, is essential for building trust and understanding in AI systems.
- Increased user awareness and training are crucial to strengthening the weakest link in cybersecurity - the user.
- Collaborative approaches involving regulators, industry leaders, and other stakeholders are necessary for effective cybersecurity strategies in the metaverse.
- Insurance products and other innovative solutions can play a significant role in mitigating emerging cyber risks.

The thematic session on fostering resilience in data protection and response strategies in the metaverse highlighted a multidisciplinary approach to tackling the complex challenges in this rapidly evolving digital environment. The insights gathered underscore the importance of a proactive, collaborative approach involving diverse stakeholders. Preparing for future challenges in the metaverse will require technological solutions, innovative regulatory frameworks, user education, and industry collaboration to ensure a secure, inclusive, and ethically sound digital future.

# FEATURED INTERVENTION BY ALVIN WANG GRAYLIN AND SPOTLIGHT ON HIS UPCOMING BOOK: OUR NEXT REALITY




Co-authored with AI and XR pioneer **Dr. Louis Rosenberg,** Alvin Wang Graylin shared the details and upcoming launch of his new book - **Our Next Reality: How the AI-powered Metaverse will Reshape the World**. Over the last 100 years, technology has changed our world. Over the next decade, it will transform our reality. As a society, we are not prepared. This book will provide a clear picture of what is coming and what we can do to push the future towards a more positive outcome. The book is published by Hachette Book Group.[3] The digital version will be released on March 5, 2024, and the hardcover on June 4, 2024.

Critical insights from Alvin's upcoming book and discussion include the following:

- The seamless integration of AI and XR in the metaverse will profoundly influence all life aspects of life, requiring robust security and privacy measures.
- The metaverse blurs physical and digital borders, creating complex security enforcement and jurisdiction challenges that demand global cooperation and innovative approaches.
- The potential for hyper-personalization in the metaverse raises significant risks of identity theft, impersonation, and manipulation, necessitating advanced countermeasures and public education to protect individuals' privacy and agency.

3. Hachette Book Group. (2023, December 21). Hachette Book Group. LinkedIn. https://www.linkedin.com/company/hachette-book-group/

In his intervention, Alvin W. Graylin of HTC emphasized the critical juncture we are at with the emergence of an AI-powered metaverse, foreseeing its significant integration into all aspects of life. He expressed concerns about cybersecurity, privacy, and the preservation of individual agency within this new realm. Drawing on his extensive background in security, AI, and XR technologies, Alvin highlighted the multifaceted nature of potential threats, ranging from network attacks to subtle, pervasive manipulation of human behavior and decision-making.

He stressed the importance of preparing for these challenges on multiple levels: consumer, corporate, and national security. Alvin pointed out that as the metaverse eliminates traditional borders, enforcement, and jurisdiction will become increasingly complex, requiring global cooperation and innovative solutions. He also discussed the potential for hyper-personalization in the metaverse to lead to unprecedented levels of identity theft and impersonation, exacerbated by the sophisticated capabilities of AI.

> *"As a society, we are not prepared. We are at a pivotal moment where the AI-powered metaverse will soon permeate our lives. It is crucial that we ensure cybersecurity, privacy, and maintain our agency, as these technologies will intricately weave through everything we do. If we manage these issues well, the beneficial effects of AI + XR will make the world increasingly abundant and equitable."*
>
> **- Alvin Graylin Wang, HTC**

In summary, Alvin W. Graylin provides a compelling and comprehensive analysis of the imminent challenges and threats posed by the convergence of AI and the metaverse. His insights underline the urgency for a multi-layered defense strategy, international collaboration, and continued innovation in safeguarding our cognitive infrastructure against cyber threats' sophisticated and evolving landscape.

# STRATEGIC INTELLIGENCE GATHERING SESSION VIA SWARM AI

In the thematic session focused on navigating the cyber frontier in the AI-powered metaverse, 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 proved pivotal in reaching a consensus during the roundtable, focusing on the imperative of accountability in AI governance. The collective intelligence process emphasized fostering a responsible organizational culture and highlighted multiple concerns. Discussions converged on the need for greater transparency and education to navigate AI's complexities in the metaverse, explicitly addressing the challenge posed by AI that acts on instructions—a testament to the nuanced understanding of AI's influence and the critical nature of its oversight.

**Question for deliberation: Which metaverse threat actor is the most worrisome?**

**Response: AI acting on instructions**

After a comprehensive evaluation and consensus-building exercise utilizing Swarm AI® technology, the roundtable participants reached a significant conclusion on the most alarming threat actor in the metaverse. The collective insight prioritized concerns over AI acting on instructions as the principal threat actor, highlighting the opaque nature of AI directives and the potential for misuse when instructions are not transparent or ethically grounded. This decision underscores the need for rigorous scrutiny of AI objectives and the development of clear guidelines

for AI behavior in the metaverse. The dialogue also highlighted the differing global stances on data privacy, emphasizing the urgency for establishing universal norms to combat intrusive data practices linked to AI operations.

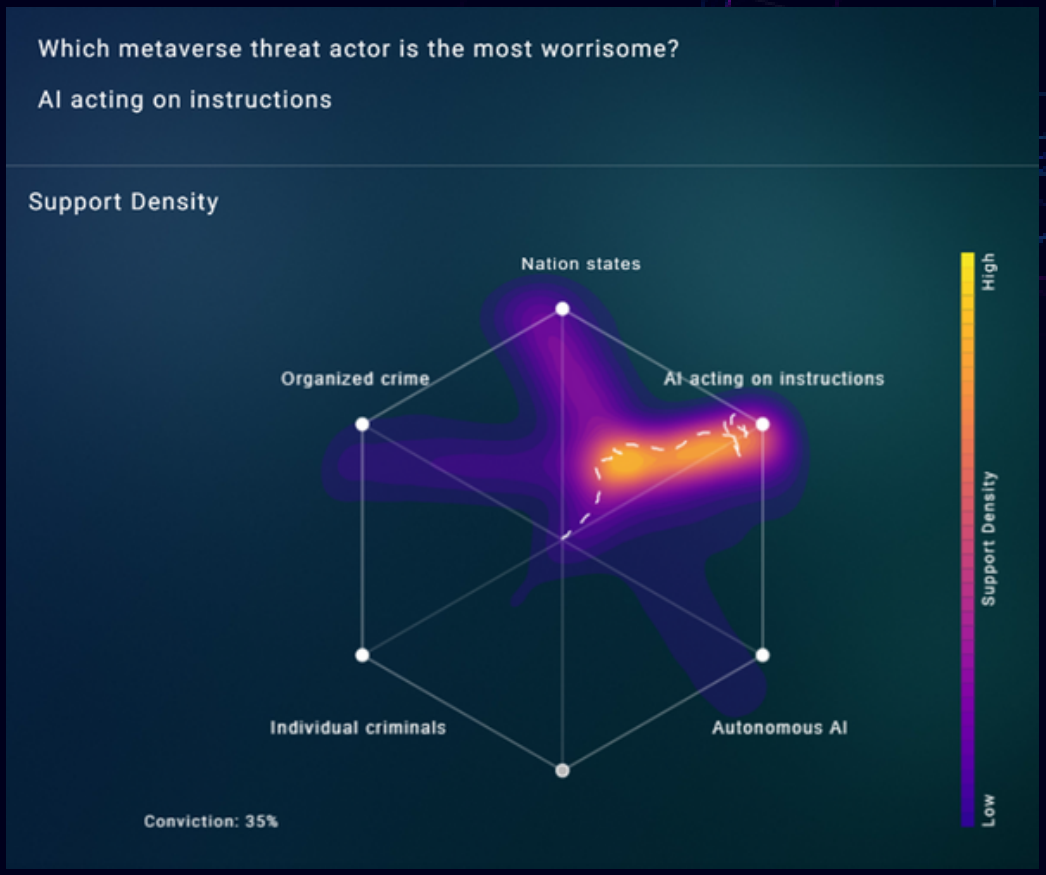



FIG -1

**Question for deliberation: What is the most common AI-related challenge organizations face?**
**Response: Lack of awareness**

After an intensive and collaborative session utilizing Swarm AI® technology, the consensus among participants is the recognition of 'Lack of awareness' as the most significant AI-related challenge organizations face today (Fig - 2). This discernment brings to light the critical gap in understanding that pervades across sectors, where misconceptions about AI's potential and limitations lead to suboptimal or even harmful deployment of AI solutions. The clear imperative is for organizations to cultivate an AI-savvy culture where knowledge of AI capabilities, risks, and ethical considerations is disseminated widely and deeply. Such an

environment will enhance the effective use of AI and safeguard against its misuse. To thrive in an AI-augmented future, pervasive education in AI literacy is beneficial and essential for all organizational stakeholders.

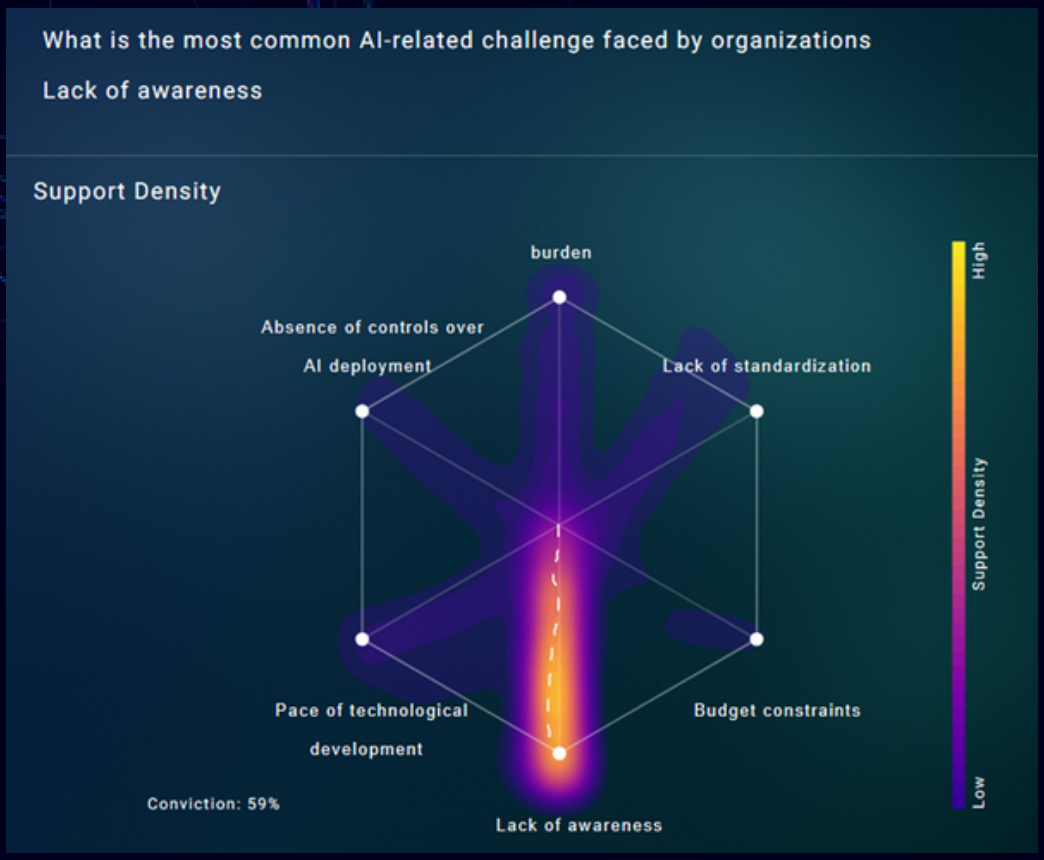



FIG -2

Utilizing Swarm AI, the roundtable harnessed a wealth of collective expertise, crafting outcomes reflective of a comprehensive and unified perspective. This collaborative intelligence was instrumental in devising robust, forward-thinking strategies to navigate the complex terrain of cybersecurity and data protection within the AI-powered metaverse, underpinning our commitment to a secure digital future that safeguards user privacy and promotes resilient infrastructures.

Fig 2

# CYBERSECURITY RISKS AND DATA PROTECTION CHALLENGES IN THE AI-POWERED METAVERSE

The fourth annual Metaverse Safety Week has highlighted growing concerns about cybersecurity risks and data protection challenges in the AI-powered metaverse. This evolving digital realm presents unique, complex issues requiring immediate and strategic attention.

## CYBERSECURITY THREATS IN THE AI-POWERED METAVERSE: CLASSIC, AI-ENHANCED, AND EMERGING RISKS

A detailed exploration of specific threats within the digital landscape of the metaverse is essential for gaining a deeper understanding of the potential risks and meaningful impacts on communities, companies, and individuals. The following section will consider example threats such as AI manipulation, issues surrounding NFTs, and the challenges posed by reality distortion.

Threats in the AI-powered metaverse come in three flavors: classic cybersecurity and privacy threats, those enhanced through AI and XR technologies, and novel threats emerging from these new technologies. The following sections briefly outline some examples of these threats and considerations for risk management in the AI-powered metaverse. This section delves into various threats within the metaverse's digital landscape, discussing potential impacts on communities, companies, and individuals.

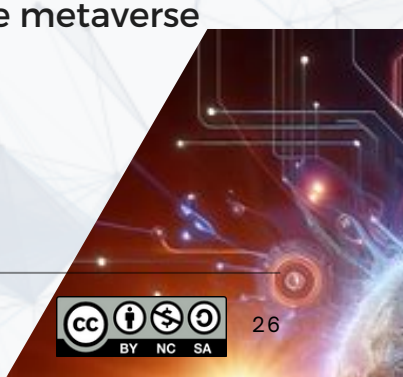
## CLASSIC CYBERSECURITY THREATS IN THE METAVERSE:

Classic cybersecurity threats, traditionally present in cyberspace, are expected to have significant implications in the metaverse environment. These classic cybersecurity threats in the AI-powered metaverse underscore the need for robust security measures and increased user education to mitigate risks and protect users in these evolving digital landscapes. These include social engineering, hardware vulnerabilities, virtual property and asset security, phishing attacks, DDoS attacks, ransomware, malware targeting VR/AR devices, insider threats, and account takeovers.

- **Phishing Attacks in Virtual Environments:** Similar to traditional phishing, this involves deceptive practices where users are tricked into revealing sensitive information within the metaverse. Attackers could use sophisticated AI to create highly realistic scenarios or avatars that mimic trusted entities, duping users into disclosing personal data, login credentials, or financial information. The immersive nature of the metaverse might make it harder for users to distinguish between legitimate and fraudulent interactions.

- **Distributed Denial of Service (DDoS) Attacks:** In the metaverse, attacks could overwhelm and incapacitate virtual platforms or services. This could involve flooding a metaverse server with superfluous requests, rendering it unable to process legitimate requests. Such attacks could disrupt the user experience significantly, causing financial loss and damaging the reputation of metaverse platforms.

- **Ransomware in Virtual Spaces:** Similar to its real-world counterpart, ransomware in the metaverse would involve encrypting a user's virtual assets or access, with demands for payment in exchange for restoration. This could include locking users out of their virtual properties, hijacking control of avatars, or encrypting valuable digital assets. The interconnectedness and value of assets in the metaverse could make ransomware especially damaging.



- **Malware and Exploits Targeting VR/AR Devices:** As users access the metaverse through various devices like VR headsets or AR glasses, these endpoints become targets for malware and exploits. Attackers could use malware to gain unauthorized access to a user's device, steal data, or manipulate metaverse experiences. The personal nature of these devices also raises concerns about privacy violations.

- **Insider Threats and Account Takeovers:** In the metaverse, insider threats could involve individuals with legitimate access to metaverse platforms or systems misusing their privileges for personal gain or to cause harm. Account takeovers, where attackers gain control of a user's metaverse account, pose a significant threat, given the likely integration of real-world identity and financial information within these accounts.

- **Social Engineering and Hardware Vulnerabilities:** The immersive nature of the metaverse opens new avenues for social engineering attacks, exploiting the interaction between users and virtual environments. Attackers could manipulate users by targeting vulnerabilities in VR hardware, leading them to divulge sensitive information.

- **Virtual Property and Asset Security:** Securing virtual assets poses unique challenges in the metaverse. Unauthorized access to or theft of virtual property, including avatars, can result in financial and experiential losses for users.

- **Virtual Currency and NFT Security:** Using virtual currencies and non-fungible tokens (NFTs) in the metaverse introduces new financial security and asset management concerns. The novelty and lack of regulation around these digital assets make them attractive targets for cybercriminals.

## ENHANCED THREATS BY AI IN THE METAVERSE:

Integrating AI in the metaverse, a realm of immersive technology, significantly heightens the complexity and impact of cybersecurity threats. Sophisticated cyber-attacks, privacy concerns heightened by immersive experiences, and the dynamic nature of threats amplified by AI necessitate agile and adaptive response mechanisms. This section explores the diverse ways in which AI magnifies traditional cybersecurity challenges, underscoring the need for sophisticated security measures and robust regulatory frameworks.

- **Enhanced Cyber Attacks:** AI-enhanced cyberattacks in the metaverse showcase unprecedented sophistication. One notable example is deepfakes for identity theft or misinformation dissemination. Attackers can create hyper-realistic scenarios or avatars to deceive users, such as AI-generated figures impersonating trusted individuals, leading to sensitive information leaks or user manipulation.

- **Privacy and Protection Concerns:** Privacy issues are significantly elevated in the metaverse due to immersive technology and AI. The voluminous data generated by user interactions necessitates dynamic and resilient privacy protection frameworks. AI algorithms can analyze extensive behavioral data, leading to potential unauthorized exploitation or surveillance.

- **Dynamic Nature of Threats:** The metaverse's evolving landscape demands adaptive mechanisms to counteract AI-enhanced threats. AI's capability to instantly modify virtual environments introduces new security challenges. For instance, AI could exploit vulnerabilities in real-time or subtly influence user behavior, demonstrating the need for adaptable security protocols.

- **NFT-Related Asset Theft and Fraud:** In the metaverse, NFTs represent a unique area of risk, further complicated by AI technologies. The potential for AI to facilitate counterfeit NFT creation or transaction manipulation is significant. An attacker might use AI algorithms to clone or forge NFTs, disrupting the virtual asset market.

- **Deepfake Exploitation:** Deepfakes can create compelling scenarios or avatars in the metaverse. This could lead to identity theft, misinformation, or manipulation of users. In the metaverse, where interactions rely heavily on virtual appearances and representations, the impact of deep fakes could be even more significant than in traditional digital spaces. This threat is particularly potent in scenarios where AI creates avatars or environments for deceptive purposes.

- **Dynamic Metaverse Manipulation:** The highly dynamic and evolving nature of the metaverse, driven by AI, could lead to new forms of manipulation and pose unique risks. This includes altering the virtual environment in real-time to deceive or disorient users, dynamically changing scenarios to exploit vulnerabilities, or creating events that have real-world consequences.

- **AI-Powered Social Engineering:** Enhanced AI capabilities amplify the effectiveness of social engineering attacks within the metaverse. By mimicking real people and utilizing behavior-based strategies, AI-driven avatars could persuasively extract sensitive information from unsuspecting users.

- **Exploitation of Real-Time Data and Surveillance:** The metaverse enables extensive surveillance and real-time data collection, with AI systems capable of analyzing this data for user tracking, profiling, or behavioral manipulation. This raises significant privacy concerns, such as AI systems using real-time data to influence user decisions or experiences without their knowledge.

- **Extensive Data Collection by Immersive Technology:** Immersive technology devices, like VR headsets and AR glasses, collect a broad spectrum of user data, including biometric information. The risk of this data being exploited is heightened without stringent privacy controls, leading to potential unauthorized use for targeted advertising or behavior influence.

- **AI-Driven Algorithmic Bias:** In the metaverse, AI algorithms might exhibit biases based on the data they are trained on. This could lead to unfair treatment of certain user groups or discriminatory practices within virtual environments, necessitating careful oversight and ethical AI design.

- **Manipulation of Virtual Economies:** AI can be used to manipulate virtual economies within the metaverse, such as artificially inflating the value of virtual assets or manipulating market trends. This could lead to economic instability and financial losses for users.

- **AI-Enhanced Network Intrusions:** The metaverse's interconnected networks are vulnerable to AI-enhanced intrusions, where AI systems can rapidly identify and exploit network vulnerabilities, leading to widespread data breaches or service disruptions.

## NOVEL AND EMERGING THREATS IN THE METAVERSE:

The metaverse, integrating advanced AI and immersive technologies, presents novel and emerging threats that challenge traditional cybersecurity paradigms. These threats not only exploit technological vulnerabilities but also raise significant ethical, psychological, and social concerns. Phantom Timeline Syndrome challenges users' ability to differentiate between virtual and real events, avatar identity threats, transparency issues in AI decision-making (Glass Box vs. Black Box), AI manipulation, reality distortion, and risks associated with Brain-Computer Interfaces (BCIs) and connected medical devices and much more.

- **Phantom Timeline Syndrome:**[4] This refers to a disorienting experience where users might struggle to differentiate between events occurring in the virtual metaverse and those in the real world. As AI technologies create more immersive and convincing virtual environments, this syndrome could lead to confusion or misremembering of events, blurring the lines between virtual experiences and real-life memories. It may also impact a person's sense of time, as they navigate between virtual and physical realities.

- **Avatar Identity Threats:** In the metaverse, avatars represent users' identities. Avatar identity threats encompass identity theft, where someone may impersonate another user's avatar or unauthorized usage and manipulation of a person's avatar. This can lead to reputational damage, privacy violations, and even psychological distress for the victim. As avatars become more personalized and tied to real-world identities, the potential impact of such threats increases.

4. Nichols, S. (2022, February 7). Metaverse rollout brings new security risks, challenges. TechTarget. https://www.techtarget.com/searchsecurity/news/252513072/Metaverse-rollout-brings-new-security-risks-challenges

- **Glass Box vs Black Box AI:** This challenge pertains to the transparency and explainability of AI algorithms. A 'glass box' AI is one where the decision-making process is transparent and understandable to users. In contrast, a 'black box' AI operates opaquely, with little to no insight into how it reaches its conclusions. In the metaverse, the prevalence of black-box AI systems can raise concerns about bias, fairness, and accountability, particularly when these systems make decisions that affect user experiences or digital assets.

- **AI Manipulation Problem:**[5] This threat involves the potential for AI systems in the metaverse to be manipulated or exploited to achieve malicious objectives. This could include altering AI algorithms to bias their outcomes, using AI for deepfakes or other deceptive practices, and even leveraging AI to manipulate user behavior or perceptions within the virtual environment. Such manipulation can have far-reaching implications, from spreading misinformation to influencing personal decisions.

- **Reality Distortion:** In the metaverse, the distinction between what is real and what is virtual can become increasingly blurred. Reality distortion is where the virtual world's experiences might significantly alter a user's real-world perception. This could manifest in altered social behaviors, changes in perception of physical spaces, or even mental health impacts. The concern is that prolonged or intense exposure to AI-driven virtual environments could lead to a persistent altered state of reality perception

- **Brain-Computer Interface (BCI) Risks:** As BCIs become more integrated into the metaverse, they pose unique security risks such as Brainjacking.[6] The primary concern is unauthorized access to a user's neural data, potentially leading to severe privacy breaches or manipulation of the user's actions. This raises profound ethical concerns about the sanctity of individual thought and control.

5. Rosenberg, L. (2023, June 19). The Manipulation Problem: Conversational AI as a Threat to Epistemic Agency. arXiv. https://arxiv.org/abs/2306.11748
6. University of Oxford. (2016, August 24). Brainjacking – a new cyber-security threat. University of Oxford. https://www.ox.ac.uk/research/brainjacking-%E2%80%93-new-cyber-security-threat

- **Medical Device Vulnerabilities in the Metaverse:** The increasing connectivity of medical devices within the metaverse, especially in applications like telehealth, exposes them to new cyber threats. Hackers targeting these devices could compromise critical health functions or steal sensitive health data. The interconnected nature of these devices means a single exploit could have widespread ramifications.

- **Virtual Addiction and Dependency:** Prolonged engagement in the metaverse could lead to virtual addiction, where users become overly dependent on virtual experiences, neglecting real-world responsibilities and relationships. This could have significant psychological and social implications.

- **AI-Induced Echo Chambers:** AI algorithms in the metaverse might create personalized experiences that inadvertently result in echo chambers, where users are exposed only to ideas and perspectives similar to their own. This could limit exposure to diverse viewpoints and reinforce biases.

- **Ethical Dilemmas in Virtual Interactions:** As AI drives more complex interactions in the metaverse, it raises ethical questions about the nature of these interactions, particularly concerning AI-driven characters that exhibit human-like behaviors. Determining the ethical boundaries for interactions with these entities becomes a critical concern.

These novel and emerging threats in the metaverse highlight the necessity for a multidisciplinary approach to security, encompassing technological solutions, ethical guidelines, and psychological and social considerations. Addressing these threats requires collaborative efforts from technologists, ethicists, psychologists, and policymakers to ensure a secure, ethical, and psychologically safe metaverse.

# CONVERGING PHYSICAL AND VIRTUAL REALITIES: RISKS AND CHALLENGES

- **Operational Risks:** Integral to the metaverse's functioning, AI systems are vulnerable to malfunctions and cyberattacks. These issues can cause significant disruptions in virtual environments, impacting virtual economies, social interactions, and digital identities. For example, malfunctions in AI moderation tools could escalate uncontrolled cyberbullying or spread harmful content.

- **Business Existential Risks:** Companies face existential threats due to significant breaches or failures to adapt to the metaverse's dynamics. Data breaches can erode user trust and result in considerable financial damages. A notable risk is the breach of user avatars, leading to a loss of user base and company credibility.

- **Strategic Risks:** Underestimating the value of digital assets or failing to engage in virtual communities effectively can lead to missed market opportunities. Businesses that overlook the potential of virtual real estate or NFTs may lose out on significant revenue streams.

- **Reputational Risks:** In the metaverse, where relationships form a core aspect of the user experience, cybersecurity incidents can lead to reputational damage. For instance, privacy breaches in social VR platforms can result in user attrition and decreased engagement.

- **Financial Risks:** Financial transactions, especially those involving cryptocurrencies and NFTs, are susceptible to fraud and theft. The volatile nature of these virtual assets poses significant investment risks, such as financial losses due to NFT fraud.

- **Regulatory Risks:** The emerging regulatory framework for the metaverse presents compliance challenges. Lack of clarity in virtual transactions and data privacy regulations necessitates careful navigation by businesses to avoid legal issues.

- **Legal Risks:** The metaverse's legal landscape includes complex issues like intellectual property disputes and liability for virtual actions. The global nature of the metaverse adds to the complexity of jurisdiction and enforcement.

- **Personal Risks:** For individuals, the metaverse brings risks to privacy and personal security. Identity theft, harassment, and emotional manipulation within virtual spaces can lead to psychological distress and reputational damage. For example, the hacking of a user's avatar for inappropriate use can have serious personal repercussions.

## PRIVACY AND DATA PROTECTION CONCERNS IN THE AI-POWERED METAVERSE

Privacy risks in the metaverse are particularly acute due to the depth and nature of data involved in immersive experiences. As users interact within these spaces, they generate an unprecedented amount of personal data, from basic demographic information to sensitive behavioral patterns. This data can include:

- **Biometric Data:** VR and AR devices often require biometric data for personalized experiences, raising concerns about using such sensitive information. This includes XR Data[7] and Biometrically inferred data.[8]
- **Location Data:** The continuous tracking of a user's location within the metaverse can lead to privacy infringements if such data is accessed or misused by unauthorized parties.
- **Communication Data:** Private conversations and interactions within the metaverse could be subject to eavesdropping or data mining, threatening user confidentiality.
- **Behavioral and Preference Data:** The metaverse's ability to track and analyze user behavior and preferences poses significant privacy risks. This data could be exploited for targeted advertising or malicious purposes like manipulation or profiling.

7. X Reality Safety Intelligence (XRSI). (2023, December 21). XR Data. XRSI. https://xrsi.org/definition/xr-data
8. X Reality Safety Intelligence (XRSI). (2023, December 21). Biometrically-Inferred Data (BID). XRSI. https://xrsi.org/definition/biometrically-inferred-data-bid

In the AI-powered metaverse, privacy risks are significantly amplified due to the extensive nature of data collection and processing. Users generate vast amounts of personal data, and the integrated AI systems continuously process and analyze this data, raising substantial privacy concerns. This extensive data collection and AI processing include risks associated with biometric data collection, location tracking, communication monitoring, behavioral data analysis, and the AI's ability to infer additional personal information from collected data. The AI's deep learning algorithms, designed to enhance user experience, can inadvertently expose users to privacy breaches if not governed by robust and adaptable data protection frameworks. These frameworks are crucial in safeguarding user privacy against potential misuse of AI-processed data in the dynamic and evolving environment of the metaverse.

Robust data protection frameworks are critical in addressing these privacy concerns. Such frameworks must be adaptable to the dynamic and evolving nature of the metaverse, ensuring that user privacy is safeguarded against emerging threats. Moreover, as the metaverse continues to grow and intersect with our physical realities, these dialogues and strategies will be pivotal in shaping a secure, inclusive, and ethically grounded digital future.

*"Our current incentives often reward hiding mistakes or misfortunes from the community for fear of inviting regulatory actions or public criticism. Mistakes are the best teachers, especially when we can learn from others. We must encourage and incentivize sharing lessons learned and hard-won knowledge, especially about cybersecurity and data usage and involving innovative technologies."*

**- Alexander McD White, Privacy Commissioner, Bermuda**

# SECURING THE DIGITAL FRONTIER: CYBERSECURITY AND DATA PROTECTION STRATEGIES FOR THE METAVERSE

In the ever-evolving landscape of technology and digital innovation, the Metaverse emerges as a transformative frontier that promises boundless possibilities. However, the promise of the Metaverse also brings a new realm of challenges, particularly in cybersecurity and data protection, which require immediate attention and action. The MSW 2023 Cybersecurity and Data Protection Focused Track, organized by GPA and XRSI, convened experts, professionals, policymakers, and enthusiasts to deliberate on these pressing issues. Through rigorous discussions and evaluation of current and emerging intelligence sources and contributions from roundtable participants, the programming committee formulated recommendations, along with specific measures aimed at safeguarding the integrity and security of the AI-powered Metaverse, as follows:

- **Advanced Threat Intelligence Sharing Hub:** Establishing a dedicated Metaverse Threat Intelligence Sharing Hub or Information Sharing and Analysis Center (ISAC) is recommended to combat dynamic cyber threats. An ISAC is an independent nonprofit organization providing a central resource for gathering information on cyber and related threats to critical infrastructure and providing two-way information sharing between the private and public sectors. Such a center will continuously monitor and analyze threats specific to the Metaverse.

- **Responsive Data Protection Frameworks:** In response to data privacy challenges, the committee advocates for creating and adopting a Data Privacy and Safety Framework. This framework should focus on balancing safety and privacy, outlining how user data is collected, processed, and protected in immersive environments. A baseline framework that is already in the works is the XRSI Privacy and Safety Framework,[9] the adoption of which could serve as the foundation for an agile framework that is crucial for complementing the regulatory efforts with various privacy and security controls and functions for various stakeholders. The Future of Privacy Forum (FPF) Risk Framework for Body-Related Data in Immersive Technologies can be used by organizations to structure the collection, use, and onward transfer of body-related data.[10]

- **Secure AI Development:** Given AI vulnerabilities, developers must prioritize creating secure AI and ML algorithms resistant to manipulation and misuse in immersive environments. Collaboration with cybersecurity experts in the development lifecycle is essential. Organizations must Integrate cybersecurity experts into AI development teams[11,12] and conduct regular code audits to identify and mitigate vulnerabilities. This is where the principle of Shift Left could be helpful in collaboration and testing that's traditionally done at a later stage of the process and perform that task at earlier stages of development.[13]

9. X Reality Safety Intelligence (XRSI). The XRSI Privacy and Safety Framework 2. XRSI. https://xrsi.org/the-xrsi-privacy-and-safety-framework-2
10. Spivack, J., Berrick, D. (2023, December). Risk Framework for Body-Related Data in Immersive Technologies. Future of Privacy Forum. https://fpf.org/wp-content/uploads/2023/12/FPF-Risk-Framework-for-Body-Related-Data-FINAL-Digital.pdf
11. ISO/IEC 42001:2023 - Artificial intelligence Management system (December 2023). https://www.iso.org/standard/81230.html
12. ISO/IEC 5338:2023 - Artificial intelligence AI system life cycle processes (December 2023) https://www.iso.org/standard/81118.html
13. Souppaya, M., Scarfone, K., & Dodson, D. (2022, February). Secure Software Development Framework (SSDF) Ver. 1.1: Recommendations for Mitigating the Risk of Software Vulnerabilities. NIST Special Publication 800-218. https://nvlpubs.nist.gov/nistpubs/specialpublications/nist.sp.800-218.pdf

- **Incident Response Strategies:** Corporate leaders should implement comprehensive Metaverse-specific Incident Response Plans to address security incidents swiftly. A common framework for incident reporting would enable global consistency, interoperability, and terminology alignment between regulatory or self-regulatory AI incident reporting in different jurisdictions ahead of implementing mandatory or voluntary incident reporting schemes, as planned notably in the European Commission's proposed AI Act.[14] The OECD is engaging with policymakers, experts, and partners from all stakeholder groups to develop a common framework for AI incident reporting,[15] which could serve as the foundation for the strategic measure.

- **Cultivating a Safety by Design Culture:** Foster a culture of security and privacy by default and design within organizations by conducting regular Metaverse security awareness training for employees.[16] This could be accomplished by launching a Metaverse and AI intersection-focused security awareness program targeting all employees. This also applies to the regulators and policymakers focused on creating and updating laws, regulations, and governance frameworks for immersive environments.

- **Regulations and Collaboration:** Legal professionals, government officials, regulators, and policymakers must develop and enforce Metaverse-specific regulations that ensure responsible innovation and protection. International collaborations are essential to establish uniform cybersecurity standards and policies for the global Metaverse community. The OECD Global Forum on Technology is a venue for foreseeing regular in-depth dialogue. It could be the hub for getting ahead of the long-term opportunities and risks presented by emerging technologies.[17]

14. European Commission. (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206
15. OECD. (2023, December 21). Expert Group on AI Incidents. OECD https://oecd.ai/en/network-of-experts/working-group/10836
16. World Economic Forum. (2023, January 18). Metaverse Privacy and Safety: A Global Risk Report. World Economic Forum https://www3.weforum.org/docs/WEF_Metaverse_Privacy_and_Safety_2023.pdf
17. OECD. (2022, December). Global Forum on Technology. OECD. https://www.oecd.org/digital/global-forum-on-technology

- **Education and Innovation:** The convergence of various emerging technologies is accurate, and ignoring the implications could be dangerous. Academics should incorporate Metaverse-specific cybersecurity and data protection challenges into curriculums and research. This can be done by revising the academic curriculums to encourage the development of innovative tools and methodologies for securing emerging technologies in the Metaverse.[18]

- **Community Engagement:** All stakeholders, including technology enthusiasts and early adopters, should actively engage in community-driven initiatives to promote awareness and safe practices in immersive environments. Organize and participate in community-led Metaverse safety workshops such as MIT Reality Hack[19] and adopt awareness campaigns such as Metaverse Safety Week as the opportunity presents.[20]

- **Strengthening Cognitive Critical Infrastructure:** In the context of the converging realms of AI and the metaverse, establishing a "Critical Infrastructure Protection"[21] framework becomes imperative for safeguarding a nation's cognitive infrastructure. This and infrastructure, crucial in the digital age, include the collective mental capacities and decision-making abilities of citizens, organizations, and governments and the information channels that fuel these cognitive processes.

18. GovTech. (2022, November 18). Metaverse Meets Higher Ed: Security, Privacy, Safety Concerns. GovTech. https://www.govtech.com/education/higher-ed/metaverse-meets-higher-ed-security-privacy-safety-concerns
19. MIT Reality Hack. (Year Unknown). About — MIT Reality Hack | AR/VR Metaverse Hackathon. MIT Reality Hack https://www.mitrealityhack.com/about
20. Metaverse Safety Week. (2023, December 10-15). About Metaverse Safety Week. Metaverse Safety Week. https://metaversesafetyweek.org/about/
21. Pereira, D. (2022, January 4). National Cognitive Infrastructure Protection: What Can We Learn from the Swedish Psychological Defence Authority? OODA Loop. https://www.oodaloop.com/archive/2022/01/04/national-cognitive-infrastructure-protection-what-can-we-learn-from-the-swedish-psychological-defence-authority

> *"One could argue we are in the business of critical infrastructure, and the most critical infrastructure is our cognitive infrastructure, so building that resilience to misinformation and disinformation, I think, is incredibly important."*
>
> **- Jen Easterly, CISA Director**[22]

As AI and the metaverse create increasingly immersive and influential digital environments, the framework should address the heightened risks of misinformation, cognitive manipulation, and data privacy breaches. It should focus on enhancing cybersecurity measures specific to AI and virtual environments, promoting media literacy that includes understanding AI-driven content and virtual interactions, and fostering critical thinking skills tailored to navigating these advanced technologies. Collaboration across sectors is essential, involving government agencies, technology companies, educational institutions, and civil society to develop strategies that protect the integrity of information and decision-making processes in this new digital frontier.

This approach ensures the resilience of cognitive infrastructure against the unique challenges posed by the AI-metaverse convergence. As we venture into the AI-powered Metaverse, these recommendations and concrete measures aim to empower stakeholders to navigate the complex cyber frontier collaboratively. By identifying and understanding cyber threats, developing responsive data protection frameworks, and fostering a culture of security, we can ensure the safety, rights, and data integrity of every individual in the digital realm.

Global communities, industry leaders, and policymakers are responsible for innovating responsibly, creating a Metaverse that is secure, equitable, and beneficial for all. This report serves as a collective call to action, urging us to rise to the challenge and build a Metaverse that thrives on safety, security, and respect for individual privacy.

22. Committee on the Judiciary. (2023, June 26). The Weaponization of CISA: How a "Cybersecurity" Agency Colluded with Big Tech to Censor Americans. Committee on the Judiciary. https://judiciary.house.gov/sites/evo-subsites/republicans-judiciary.house.gov/files/evo-media-document/cisa-staff-report6-26-23.pdf

# 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 adopt MSW as an annual awareness campaign, 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 GPA

The Global Privacy Assembly (GPA) is an international network of data protection and privacy authorities that aims to provide leadership and guidance in this rapidly evolving field. Founded in 1979, the GPA has grown to include more than 130 members from across the globe, representing diverse legal systems and cultures. The GPA facilitates cooperation and information exchange among its members, as well as engaging with other stakeholders and international organizations on data protection and privacy issues.

GPA
Global Privacy Assembly

The GPA also organizes an annual conference, where members adopt resolutions, declarations, and reports on various topics of common interest. The GPA's website serves as a permanent repository of its core documents and activities and a platform for sharing news and updates from its members and working groups. The GPA envisions a world where privacy and data protection authorities can effectively fulfill their mandates and promote the rights and interests of individuals.

# CO-ORGANIZER DETAILS

## ABOUT XRSI

Headquartered in the San Francisco Bay Area in the United States and Torino, Italy, in Europe, X Reality Safety Intelligence (XRSI) is the world's leading organization dedicated to providing intelligence and advisory services that are vital for the protection and well-being of emerging technology ecosystems. With a strong emphasis on critical aspects such as safety, privacy, security, human rights, human well-being, responsible innovation, governance, and regulation, XRSI offers comprehensive expertise to ensure the responsible and ethical advancement of emerging technologies.

XRSI™
Human Intelligence In The Loop

By placing the emphasis on Human Intelligence, XRSI brings together a global network of experts and thought leaders committed to shaping the future of technology in a way that prioritizes the welfare of individuals and society as a whole. We offer standardization, certification, policymaking, and workforce development professional advisory services for the emerging technology domain.

> *"As the Events Director of MSW 2023, I'm thrilled by the resounding success of our campaign, bringing together brilliant minds dedicated to safeguarding the intersections of AI and emerging technologies. Our achievements this year stand as a testament to the collective dedication of XRSI advisors, our volunteers, community partners, supporters, and our sponsors. I'm excited to continue this vital work beyond 2023, forging a future where innovation thrives hand-in-hand with safety and responsibility."*
>
> **- Bhanujeet Choudhary, XRSI | MSW 2023 Events Director**

# APPENDIX 1: OUTPUT FROM SWARM AI SESSIONS

| Question & Answer | Conviction |
|---|---|
| What is most critical for implementing responsible AI governance?<br>**Answer: Accountability** | **58%** |
| How strongly do we Agree or Disagree with this statement?<br>**Answer: Disagree** | **65%** |
| Initial focus for orgs starting their responsible AI adoption journey:<br>**Answer: Cultivate a Culture of Responsibility** | **69%** |
| Which newer strategy promises to best manage metaverse risks?<br>**Answer: Virtual Reality (VR) Simulations for Risk Scenarios** | **51%** |
| What is the best way to mitigate risks in the AI-driven metaverse?<br>**Answer: BRAIN FREEZE!** | **0%** |
| What is the most common AI-related challenge faced by organizations<br>**Answer: Lack of awareness** | **59%** |
| Which AI harm has the most significant impact on society today?<br>**Answer: Manipulation and Misinformation** | **65%** |
| Which metaverse threat actor is the most worrisome?<br>**Answer: AI acting on instructions** | **35%** |
| Was the call for "pausing" AI a good way to reduce risk?<br>**Answer: No, it's too late to put the genie back in the bottle** | **53%** |
| Which of these areas of metaverse-related risk is the most dangerous?<br>**Answer: BRAIN FREEZE!** | **0%** |

## LEAD AUTHORS

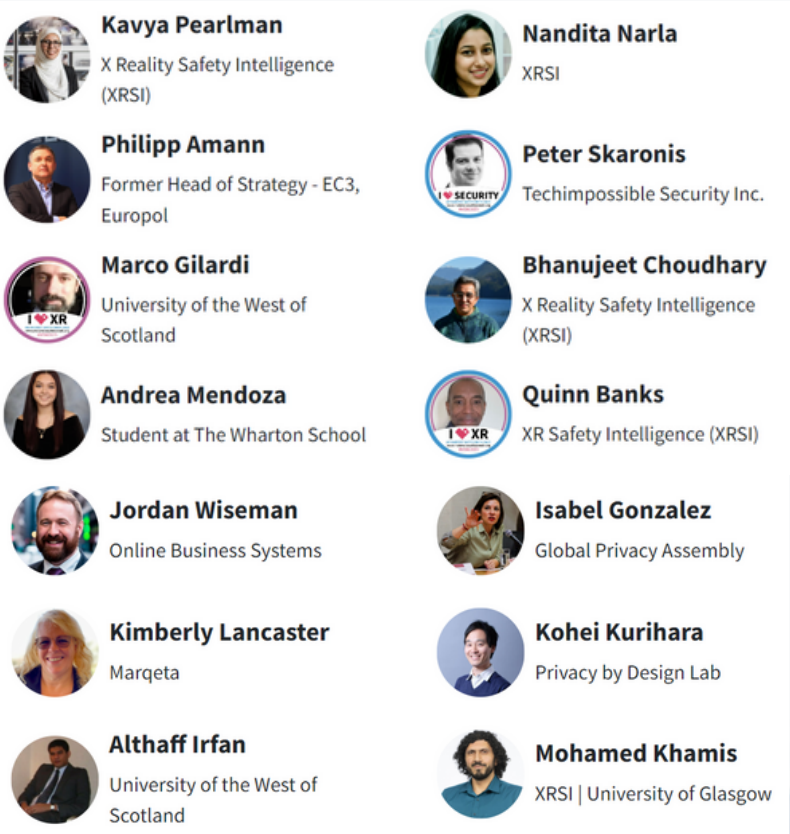
**Kavya Pearlman**
Founder & CEO, XRSI
Chair - Cybersecurity and Data Protection, MSW2023

**Nandita Narla**
Advisor, XRSI
Co-Chair - Cybersecurity and Data Protection, MSW2023

**Kimberly Lancaster**
Advisor, XRSI | Director of Privacy & Data Protection, Marqeta

**Peter Skaronis**
CEO, Techimpossible Security Inc.

# PROGRAMMING COMMITTEE MEMBERS: CYBERSECURITY AND DATA PROTECTION



# ACKNOWLEDGEMENTS

**Florian Krueger**
Unyted

**Scott Phillips**
Vaulted Ventures

**Will Kreth**
HAND (Human & Digital)

**Liam Coffey**
AMD

**Paul Lanois**
Fieldfisher (also adjunct faculty - University of California College of Law in San Francisco

**Jaime Schwarz**
XRSI & Brand Therapy LLC

**Alathaff Mohideen**
University of the West of Scotland

**Kimberly Lancaster**
Marqeta | XRSI

**Quinn Banks**
ARCH Sahn Inc. | XRSI

**Vance Lockton**
Office of the Privacy Commissioner of Canada

**Anna Collard**
KnowBe4 Africa

**Jordan Wiseman**
Online Business Systems | XRSI

**Tomas Petru**
Goldilock Secure Ltd.

**Joshua Sipper**
Air Command and Staff College

**Kimberly Hieftje**
Yale School of Medicine

**Peter Skaronis**
Techimpossible Security Inc. | XRSI

**Kohei Kuirhara**
Privacy by Design Lab | XRSI

**Andreea Ion CojoCaru**
NUMENA

**Anmol Agarwal**

**Francesco Pagano**
XRSI

**Selin Fidan**

# PARTICIPATING ENTITIES AND PARTNERS



















# PARTICIPATING ENTITIES AND PARTNERS

METAVERSE SAFETY WEEK
4th annual edition | DEC 10-15, 2023
www.metaversesafetyweek.org

**DISCLAIMER:** This report, edited and published by XRSI - X Reality Safety Intelligence, originates from the Cybersecurity and Data Protection roundtable held during Metaverse Safety Week 2023, centered on "Navigating the Cyber Frontier in the AI-Powered Metaverse", convened on December 13th, 2023. The information contained herein is intended solely for informational purposes.



ORGANIZED BY
XRSI
Human Intelligence In The Loop
www.xrsi.org

CYBERSECURITY AND DATA PROTECTION