METAVERSE SAFETY WEEK

4th annual edition | DEC 10-15, 2023

HUMAN RIGHTS IN THE AI-POWERED IMMERSIVE WORLDS

DECEMBER 10, 2023

THE ROUNDTABLE REPORT

CO-ORGANIZED BY:

Australian Human Rights Commission

XRSI
Human Intelligence In The Loop

SUPPORTED BY:

Meta

www.xrsi.org

www.Metaversesafetyweek.org

TABLE OF CONTENTS

01 Executive Summary

02 Introduction

03 The Roundtable Overview

04 Thematic Session 1: Fostering cultural diversity and tackling discrimination in the AI-powered immersive worlds

05 Thematic Session 2: Digital identity and data dignity: Extending Human Rights for access, ownership, and mental autonomy

06 Spotlight : UN's report on governing AI for humanity

07 Strategic intelligence gathering sessions via Swarm AI

08 Balancing Human and Artificial Intelligence: Challenges to Human Rights in the Metaverse

09 Shaping a responsible digital future with Human Rights at the heart of the metaverse

10 Strategic recommendations for human-centric digital innovation

11 Call to Action: Adopt MSW

12 Co-Organizer Details: About AHRC and XRSI

13 Appendix 1: Output from Swarm AI sessions

14 Programming committee members: International Human Rights Day

15 Contributor Acknowledgements

16 Participating Entities and Partners

EXECUTIVE SUMMARY

As we honor the 75th Human Rights Day, we delve deep into the intertwining of AI-powered immersive worlds and their profound implications on human rights. In an era marked by rapid technological convergence, the Metaverse and AI jointly revolutionize industries, necessitating a fresh perspective on established human rights doctrines. Bringing together global thought leaders, our focus sharpens on the emerging concepts of neurorights and the transformative potential of Brain-Computer Interfaces (BCI). The pivotal task in a dominantly AI-orchestrated Metaverse is to fortify and safeguard human rights. We engage in robust discussions surrounding the multifaceted challenges posed by privacy, autonomy, and the looming risks of algorithmic biases. Drawing insights from the collective intelligence, we accentuate the urgency for erecting standardized guardrails.

The Dawn of a New Era in Human Rights: This period marks a significant shift in how we perceive and protect human rights in the age of AI and the Metaverse. The advent of neurorights and the implementation of BCIs signify a transformative era, necessitating a robust defense of human rights against potential infringements by technological advancements.

Challenges and Prospects in AI-Driven Realities: The intricate challenges of AI within the Metaverse demand a comprehensive approach. There is a recognized need for standardized frameworks to manage these complexities effectively. The insights gathered from diverse experts underscore the necessity for regulatory frameworks that balance innovative progress and the unwavering commitment to human rights and dignity.

Strategic Intelligence Gathering by AHRC and XRSI: The Human Rights roundtable, the first track of the 4th Annual Metaverse Safety Week 2023, was co-organized by the Australian Human Rights Commission (AHRC) and X Reality Safety Intelligence (XRSI). This roundtable brought together

a diverse assembly of global experts, including human rights advocates, legal scholars, technologists, and policymakers, to delve into the complex interplay between human rights and AI-driven Metaverse environments. The discussions focused on integrating human rights principles within these emerging digital spaces, with particular attention to neurorights, the ethical use of Brain-Computer Interfaces (BCI), and adapting traditional human rights frameworks to AI's challenges and opportunities. This event signified a critical milestone in forging a comprehensive, rights-based approach for the future of AI and immersive technologies, aiming to ensure that human dignity and fundamental rights remain central in the evolution and application of these transformative technologies.

Despite the complexities and risks involved, the dialogue highlighted the potential to harness AI-powered immersive realities positively. A key outcome was the collective resolve to anchor human rights at the forefront of technological innovation. This commitment is vital for preserving mental autonomy, promoting inclusivity, and giving to the underrepresented.

The Human Rights Roundtable at Metaverse Safety Week 2023 encapsulates the concerted effort to weave human rights considerations into the ongoing technological evolution. It marks a pivotal moment in the journey towards an equitable and responsible digital future, where human rights are preserved and integral to technological development and application.

Monika Manolova
Advisor, XRSI
Chair - International Human Rights Day, MSW2023

Patrick Hooton
Human Rights Advisor, AHRC
Co-Chair - International Human Rights Day, MSW2023

INTRODUCTION

As the world commemorated the 75th anniversary of the Universal Declaration of Human Rights[1] on December 10, the significance of this milestone resonated through the Global Roundtable Discussion on Human Rights. Hosted during the World Metaverse Safety Week in 2023, this event, co-organized by the AHRC and XRSI, marked a crucial juncture in the discourse on digital rights in the age of AI and Metaverse technologies.

Set against the backdrop of the unprecedented adoption of large language models (LLMs) and the growing integration of AI into various products and services, the roundtable sought to delve into the nuanced impactts of AI-augmented realities on human rights. In a world increasingly dominated by AI, safeguarding human rights within these new digital landscapes has become a pivotal challenge.

The exchange of ideas, propelled by calls to action from various global entities, catalyzed a shared vision for collaborative progress and creating safer, more inclusive digital environments. Marking a significant milestone in the ongoing evolution of digital rights, this Global Roundtable Discussion on Human Rights laid the groundwork for a future where human rights are seamlessly integrated into our digital existence.

KEY OBJECTIVES INCLUDED

- Mapping the aspects of human rights within the **current convergence tech environment.**
- Defining key next steps for global digital communities to align behind in **the pursuit of equitable digital spaces.**
- Providing recommendations on the importance of digital dignity in **human-centric digital environments.**

1. United Nations. (1948, December 10). Universal Declaration of Human Rights. United Nations. https://www.un.org/en/about-us/universal-declaration-of-human-rights

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 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 10, 2023, the Australian Human Rights Commission(AHRC) and X Reality Safety Intelligence (XRSI) organized a strategic intelligence roundtable focused on how the fight for human rights can and must expand to account for the capabilities of emerging and immersive technologies. 2023 Metaverse Safety Week, an annual awareness campaign to promote a safe and positive experience within immersive environments commenced with highlighting a gap between existing human rights and the need for additional rights and considerations for human well-being in digital 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 and human rights within the Metaverse. The Roundtable was three hours of exciting and engaging discussions from experts in various areas who brought their experience about "Human Rights in the AI-powered Immersive Worlds" to the audience.

The hosts, **Monika Manolova** and **Kavya Pearlman** were joined by the following individuals for the discussion. **April Boyd-Noronha** also supported the discussion, with support from several XRSI Team Members and Advisors in the background.

On the notable occasion of the 75th anniversary of the Universal Declaration of Human Rights, **Lorraine Finlay** from the Australian Human Rights Commission addressed the profound intersection of neurotechnology and AI-powered immersive worlds with human rights. She emphasized that these technologies, in their essence, are neutral; their impact hinges on their application. Central to her address was advocating for human rights to be at the forefront of developing and deploying neurotechnologies and related AI-powered products. Finlay expressed concerns about the lack of sufficient safeguards and regulations around these evolving technologies, particularly highlighting the potential erosion of privacy rights, a fundamental human right. She acknowledged the immense potential of neurotech in enhancing the

acknowledged the immense potential of neurotech in enhancing the inclusion of disabled individuals in digital spaces and expanding human capabilities. However, she stressed the importance of continuing efforts towards establishing robust safeguards and recognizing neurorights as integral to human rights.[2]

The following distinguished panelists each brought forth their knowledge and explored actionable strategies and collaborative initiatives to ensure that AI-powered immersive worlds are designed and governed with a central emphasis on human rights, integrating diverse perspectives and ethical considerations to foster inclusive, equitable, and respectful digital spaces.

Melodena Stephens of the Mohammed Bin Rashid School of Government followed with thought-provoking questions about human rights in digital environments where refined replicas of people are possible. With sufficient data, it may be possible to recreate individuals' voices, images, and works. An emphasis was placed on rights that people have over these representations of their likenesses and skills, noting that to foresee a future where companies replace their workforce with AIs or utilize reproductions of dead individuals against the wishes of their

2. Human Rights in AI-powered Immersive Worlds. (2023, December 11). Australian Human Rights Commission. https://humanrights.gov.au/about/news/speeches/human-rights-ai-powered-immersive-worlds

families. A few winners often dominate the digital economy, while many creators struggle to earn a fair share of their value. Roblox, a platform that hosts millions of user-generated games, is a case in point. Only a small fraction of its creators can make a living from their creations, while the rest are left out of the revenue stream.

Judith Okonkwo of Imisi 3D, an XR creation lab in Africa, discussed the challenges to equitable access to the Metaverse. XR technologies could provide huge benefits for developing nations, particularly in areas such as education, where they could help fill gaps in facilities and access to equipment often taken for granted. These technologies need significant guardrails to keep humans safe, respect privacy, and an informed populace that understands the benefits and risks involved. She stated that equitable access to technology is critical in educating the wider populace.

Dylan Fox of XR Access made the case that investing in accessibility for XR technologies[3] lays the groundwork for many advanced features. For example, much of the infrastructure used in displaying captions can also be used for real-time translation, and text alternatives for objects enable screen readers and improve search and localization capabilities. He emphasized that the design and development of these technologies should always involve the people directly affected by them, not just those who are able-bodied. It is the essence of the "Nothing About Us, Without Us" design approach,[4] which he considered crucial.

Marisa Zalabak, an IEEE contributor and educational psychologist, spoke about the dangers inside and outside the headset. In addition to emotional abuse suffered in VR, there are risks of addiction and overreliance; too much time in VR could potentially affect a user's sense of reality and social skills. She highlighted The need for comprehensive digital education and literacy programs to help combat these adverse effects and the importance of reporting systems for abuse that do not simply pass the responsibility to local organizations that are ill-equipped to deal with it.

3. XR Access Symposium. (2023). Highlights from the 2023 XR Access Symposium: Pioneering Accessibility in Virtual Realms. Symposium Insights Journal. https://xraccess.org/wp-content/uploads/2023/11/2023-Symposium-Report.pdf
4. BOIA. (2023, February 24). Nothing About Us, Without Us: Starting Digital Accessibility Conversations. Bureau of Internet Accessibility. https://www.boia.org/blog/nothing-about-us-without-us-starting-digital-accessibility-conversations

Finally, **Marisa Borsboom** of Humanity of Things pointed out that securing human rights in a system more beholden to investors than ethicists is challenging. She highlighted that with XR regulations still in an embryonic state, the critical importance of civil society involvement and inclusion of independent knowledge and voices in the discussion, both in order to prevent harm to children and neurodiverse people and in order to demand accountability.

The participants in the Human Rights roundtable collectively addressed pressing issues such as the ethical implications of neurotechnology, privacy concerns in AI-augmented environments, and the broader societal impacts of emerging technologies. The discussions also emphasized the critical need for ongoing collaboration, education, and public awareness in integrating human rights principles into the rapidly evolving landscape of digital technologies and immersive worlds.

Emphasizing the need for a human rights-centered approach, the discussion traversed various dimensions of emerging technologies, from the potential of Brain-Computer Interfaces (BCIs) to enhance lives to the profound challenges they pose to privacy and mental autonomy. Experts highlighted the imperative of integrating human rights into the development and deployment of these technologies, addressing the rapid evolution without adequate safeguards and regulations. It marked a critical moment for advocating the reinforcement of human rights in the digital age.

KEY MESSAGES FROM THE DISCUSSION:

- Speakers at the session highlighted the complex interplay between human rights and technological advancements, focusing on the **implications of neurotechnology and AI**. They stressed the importance of maintaining mental privacy and developing neurotechnologies that respect human dignity and rights.

- The conversation raised concerns about **diminishing privacy rights** in the face of advancing technologies, calling for active steps to safeguard against potential infringements. The necessity of establishing strong regulatory frameworks and fostering international cooperation to tackle these issues was consistently emphasized.

- The dialogue sheds light on the **societal and ethical ramifications of AI-enhanced environments,** touching on their effects on employment, accessibility, and social skills. Discussions delved into the concept of digital identity, underlining the legal and moral considerations necessary in developing and implementing these technologies.

- The **importance of integrating diverse viewpoints in defining the trajectory of AI and immersive worlds** was a key point of emphasis. The dialogue underscored the need for inclusive technology access, highlighting the crucial role of civil society and personal agency in promoting ethical innovation and ensuring the benefits of emerging technologies are accessible to all.

In conclusion, the roundtable set the stage for ongoing exploration and collaboration. It called for a unified approach toward integrating human rights principles in the rapidly evolving digital landscape, ensuring that technological advancements are harmoniously aligned with preserving and promoting human rights and dignity. This session marked a significant step toward a more equitable and responsible digital future where human rights are central to the evolution of AI-powered immersive worlds.

> *"Human rights aren't just for some, but need to apply to everybody, everywhere, every day. It's crucial that AI technologies, especially in their combination with neuro tech, prioritize human rights. The future should champion humanity and safeguard our fundamental rights, including the right to privacy, ensuring these technologies benefit us all"*
>
> **- Lorraine Finlay, Australian Human Rights Commission**

THEMATIC SESSION 1: FOSTERING CULTURAL DIVERSITY AND TACKLING DISCRIMINATION IN THE AI-POWERED IMMERSIVE WORLDS

The thematic discussion on fostering cultural diversity and tackling discrimination in the AI-powered immersive worlds examines the integration of diverse cultural perspectives in AI technology within virtual environments. It highlights the significant challenges associated with biases and discrimination that can be embedded in AI algorithms, and the consequent impacts on different cultural groups.

A key theme element was the emphasis on inclusive design and diverse representation in developing extended reality systems with integrated AI. The discussion explored the development of effective strategies and policies to not only counteract discrimination but also to foster cultural diversity and inclusivity actively, ensuring these considerations keep pace with the dynamic advancements in virtual and augmented reality technologies. It was focused on several critical themes related to challenges in bias, accessible design, universal accessibility, and the infringement of human rights within immersive realities.

Nico Fara, a tech executive, set the stage for the thematic discussion by highlighting the challenges of representation in immersive technologies. As a woman of color who has witnessed and experienced the challenges women face in Iran, she emphasized the power of technology to bring about positive change. She noted that her years of developing a diverse representation ecosystem have been the most significant setback to the incredible potential of disruptive technologies. She also emphasizes the incredible potential to harness these technologies to unite humanity and celebrate diversity.

"The destiny of the Metaverse is unwritten, and we can still build these new worlds to empower voices both loud and unheard."

- Nico Fara, Let's Go!

The session brought about several key recommendations, as follows:

- **Promote Diversity and Inclusivity:**
  - Advocate for diversity mandates across standard bodies to foster universally accessible and equitable digital spaces.
  - Encourage the participation of underrepresented groups in developing and auditing AI systems to identify and address biases.

- **Comprehensive Bias Audits:**
  - Regularly audit AI systems for prejudices, with active involvement from diverse perspectives.
  - Challenge the notion of innovation masking discrimination, ensuring transparent and inclusive technological evolution.

- **Hiring for Diverse Perspectives:**
  - Emphasize hiring individuals with varied life experiences and cultural competencies.
  - Recognize that diversity in thought leads to breakthroughs and a balanced representation in digital environments.

- **Global Values and Inclusivity:**
  - Advocate for a global values approach, recognizing the distinct cultural nuances and needs across different regions.
  - Focus on user-centric inclusivity by design rather than relying solely on frameworks.

- **Systemic Change and Executive Prioritization:**
  - Stress the need for systemic change with executive-level commitment to diversity and accessibility.
  - Explore innovative roles, such as "CTOs - Chief Thinking Officers," to prioritize human-centric design and address emerging ethical challenges.

- **Survivor-Informed Approaches:**
  - Highlight the importance of survivor-informed approaches to technology for creating safer digital spaces.
  - Emphasize the seriousness of harassment in digital spaces and the need for practical, informed strategies to mitigate risks.

The session advocated for a comprehensive approach that marries the innovation of AI-augmented environments with a steadfast commitment to human rights. This strategy aims to cultivate digital spaces that foster intellectual and cultural growth and safeguard the diverse spectrum of human dignity and freedoms. The dialogue underscored the collective responsibility of all stakeholders to shape a digital future that prioritizes the protection and enhancement of human rights in every aspect of emerging technologies. Recommendations from the session are geared toward striking a balance between the forward march of innovation and the imperative to ensure these advances honor and protect human values. As our digital landscapes evolve, they remain spaces where every individual's rights are recognized and respected. This endeavor is pivotal in building a future where technology serves humanity, enhancing lives while vigilantly guarding against any erosion of the rights that form the bedrock of our societies.

THEMATIC SESSION 2: DIGITAL IDENTITY AND DATA DIGNITY: EXTENDING HUMAN RIGHTS FOR ACCESS, OWNERSHIP, AND MENTAL AUTONOMY

The thematic discussion on digital identity and data dignity: Extending human rights for access, ownership, and mental autonomy delves into the crucial aspects of human rights within the context of AI-powered immersive worlds. The discussion offers an in-depth exploration of the challenges and opportunities in the ever-expanding Metaverse, focusing on ensuring equitable access to these technologies for all individuals. This theme examines the importance of maintaining user autonomy in managing digital identities and the need for establishing clear and effective data ownership rights. Additionally, it addresses the burgeoning area of neurorights, highlighting the significance of protecting mental autonomy from possible encroachments by AI and brain-computer interface technologies.

Dr. Divya Chander initiated an in-depth exploration into the nuanced subtopic of neurorights within immersive realities, emphasizing the burgeoning challenges as technology evolves. She highlighted the critical issue of "Human Autonomy," sparking a debate on the concept of "Neural Sovereignty" and its implications for individual selfhood. She also addressed the advanced capabilities of neurotechnology, which can interact with the brain's neural code. While these advancements hold promise for medical healing and addressing disorders, they pose significant concerns regarding their influence on human behavior and perception. Acknowledging the complexity of regulating such digital and cross-jurisdictional technologies, she suggested that their designation as human rights might offer a viable solution, as recognizing neurorights as human rights also transcend jurisdictions. This discussion spotlighted the urgent need for robust protective measures and ethical considerations in the rapidly advancing field of neurotechnology, ensuring the safeguarding of human autonomy and security in the face of such transformative capabilities

The session brought about several key recommendations, as follows:

- **Cybersecurity and Informed Consent in Neurotech:**
  - Emphasize the importance of cybersecurity and encryption in neurotechnology.
  - Advocate for informed consent processes, ensuring users understand the implications and risks.

- **Consumer Awareness and Rights:**
  - Enhance consumer education about neurotech risks and implications.
  - Establish the right for consumers to revoke data usage, reinforcing control over personal information.

- **Addressing Surveillance Concerns:**
  - Link neural rights to fundamental human rights, focusing on mental privacy and freedom from technological manipulation.
  - Highlight the need for protective measures against invasive surveillance practices in neurotech.

- **Identity and Data Control:**
  - Explore the implications of data storage and usage on individual identity.
  - Advocate for user rights to grant or revoke access to personal data, emphasizing autonomy.

- **Privacy and Data Security in Development:**
  - Integrate robust cybersecurity and privacy protections at every neurotech development level.
  - Prioritize safeguarding personal and sensitive data throughout technology design and deployment.

- **Alignment of Interests and Comprehensive Security:**
  - Foster alignment between end-user interests and company objectives.
  - Implement user-centric controls and comprehensive cybersecurity across all data interaction levels.

Emphasizing stakeholders' shared responsibility, the dialogue highlighted the importance of transdisciplinary approaches, involving mental health professionals in developing AI guardrails and creating EULA-like documentation[5] to protect data producers and owners. Additionally, integrating "data vaults" with robust cybersecurity measures at various levels was proposed to ensure comprehensive data protection. These recommendations, together with the innovative use of technology like data vaults, are designed to ensure that as digital landscapes evolve, they remain environments where individual rights are recognized and fiercely protected.

The session's discussion culminated in advocating for a comprehensive approach that balances advancing AI-augmented environments with a firm commitment to human rights. This strategy aims to cultivate digital spaces that encourage intellectual and cultural growth while protecting the diverse spectrum of human dignity and freedoms. Establishing a future where technology serves humanity, the session underscored the need for enriching lives and vigilantly safeguarding fundamental human values with every step forward in emerging technologies.

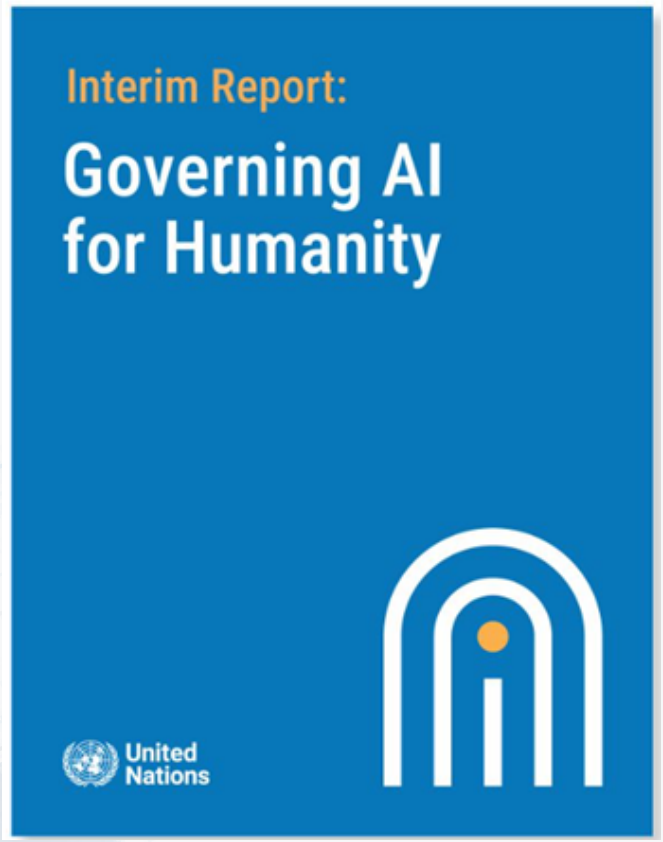
"Neurotechnologies hold the promise of a future where human well-being is significantly enhanced and suffering is greatly reduced. As we unlock the mysteries of the brain, we are developing tools that can not only monitor but also modulate neural activity. These advancements have the potential to revolutionize healthcare, offering new treatments for neurological disorders and improving our understanding of consciousness. However, as we tread this new frontier, it is crucial that we do so responsibly, ensuring that the rights and dignity of individuals are respected. The work we do today will shape the future of neurotechnology and its impact on humanity, and it is our responsibility to ensure that this future is one that upholds our fundamental human rights."

- Dr. Divya Chander, XRSI | Atlantic Council | Stimson Center

5. ServiceNow. (2023, December 21). What is an End-User License Agreement (EULA)? ServiceNow. https://www.servicenow.com/products/it-asset-management/what-is-eula.html

SPOTLIGHT : UN'S REPORT ON GOVERNING AI FOR HUMANITY

The UN AI Body was convened by the Office of the UN Secretary-General's Envoy on Technology at the end of October and, in the intervening weeks, they have sprinted to produce this initial report reflecting their collective analysis.[6] The central piece of the report is a proposal to strengthen international governance of AI by carrying out seven critical functions, such as horizon scanning for risks, supporting international collaboration on data, and computing capacity and talent to reach the Sustainable Development Goals (SDGs). It also includes recommendations to enhance accountability and ensure an equitable voice for all countries.



The UN AI Body was convened by the Office of the UN Secretary-General's Envoy on Technology at the end of October and, in the intervening weeks, they have sprinted to produce this initial report reflecting their collective analysis.

6. United Nations. (2023, August 31). High-level Advisory Body on Artificial Intelligence. United Nations. https://www.un.org/en/ai-advisory-body

The central piece of the report is a proposal to strengthen international governance of AI by carrying out seven critical functions, such as horizon scanning for risks, supporting international collaboration on data, and computing capacity and talent to reach the Sustainable Development Goals (SDGs). It also includes recommendations to enhance accountability and ensure an equitable voice for all countries.

GUIDING PRINCIPLES

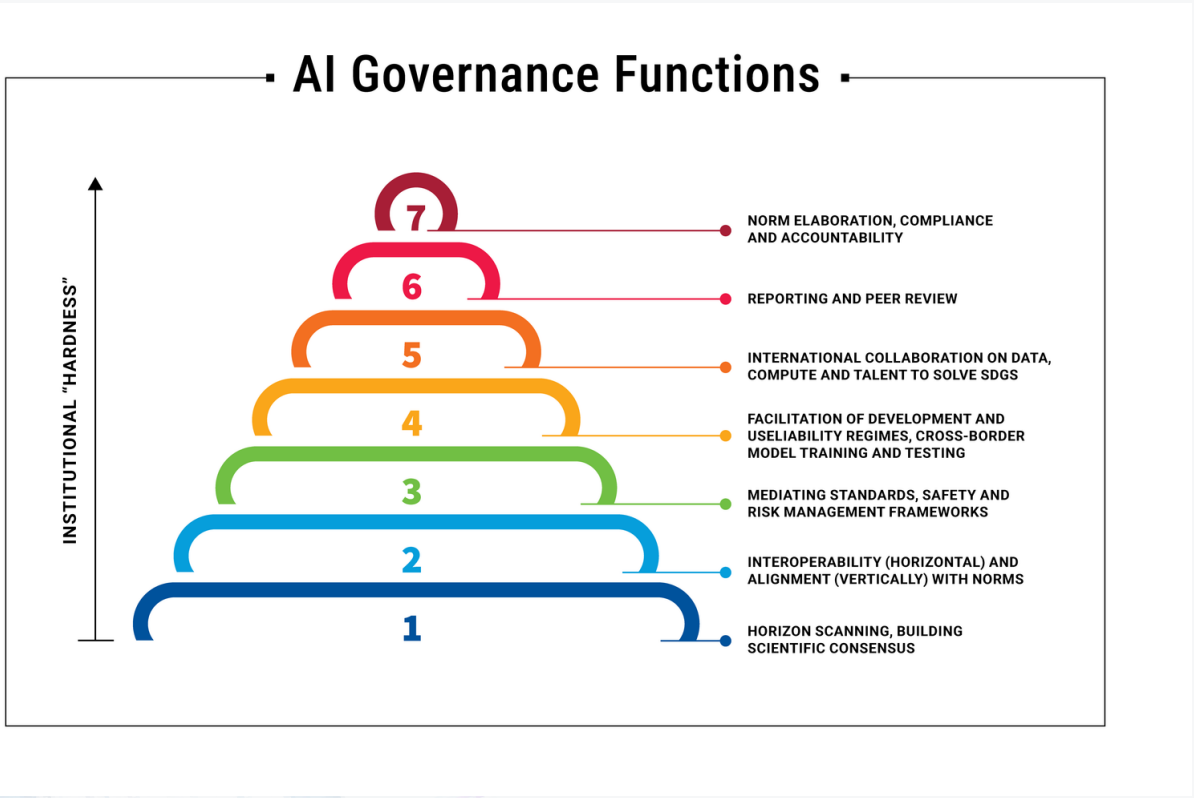
The interim report identifies the following principles that should guide the formation of new global AI governance institutions:

- **Inclusivity:** All citizens, including those in the Global South, should be able to access and meaningfully use AI tools.
- **Public interest:** Governance should go beyond the do-no-harm principle and define a broader accountability framework for companies that build, deploy, and control AI and downstream users.
- **Centrality of data governance:** AI governance cannot be divorced from the governance of data and the promotion of data commons.
- **Universal, networked, and multistakeholder:** AI governance should prioritize universal buy-in by countries and stakeholders. It should leverage existing institutions through a networked approach.
- **International Law:** AI governance needs to be anchored in the UN Charter, International Human Rights Law, and the Sustainable Development Goals.

INSTITUTIONAL FUNCTIONS

The Body identified what some functions of AI governance should entail. Among others, they include regularly assessing the state of AI and its trajectory, harmonizing standards, safety, and risk management frameworks, promoting international multistakeholder collaboration to empower the Global South, monitoring risks and coordinating emergency response, and developing binding accountability norms.

Each of these functions requires international cooperation among all actors involved.



The following steps include consultations and conversations with diverse stakeholders worldwide about the interim report and engagement with governments, the private sector, civil society, and research and technical communities.

STRATEGIC INTELLIGENCE GATHERING SESSION VIA SWARM AI

In the thematic session focused on Human Rights in the AI-powered immersive worlds, 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. Questions about user autonomy, data ownership, digital identity, human rights in AI-powered immersive worlds, and mental autonomy were raised during the Swarm AI discussion.

Question for deliberation: Most essential to prevent human rights violations by AI in immersive worlds

Response: Transparent AI Algorithm Policies

When discussing the prevention of human rights violations by AI in immersive worlds, the paramount solution identified through Swarm AI technology is the implementation of transparent AI algorithm policies (Fig -1). This approach ensures that the mechanics behind AI decision-making are open to scrutiny, enabling effective auditing and upholding the right to privacy and freedom from discrimination. To make transparency actionable, participants highlighted the essential role of public education, empowering individuals to recognize and challenge potential AI biases. The heatmap's intensity further suggests a consensus on the need for precise enforcement mechanisms, ensuring these transparent policies have the teeth to protect human rights in the burgeoning landscape of immersive technologies.

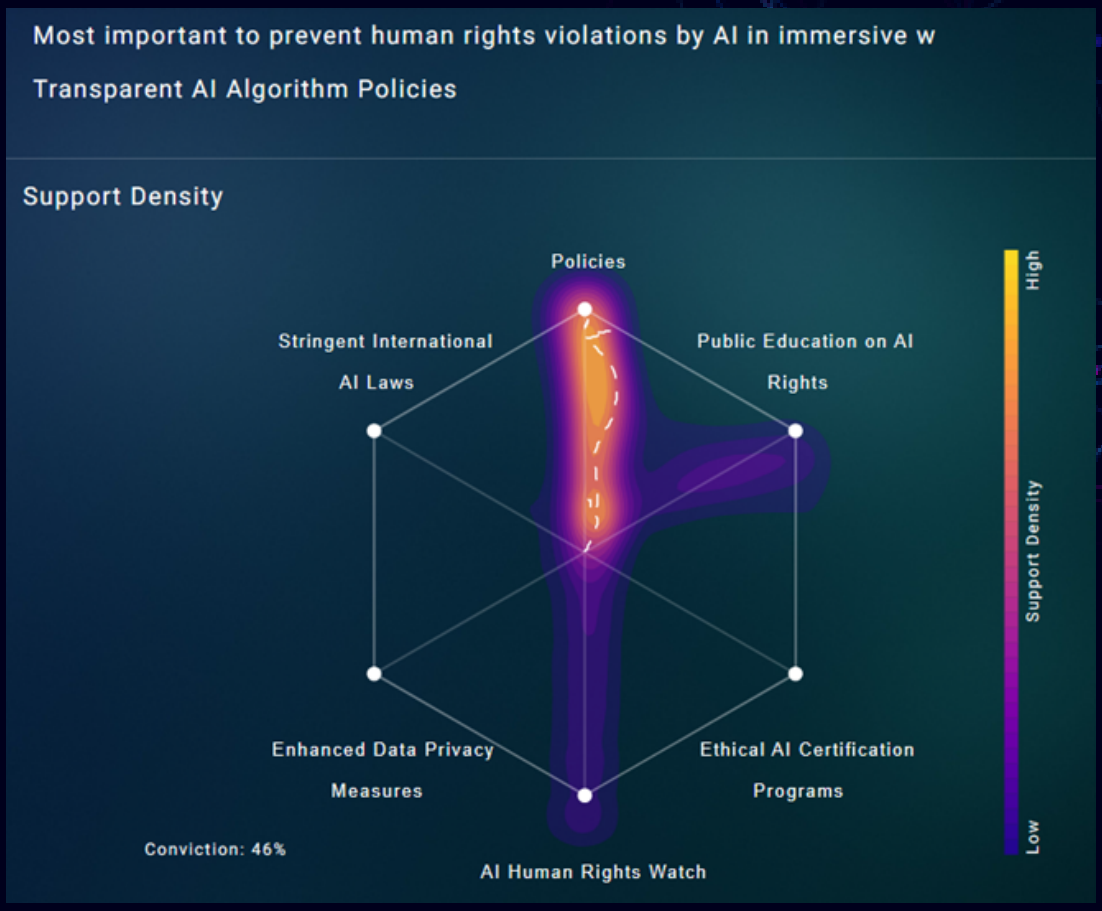


FIG -1

Question for deliberation: Human right that needs the most strengthening to ensure mental autonomy

Response: Right to Mental Privacy

A 56% of the discourse recognized this right as fundamental yet underdeveloped, necessitating urgent attention and reinforcement. (Fig -2). The right to mental privacy, although not universally codified, is increasingly relevant as neurotechnological advancements blur the lines between the sanctity of the human mind and the reach of digital tools. The consensus acknowledges the innate value of mental privacy as a human right that underpins the autonomy of thought and the integrity of individual consciousness in the face of AI's pervasive growth.

Fig 1

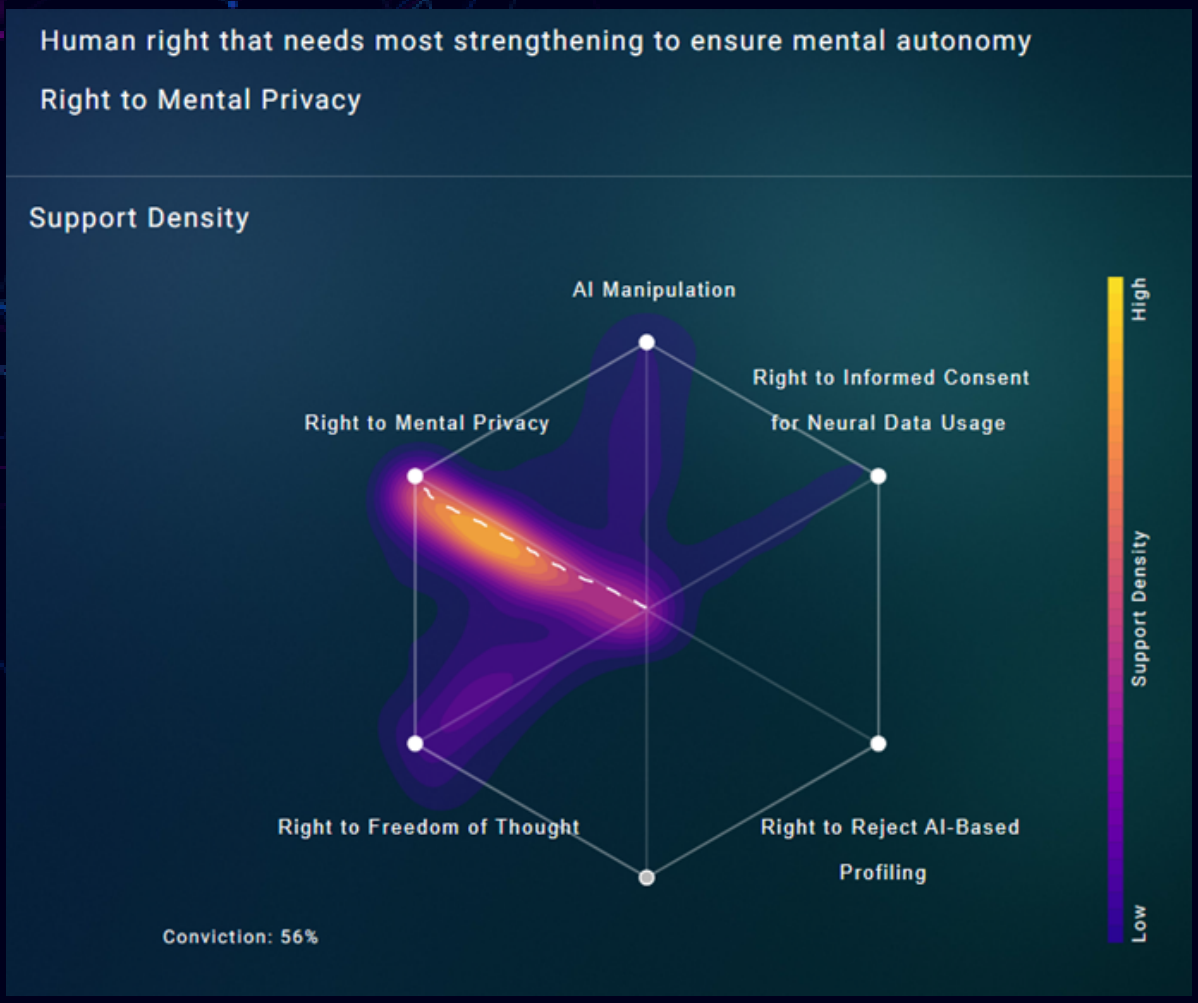


FIG -2

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 yielded comprehensive viewpoints on key issues like digital autonomy, rights integration, and equitable access in AI-driven virtual worlds. The collective intelligence gathered emphasized the importance of aligning AI development with human rights principles, ensuring technology advances respect individual dignity and freedoms. These insights form a vital guide to navigating the complexities of AI and neurotechnology, balancing innovation with the imperative of human rights protection.

Fig 2

BALANCING HUMAN AND ARTIFICIAL INTELLIGENCE: CHALLENGES TO HUMAN RIGHTS IN THE METAVERSE

The section focuses on the delicate balance between advancing neurotechnology and preserving personal privacy, emphasizing the crucial need for diversity in technological development. This section highlights the necessity of digital education to mitigate risks and ensure equitable Metaverse access, particularly for underrepresented communities. The discussions settled on the following significant risks and challenges associated with Human rights in AI-powered immersive realities.

- **Neurotechnology and Human Autonomy:**[7]
  - The interconnectedness of neurotechnology and advanced motion tracking threatens human autonomy.
  - The significant advancement of neurotechnology raises concerns about the potential endangerment of personal information and erosion of the cornerstone right to freedom of speech

- **Representation and Diversity:**
  - Lack of representation technology, especially for diverse groups, is identified as a significant setback to the robust development and innovation in the context of AI-powered immersive realities.
  - The discussion emphasizes the need for diversity in development teams and calls for actions such as advocating for diversity requirements and including individuals with varied life experiences in the design and development of technologies.

- **Equitable Access:**
  - There are challenges to ensuring equitable access to the Metaverse, particularly in nations from the Global South.
  - XR technologies, while offering benefits, require significant guardrails to protect end users, respect privacy, and inform the public.

7. X Reality Safety Intelligence (XRSI). (2023, December 21). Human Rights in the Age of Neurotechnology: Impact, Opportunities & Measures. X Reality Safety Intelligence. https://xrsi.org/publication/human-rights-in-the-age-of-neurotechnology-impact-opportunities-measures

- **Lack of Digital Education and Literacy:**
  - The risks of emotional abuse, addiction, and overreliance in VR environments underscore the urgent need for comprehensive digital education and literacy programs.
  - Reporting systems for abuse are essential to combat adverse effects, and responsibility should not only be passed to local organizations, which are often ill-equipped to handle the issues.

- **Human Rights Violations in AI-Powered Immersive Worlds:**
  - Biases and discrimination embedded in AI algorithms within immersive worlds threaten human rights, with a particularly significant effect on different cultural groups and individuals.
  - The need for inclusive design and diverse representation in AI systems development is emphasized to counteract discrimination actively.

- **Digital Identity and Data Dignity:**
  - The discussion highlighted concerns about risks associated with digital identity, including human autonomy, neurotech capabilities, and security risks.
  - The importance of cybersecurity, encryption, and informed consent in the context of neurotechnology is paramount, as well as the need for awareness and education for humans and societies about the varied impacts of these technologies.

- **Human Rights in AI Development:**
  - Ensuring user autonomy over digital identity is a complex issue, requiring user-friendly controls to be integrated into platforms.
  - Establishing effective data ownership rights is seen as something beyond reliance on international data protection regulations, but also providing education and programmatic mechanisms, which would be crucial to the success of the regulatory frameworks.

- **Human Rights Concerns in Neurotechnology within the Metaverse:**[8]
  - Right to Personhood and Selfhood: Risks of personality or ability changes due to neuromodulation.
  - Freedom from Discrimination: Potential misuse of neuro mapping in employment and healthcare.
  - Right to Mental Privacy: Protecting thoughts and biometric data from neuro wearable devices.
  - Right to Interactive Agency: Safeguarding autonomy against neurotechnology manipulation.
  - Freedom from Persecution: Protection against persecution for thoughts or beliefs revealed by neuro technologies.
  - Right to Harm Protection: Preventing harm or manipulation through direct brain interaction.
  - Right Against Self-Incrimination: Ensuring neuro mapping is not used for self-incrimination in legal settings.

In summary, the risks and challenges encompass privacy, representation, equity, digital literacy, biases in AI, and the need for comprehensive strategies to ensure the responsible development and deployment of immersive technologies in the Metaverse.

8. see Footnote 7

SHAPING A RESPONSIBLE DIGITAL FUTURE WITH HUMAN RIGHTS AT THE HEART OF THE METAVERSE

In this section, we explore significant transformative aspects within AI-augmented realities. Key recommendations revolve around promoting diverse representation in technology, advocating for inclusive design principles, and emphasizing the need for digital literacy. Establishing ethical guardrails and public education on technological risks are crucial steps toward responsible innovation. Additionally, the development and enforcement of legal and ethical frameworks are highlighted, focusing on protecting digital identity, ensuring data ownership, and preserving mental autonomy. These themes collectively demonstrate a commitment to creating an immersive world that is safe, equitable, and respectful of human rights. The topics discussed are significant and outlined within three pillars:

- **Empowering Diverse Representation:**
  - **Emphasizing the Need for Diversity:** The discussions highlighted the critical impact of diverse representation in technology development. Acknowledging the need for individuals from various backgrounds and lived experiences who can contribute to more innovative and inclusive solutions.
  - **Advocacy for Diversity Requirements:** The emphasis on advocating diversity requirements underscores the importance of proactive measures to ensure diverse participation in technology teams, fostering a richer and more varied perspective.
  - **Promoting Inclusive Design Principles:** The call for promoting inclusive design principles reflects a commitment to creating technologies that cater to a broad range of users, considering diverse needs, and avoiding biases in the development process.

- **Establishing Guardrails and Enabling Education:**
  - **Establishing Guardrails:** The discussions recognized the need for guardrails in technology development, indicating a proactive approach to set ethical and safety standards. This suggests a commitment to preventing potential harm and ensuring responsible innovation.
  - **Public Education on Technology Risks:** The acknowledgment of opportunities in ensuring public education signifies a commitment to transparency. Educating the public on technology risks fosters awareness and empowers individuals to make informed decisions about their engagement with immersive technologies.
  - **Advocating for Digital Literacy Programs:** Recognizing the need for comprehensive digital literacy programs indicates a commitment to equipping individuals with the skills and knowledge to navigate the digital landscape responsibly. This aligns with minimizing risks associated with emerging technologies.

- **Broadening the Impact of Legal and Ethical Frameworks:**
  - **Development and Enforcement:** The discussions underscore the importance of actively developing and enforcing legal and ethical frameworks. This impact vector suggests a commitment to creating a regulatory environment that safeguards digital identity, data ownership, and mental autonomy within immersive worlds.
  - **Protection of Digital Identity:** The focus on legal and ethical frameworks highlights the intention to protect digital identity. This includes measures to prevent unauthorized use or misuse of personal information in immersive technologies.
  - **Safeguarding Data Ownership:** The emphasis on legal and ethical frameworks for data ownership indicates a commitment to defining and enforcing rules that ensure individuals have control over their data, preventing unauthorized access or exploitation.
  - **Preserving Mental Autonomy:** The discussion recognizes the need to protect mental autonomy within immersive worlds. This involves developing ethical guidelines and legal safeguards to prevent undue influence or manipulation through advanced technologies.

The discussions within this segment have illuminated a path forward, marked by the fusion of innovation and ethical consciousness in AI-augmented realities. By advocating for diverse representation and inclusive design, we pave the way for technologies that resonate with a broader spectrum of humanity and enrich the digital experience for all.

STRATEGIC RECOMMENDATIONS FOR HUMAN-CENTRIC DIGITAL INNOVATION

Drawing from the insights of the Human Rights Roundtable, this section outlines key measures and recommendations focused on integrating human rights, privacy, and digital dignity into AI, neurotechnologies, and immersive environments. These strategies, developed through collaborative discussions, highlight the importance of ethical development, inclusive design, and robust regulatory frameworks. They aim to ensure that the advancement of these technologies aligns with the protection and promotion of fundamental human values, shaping a future where innovation and human rights coexist harmoniously.

- **Regulatory Development for AI and Neurotech:**
  - Policto minimizations that ensure responsible AI and neurotechnology deployment, protecting human rights and digital dignity.
  - Ongoing efforts are required to establish safeguards against the unintended impacts of these technologies, with an emphasis on privacy and mental autonomy.

- **Ethical Considerations in Development and Design:**
  - Developers must prioritize ethical considerations, addressing biases in AI algorithms and ensuring accessibility for all, including those with disabilities.
  - Regular audits and diverse hiring practices are vital for inclusive technology development.

- **Inclusive AI Systems and Immersive Environments:**
  - Organizations should focus on creating AI systems and immersive environments that represent diverse cultural backgrounds.
  - Active participation in policy discussions about AI-powered immersive worlds can promote standards that support diversity and human rights.

- **Awareness and Research Investment:**
  - Raising awareness about human rights, privacy, and digital dignity in emerging technologies is crucial.
  - Investing in research and development to create safeguards against potential AI and neurotechnology harms is necessary.

- **Specific Measures for Human Rights in the Metaverse:**
  - Prioritizing human rights in AI development, including conducting impact assessments and ensuring compliance with human rights guidelines.
  - Protecting privacy in neurotechnology through regular audits and reducing privacy-related incidents.
  - Promoting inclusive design principles in AI and immersive technologies, including diverse user testing and inclusive design training sessions.
  - Ensuring accessibility standards in VR and AR products, focusing on features that meet the needs of all users.
  - Conducting regular bias audits in AI systems to mitigate identified biases.
  - Diversify technology development teams and increase hiring from underrepresented groups.
  - Focusing on digital dignity through policy implementation and user satisfaction.
  - Engaging in policy and regulation discussions, contributing to regulatory discourse on AI and immersive tech.
  - Collaborating with community groups and incorporating community feedback into product development.
  - Supporting initiatives focusing on justice and equality, measuring their impact on marginalized communities.

The recommendations and measures discussed emphasize a human rights-centered, inclusive, and ethical approach to developing and deploying AI, neuro technologies, and immersive environments. They aim to address these technologies' challenges and potential risks while ensuring they benefit and protect all individuals' rights. The emphasis on collaboration, policy engagement, and community involvement highlights the need for a collective approach to shaping a digital future that upholds human dignity and fundamental 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 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 AHRC

The Australian Human Rights Commission (AHRC) is Australia's National Human Rights Institution (NHRI), which plays a crucial role in promoting and monitoring the effective implementation of international human rights standards at the national level and investigates complaints about discrimination and human rights breaches in Australia. It was established in 1986 as the Human Rights and Equal Opportunity Commission (HREOC) and renamed in 2008. The AHRC provides resources, reports, and projects on various human rights issues, such as ageism, disability, race, sex, asylum seekers, and LGBTIQ+ rights. It is Australia's National Human Rights Institution.

Australian Human Rights Commission

The AHRC is committed to promoting and protecting human rights in Australia. It works towards achieving this goal by investigating and conciliating discrimination and human rights complaints, advocating for human rights to be considered in laws and policy making, and providing guidance to the government on promoting and protecting human rights. The complaints process is free and confidential, allowing individuals to resolve disputes quickly and effectively.

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
Human right that needs most strengthening to ensure mental autonomy **Answer: Right to Mental Privacy**	**56%**
Most crucial right for digital ID management in AI immersive worlds **Answer: Right to Control Over Digital Footprint**	**69%**
Most crucial for upholding rights in AI-powered immersive worlds? **Answer: Integrating Human Rights in Tech Development**	**39%**
Primary strategy for establishing effective data ownership rights? **Answer: International Data Protection Regulations**	**39%**
Most important factor in maintaining user autonomy over digital identity **Answer: BRAIN FREEZE!**	**0%**
What is the best method to achieve equitable access in AI-driven virtual worlds? **Answer: Global Digital Literacy Programs**	**73%**
Most important to prevent human rights violations by AI in immersive worlds **Answer: Transparent AI Algorithm Policies**	**46%**
Most important factor to minimize AI discrimination in VR environments **Answer: Ethical AI Development Training**	**53%**
Most effective design principle for inclusivity, AI-powered Metaverse **Answer: BRAIN FREEZE!**	**0%**
What is the best approach to integrate diverse cultural perspectives in AI? **Answer: Diverse AI Development Teams**	**70%**

LEAD AUTHORS

Monika Manolova
Advisor, XRSI
Chair - International Human Rights Day, MSW2023

Kavya Pearlman
CEO & Founder, XRSI
MSW Strategic Lead

Reginé Gilbert
Designer & Educator
NYU Tandon School of Engineering

PROGRAMMING COMMITTEE MEMBERS: INTERNATIONAL HUMAN RIGHTS DAY

Monika Manolova
XRSI Europe

Patrick Hooton
Australian Human Rights Commission

Kavya Pearlman
X Reality Safety Intelligence (XRSI)

April Boyd-Noronha
XRSI

Bhanujeet Choudhary
X Reality Safety Intelligence (XRSI)

Reginé Gilbert
NYU Tandon School of Engineering

Janaé N. McPhaul
XRSI/CyberXR Coalition

Dylan Fox
XR Access

ACKNOWLEDGEMENTS

Francesca Rowlinson
eSafety Commissioner

Oscar Cartagena
Augmented Experiences

Katherine Sessions
eSafety Commissioner

Patrick Hooton
Austrlian Human Rights Commission

Regine Gilbert
NYU

Troy Latter
Tech 4 Humanity

Emmie Hine
University of Bologna

Andres Leon-Geyer
LabXR PUCP

Ren Tyler
XR Access

Ingrid Vasiliu-Feltes
Institute SEI

Jim St.Clair
MyLigo

Hassan Abou Seada
Tadabour Space Innovation

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 International Human Rights Day Roundtable held during Metaverse Safety Week 2023, coinciding with the 75th Human Rights Day on December 10th, 2023. The information contained herein is intended solely for informational purposes.

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