conectys
global BPO solutions

INVENTING NEW TRUST & SAFETY APPROACHES FOR THE METAVERSE

As the metaverse continues to expand and evolve, it is critical to build robust Trust & Safety strategies to protect your brand and customers.

A Conectys made white paper

Contents

Click on a chapter title to navigate

Intro

This paper examines the role and importance of Trust & Safety (T&S) within the emerging metaverse landscape. It discusses integrating and balancing concepts like content moderation, freedom of expression and data protection to establish a robust virtual reality T&S framework that does not limit human voice and creativity, but also protects brands and users.

Although embryonic, the metaverse is growing rapidly and is set to become significant in all aspects of our daily lives. From a business perspective, the metaverse is increasingly attractive to companies and brands for engaging with digitally conversant customers, expanding their reach, generating new revenue, and fostering innovation by developing a more personalised customer experience.

However, embracing virtual poses critical challenges for organisations, especially in the metaverse, which allows unregulated interaction without filters while enabling individuals and companies to share information and content with little or unsatisfactory controls. This, of course, allows for great freedom but also leaves people, brands and workers vulnerable to various forms of miscommunication, toxicity, harassment and abuse, fraud and cyber-attack, which can damage individual and company reputations.

Implementing preventive and strategic Trust & Safety measures are essential to creating and developing healthy metaverse communities. It allows businesses to mitigate the risk of security breaches that can cause more harm than the traditional online platforms used in the last decades.

The role of Trust & Safety in metaverse protection

For however long communities have existed, T&S has undoubtedly been present in some form or other. The term "Trust & Safety", as it is commonly used today, refers specifically to the efforts made by companies and online platforms' operators to build secure and friendly virtual environments.

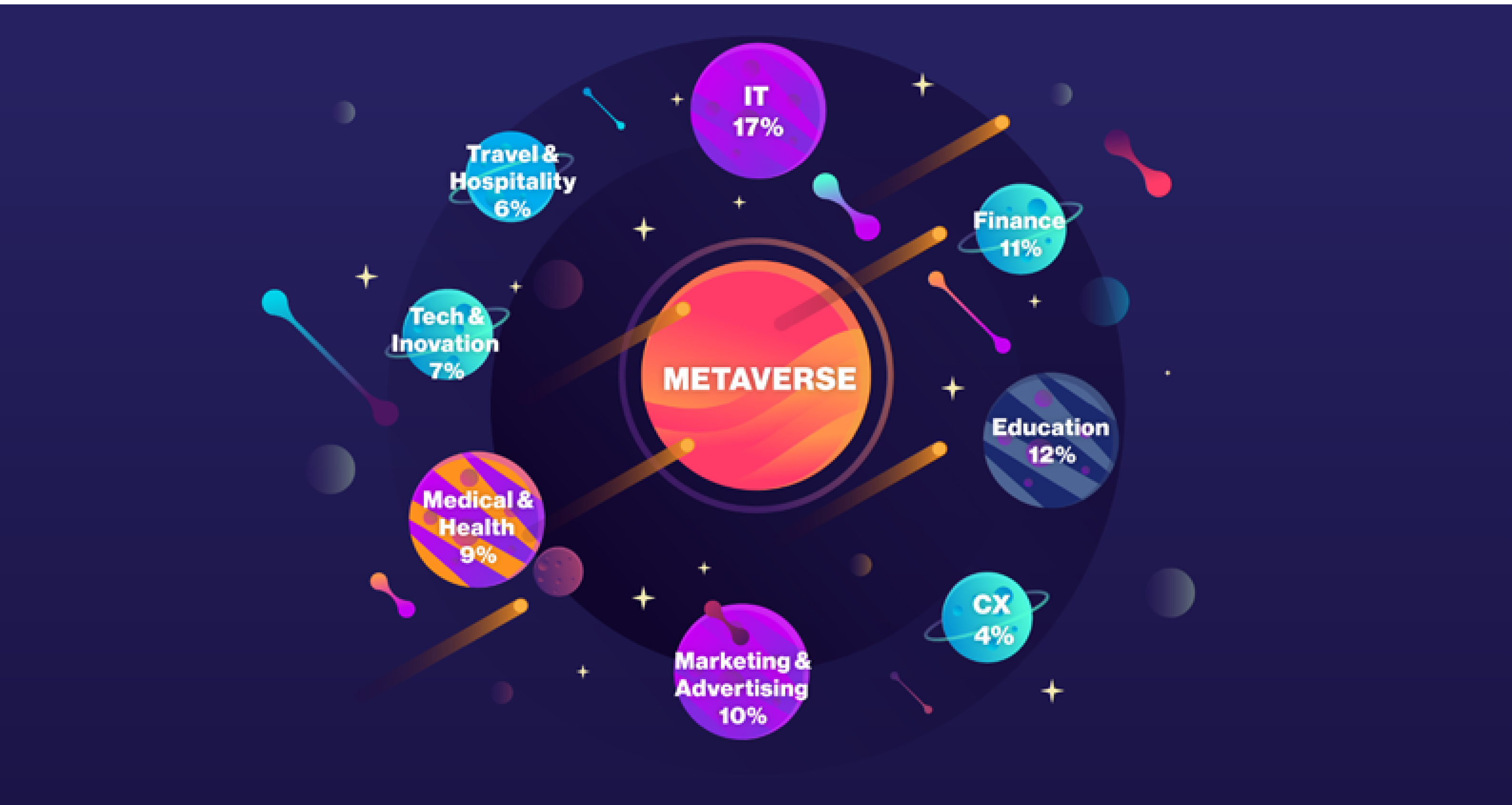
Trust & Safety typically includes a set of practices, processes and policies that can be developed and implemented to ensure the lawful use of digital platforms and make online communities welcoming. In the metaverse, where users interact in immersive and often anonymous digital spaces on an unprecedented scale, it is essential to establish protocols and measures to prevent malicious activities and protect visitors from abuses.

An appropriately designed Trust & Safety strategy can also enhance user experience, build trust and credibility, and contribute to the growth of the metaverse ecosystem. To fully grasp the significance of Trust & Safety in the metaverse, it is essential to understand the exceptional nature of virtual reality alongside its current developments and potential future advancements.

From Science Fiction to Reality

The term "metaverse" comes from the science fiction novel Snow Crash written by Neal Stephenson and published in 1992. It combines the words "meta" and "universe" to mean "beyond the world", or what is often referred to as a parallel universe. Since the release of Snow Crash, and with the advent of technology such as the internet, smartphones, social media, augmented reality and more, the potential for exploring the metaverse has grown exponentially. As per Statista1, by the end of 2021, the metaverse market size reached $38.5 billion. It was expected to exceed $47 billion in 2022, while projections for 2030 estimate that the figure will reach $678.8 billion.

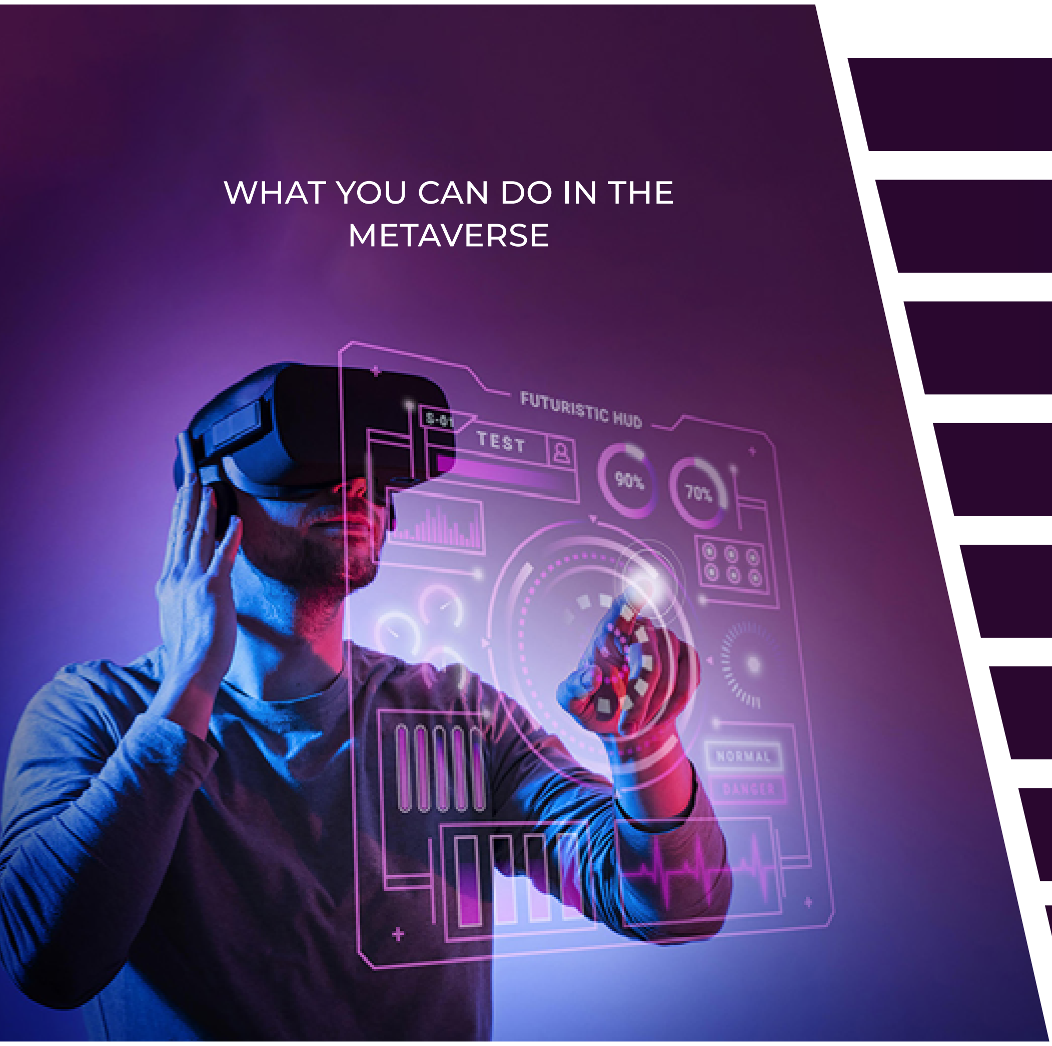
With no physical and geographical limitations, the metaverse opens the door to substantial business growth and unlimited opportunities across various spheres. Virtual, immersive realities present opportunities for individuals, communities, businesses, and industries to interact socially, entertain, educate, advertise, sell and much more.



DEFINITIONS OF THE METAVERSE

There are many definitions of what the metaverse is and how it works. We have selected a few to try and encapsulate an easily digestible overview of the virtual environment:

- According to the Cambridge University Dictionary, "The metaverse is a simulated reality where individuals represented by avatars interact with one another in a 3D environment that closely resembles the real world"2.
- Following Ed Greig, Chief Disruptor at Deloitte, "The metaverse is the internet, but in 3D, and a form of digital interaction where connected, virtual experiences can either simulate the real world or imagine worlds beyond it"3.
- As per Statista, "The metaverse is where the physical and digital worlds come together. As an evolution of social technologies, the metaverse allows digital representations of people and avatars to interact with each other in different ways. No matter where (at work, in an office etc.) or what you do virtually (go to concerts or sports events, or even try on clothes etc.), the metaverse provides a space for endless, interconnected virtual communities. All is done with the use of virtual reality (VR) headsets, augmented reality (AR) glasses, smartphone apps, or other devices"4.



WHAT YOU CAN DO IN THE METAVERSE

- TRAVELING & EXPLORING
- PARTICIPATING IN TRANING & SIMULATION
- VISITING EVENTS & EXHIBITIONS
- SOCIAL NETWORKING
- SHOPPING & TRADING
- GAMING & ENTERTAINING
- HAVING MEDCAL CONSULTANCY
- ATTENDING MEETING & CONFERENCES

Why does the metaverse need protection?

While being a digital space where people interact with each other, share content and engage in commercial or social activities, the metaverse is also endangered by different forms of cybercrime. Without effective T&S measures, the metaverse is a breeding ground for abuse, potentially harming users, brands, and overall communities. These include, for instance, personal information and intellectual property theft, financial fraud, hate speech, assault, and many other forms of online harassment.

These unique digital environments will likely become the norm as internet access and technology expand to fuel metaverse growth. To brace themselves for an exciting but unknown future, brands, technology providers, and specialist service partners must prioritise the development of secure and safe experiences for users and customers. They will capitalise if they are to understand these dynamics and implement Trust & Safety strategies that place them ahead of the pack, protecting all parties involved and making the metaverse a positive phenomenon.

Quoting The New York Times article on the metaverse's dark side, you can find out that "bad behaviour in the metaverse can be more severe than today's online harassment and bullying. That's because virtual reality plunges people into an all-encompassing digital environment where unwanted touches in the digital world can be made to feel real, and the sensory experience is heightened. In one popular virtual reality game, VRChat, a violating incident occurs about once every seven minutes".

How can Trust & Safety ensure security in the metaverse?

Trust & Safety can ensure security in the metaverse through various measures, policies and guidelines for acceptable behaviour, monitoring for security breaches and malicious activity, as well as addressing any incidents that may arise. It can also engage the technology providers to integrate security features into the metaverse infrastructure and ensure that data and user information are protected through encryption and other means.

NEW TRUST & SAFETY APPROACH FOR THE METAVERSE

The metaverse offers an immersive experience that creates a seemingly natural environment. It is likely to represent much graver dangers than the internet as we know it. Therefore, it is necessary to develop a metaverse Trust & Safety strategy to deal with the specific security challenges of immersive virtual reality while simultaneously facing the dilemma of balancing safety and freedom of expression. These include large-scale, complex, borderless environments, user anonymity and identity theft, and a continuously evolving technology set. Below we note what primarily distinguishes the metaverse:

- Multiple dimensions and interactions: such as enhanced interactive physical movements and gestures in addition to text and voice, make tracking and monitoring activity difficult.
- Content lifespan or persistence: where user-generated content and interactions remain in the virtual space after the user logs out.
- Ambiguous or lack of legislation: a virtual world that knows no geography or physicality and where local laws and regulations may or may not apply leads to ambiguous interpretations of what is or is not allowable.
- Technological limitations: evolving technology used to build and operate the metaverse could contradict the effectiveness of safety measures.
- Interoperability: the ability for users to move freely between different metaverse environments and have their experiences and assets transferred seamlessly.
- Financial transactions: covering virtual goods, services, and experiences that can have real-world value and be bought and sold using various forms of currency.
- Consternation about how to balance: between allowing freedom of expression and enforcing rules to prevent harm.

POTENTIAL TRUST & SAFETY STRATEGIES IN THE METAVERSE

As the metaverse becomes uber-democratic, its communities will be shaped into a mosaic of users representing different cultures and individual codes of conduct. By implementing strong T&S measures, diverse groups can unite, aligning their cultural values and beliefs.

Trust & Safety can help to protect individuals from being easy prey to people with intentionally harmful actions that violate community guidelines. It may prevent clashes, bullying, abuse, cheating, and even grooming in environments where user identity is currently easily manipulatable and difficult to control.

This is why all-encompassing, specialised content moderation is fundamental for leading brands. However, creating a secure and desirable metaverse environment requires a holistic and multi-dimensional strategy.

"In 2022, industry consultants from the Everest Group shared a study exploring potential Trust & Safety challenges and recommended metaverse risk mitigation strategies to help tackle them. The report suggests T&S strategies for virtual avatar abuse around data privacy, the safety of virtual assets, content moderators' well-being, and regulatory ambiguities."

A comprehensive Trust & Safety Metaverse Strategy should include the following:

Multi-dimensional content moderation, which combines the human touch with AI-driven technology while allowing for reviewing and removing inappropriate user-generated content, harassing or violating established guidelines, laws, or community standards.

Clear & enforced community guidelines & terms of service to inform what is acceptable behaviour within the virtual environment and what activities are prohibited. Continuous monitoring and improvement of T&S policies.

Real-time monitoring available with the support of technology and human review to identify potential violations as quickly as possible and respond promptly by taking appropriate actions to address them.

User reporting & incident response systems covering the processes and tools put in place to create secure and tamper-proof records of the individual's activity in the metaverse and allow users to report potential violations and for T&S teams to respond to and address these incidents.

Cybersecurity measures to prevent hacking, theft, data breaches and other forms of malicious activity. This can include implementing encryption algorithms to protect data, regularly conducting security audits and vulnerability assessments, and using multi-factor authentication to establish the authenticity of users.

Virtual asset protection and fraud detection algorithms and systems to identify and prevent unauthorised transactions, regularly review user activity for suspicious behaviour and take steps to secure the storage and transfer of virtual assets.

Metaverse law enforcement & transparency to help create virtual police forces and judicial systems to address offences in the metaverse, as well as provide clear rules and consequences for violations.

Third-party engagement to prevent misinformation, which involves the “fact-checkers” who have expertise in specific areas or industries. And who can effectively evaluate the accuracy of the information and provide a context where needed.

Collaboration between internal and external entities to share information and intelligence across different platforms and jurisdictions, as well as involve individuals in developing and designing Trust & Safety systems.

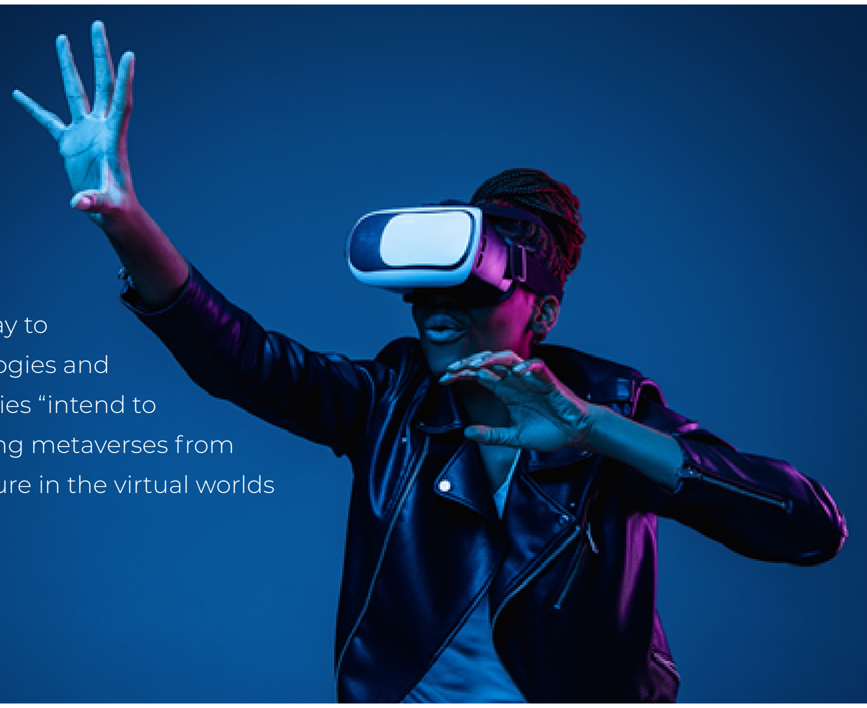
Education and awareness to provide training and resources for users. All to increase their knowledge of security and promote positive behaviours.

It is worth remembering that the metaverse is constantly evolving. This means continuously assimilating policy changes and implementing prevention and strategic measures into processes. What works today for the virtual space may not be effective later.

The European Union’s metaverse-focused regulation initiative

In the State of the Union Letter of Intent published in September 2022, the EU’s president, Ursula von der Leyen, confirmed the Commission would put forward an “Initiative on virtual worlds, such as the metaverse” in 2023. It will present several initiatives to clarify Europe’s rules and expectations for the metaverse.

Additionally, the EU Commissioner Thierry Breton announced that the “European Commission would continue looking at new digital opportunities and trends, such as the metaverse. The European way to foster the virtual worlds is threefold: people, technologies and infrastructure.” According to Breton, the EU authorities “intend to shape the development of genuinely safe and thriving metaverses from the outset. The one where people should feel as secure in the virtual worlds as they do in the real one”.



The initiative will impact current metaverse Trust & Safety strategies by setting standards and guidelines for responsible and secure virtual space operations in the future. This regulatory framework will likely include additional metaverse providers' responsibilities in ensuring they meet the requirements.

Content moderation, a major T&S challenge for the metaverse

Within the context of Trust & Safety, one of the most common forms of security governance in virtual worlds is content moderation. Its goal is to maintain integrity, quality, and safety across social media platforms, digital marketplaces, sharing economy, dating sites, communities and forums, and many more.

Content moderation can cover a wide range of content created by users, referred to as 'user-generated', be that written, audio or visual content. These may include hate speech and imagery, graphic or violent material, pornography, harassment and bullying, terrorist content, fake news and misinformation, spam and scam content, personal information and privacy violations or intellectual property violations. The specific types of content that need moderating also depends on the platform's terms of service and the jurisdiction under which it operates.

Types of Content Moderation: Pre-publication and Post-publication

There are typically two types of content moderation: proactive and reactive. Both types play essential roles in maintaining a safe and respectful online community.

The first one focuses on preventing harmful or inappropriate content from being published in the first place. It can be identified through filters, algorithms, and human moderators, who pre-screen content before it becomes publicly available.

Moderating post-published content involves responding to user reports and removing them. This type of moderation is usually necessary when proactive measures are not enough to prevent harmful content from reaching the public. In terms of reactive moderation, it can be achieved through user reporting systems, community guidelines, and service agreement terms.

A perfect combination: of human touch and technology

Metaverse content moderation can be undertaken through human, automated, and AI-driven methods. Effective moderation involves balancing humans and technology, as each offers unique benefits. Human moderators provide the personal touch and context-based decision-making, while technology offers speed via automation, scalability, and real-time monitoring. Community-based moderation allows for empathy and understanding in sensitive situations, while AI-based systems quickly scan and identify potential violations. Technology also provides transparency through tamper-proof records of user activity.

> "Creating a secure metaverse experience requires combining human content moderation expertise, a deep understanding of cultural and regulatory frameworks, and cutting-edge technology. By bringing these critical elements together, businesses can establish a robust Trust & Safety operation that enhances user experiences and ensure their safety in the virtual world."
>
> - Iulian Bacain, Sales and Marketing VP at Conectys

A famous technology and human rights expert, Brittan Heller, is going even further. According to her, VR-virtual reality and AR-augmented reality platforms need specific terms of service for immersive environments, based on how this technology interacts with our brains. We cannot simply apply rules from existing social media to the metaverse. This is of the utmost importance, as platform governance in digital worlds will regulate behaviour in addition to content.

Personal data and privacy protection in the metaverse

Personal data protection and security should always be a primary concern for any company, organisation or community operating in the virtual space. The metaverse offers a unique environment where data can be collected and processed on a large scale. It is, therefore, essential to ensure that adequate measures are in place to prevent unauthorised access, misuse, or theft of personal data. The data may include, for instance, information referring to demography, contact, behaviour, user-generated content, social media, and biometric or virtual asset data.

To guarantee that users' rights are respected, entities must be transparent about data processing practices. They should treat all individuals equally, limit data collection to what is strictly necessary, avoid manipulative or invasive procedures in data-driven profiling, and always ask for explicit consent to collect, store and process personal data.

Users should be provided with information about the type and use of the collected data and be allowed to access, modify, and delete it if required.

It is essential for regular reviews and updates to privacy policies to be carried out to remain compliant with changing regulations. Taking guidance from legal and privacy experts to stay informed of changes in data protection regulations is also highly recommended.

Data protection vs outsourcing

When delegating the services to a third party, like a Business Process Outsourcing (BPO) partner, it is essential to ensure complete alignment with your best practices. Alongside this, relevant laws and regulations regarding data protection and privacy, such as the General Data Protection Regulation (GDPR), should be included. The T&S provider must also secure adequate security measures to protect customer data from unauthorised access or breaches. This includes implementing robust security protocols and regularly testing and updating them to stay ahead of potential threats.

Advocating for improved privacy in the metaverse

It is worth emphasising that the importance of maintaining a secure and law-abiding virtual environment is regularly raised as a concern by all involved parties. In the metaverse, various stakeholders may increasingly call for such protection as creators and administrators, users and community members, government agencies, investors, civil rights and advocacy groups.

One example is the work created by organisations such as Access Now and Electronic Frontier Foundation (EFF), which fights for human rights protection in the digital world. Both organisations called for governments and other stakeholders to "address human rights in virtual and augmented reality and ensure that these rights are respected and enforced". They claimed that "like any other technology, the metaverse can have many positive impacts on our daily lives, but also pose substantial risks to human rights, exacerbating already severe intrusions on our private sphere."

The question then arises, do human rights exist in a virtual world? KPMG Global Head of Infrastructure provides an insightful answer: *"Of course, they do. Behind every avatar is a human. And they deserve the same rights in a virtual world as they are afforded in the physical world. A human operating an avatar has the right to feel safe and secure. They have the right to be free from discrimination. The applicability of human rights should not depend on the channel through which a human interacts with the world"*.

Privacy vs freedom of expression

Equally important as privacy protection is freedom of expression, a fundamental right to be protected in the metaverse, allowing users to express themselves openly without fear of censorship or retribution.

To strike a balance between privacy protection and freedom of expression, metaverse platforms and operators must implement appropriate data protection measures, establish clear and enforceable policies, including guidelines for acceptable speech and content, and provide mechanisms for reporting and addressing abusive or harmful behaviour.

By combining privacy protection, freedom of expression, and compelling content moderation, the metaverse can become a thriving and inclusive community where all users feel safe and respected.

How to protect children in the metaverse?

Protecting children in the metaverse is evident and necessary, as inappropriate content, cyberbullying, and grooming are unfortunately still prevalent. Metaverse operators must prepare appropriately for younger users, who are not always aware of the threat.

According to Common Sense, a leading non-profit organisation working to improve the digital world for kids and families, “the use of the metaverse presents various advantages for kids and teenagers, including opportunities for imaginative play and engaging education. However, the potential risks associated with its usage are not yet fully understood”. Common Sense explored the existing risks, among all you can find physiological dangers, which may cause “cybersickness” among kids.

One of the apparent measures here is an age verification system that restricts access to content via stronger privacy measures that protect children’s personal information and content moderation, enabling prompt removal of harmful or inappropriate material. Equally critical is promoting educational activities that increase the awareness of children and families about the dangers of the metaverse and how to stay safe online.

Finally, parents must be encouraged to take control and watch over children’s presence in the metaverse, setting restrictions on the amount of time children spend in the metaverse and the types of content they access.

Future of T&S outsourcing vendors in the metaverse ecosystem

The need for integrated Trust & Safety will inevitably arise as the metaverse grows and evolves. The industry faces considerable challenges and great opportunities to define a new generation of T&S services. This is where specialised BPO, and end-to-end Trust & Safety specialists, such as Conectys, make a difference.

Companies entering the metaverse will expect a range of integrated outsourcing services, such as content moderation, customer experience, automation, monitoring, and AI-driven solutions, in order to create a secure and trustworthy meta-environment.

The Everest Group report states: “Currently, T&S services are among the fastest-growing segment of the Business Process Services market, expected to grow 35-38% through 2024 and reach US$15-20 billion. Mitigating threats to user security, increased abuse, the proliferation of objectionable content and financial fraud represents an ample opportunity for the third-party T&S services market in the coming years. The market is expected to accelerate to 60-68% growth beyond 2024 as technology and infrastructure advance beyond the nascent stage.

By outsourcing Trust & Safety services, metaverse operators can concentrate on their core activities. They can free up resources, reduce costs, gain specialised skills and assets, and leverage expertise and innovation from expert partners. Entirely in touch with technological and regulatory developments in the metaverse and other digital realms, knowledgeable BPO players are equipped to supply robust, dependable, and trustworthy Trust & Safety solutions.

SUMMARY

The growing volume and diversity of user-generated content on digital platforms, the emergence of the metaverse and the legislation gap represent colossal Trust & Safety (T&S) challenges for businesses.

If your organisation plans to enter the metaverse, it is essential to think strategically and to be proactive about implementing preventive Trust & Safety measures tailored for the virtual world. This will dictate if you deliver the type of safety needed to prosper in the metaverse. Even if you are still getting ready, you must ensure that your team is up to date with the latest developments, including new regulations and best practices.

We have a few suggestions to help in preparing your company for upcoming changes in Trust & Safety:

1. Review your policies and procedures and align them with your strategy for the metaverse.
2. Understand internal stakeholders' and external customers' brand safety expectations for the metaverse.
3. Prepare an investment plan, including technology, tools and services that may help you monitor and manage your brand safety in the metaverse.
4. Provide training to your team on the new standards for Trust & Safety in the metaverse.
5. Seek the assistance of specialised partners that can help to facilitate your success.

REFERENCES

- Statista report "Metaverse market revenue worldwide from 2021 to 2030".
- Online Cambridge University Dictionary.
- Deloitte NSE LLP "Ed on the metaverse. Can it bring people closer than ever?".
- Statista insight, published by J. Clement, Research lead covering internet and gaming.
- The New York Times, "The Metaverse's Dark Side: Here Come Harassment and Assaults" by Sheera Frenkel and Kellen Browning.
- Everest Group report "Taming the Hydra: Trust and Safety (T&S) in the Metaverse".
- LinkedIn announcement published by the EU Commissioner Thierry Breton "People, technologies & infrastructure – Europe's plan to thrive in the metaverse" September 2022.
- Based on the World Economic Forum Insights "How to address digital safety in the met".
- An article by Access Now and the Electronic Frontier Foundation (EFF) "Virtual worlds, real people: human rights in the metaverse".
- The KPMG blog article "Protecting human rights in the metaverse" by Richard Threlfall,18 January 2023.
- Common Sense insights on "What Are Kids Doing in the Metaverse?" and the new report "The Common Sense Census: Media Use by Tweens and Teens, 2021.
- Everest Group press release "Growth of Metaverse Increases Trust and Safety (T&S) Risks to Enterprises, Users; Implications for the Third-Party T&S Services Market, September 2022.

A few words about Conectys

Conectys is a digital-first, Customer Experience and Trust & Safety specialist firm that delivers cost efficiencies and increases the speed of implementation for companies facing challenges like hypergrowth, market disruption and globalisation. Unlike traditional punch-in, punch-out service providers, Conectys co-creates flexible, strategic and digitally inclusive approaches that overcome our clients' extraordinary challenges.

Key Conectys' services are:

Trust and Safety: hybrid content moderation across online platforms, social media, gaming, and the metaverse according to brand and social safety guidelines and regulatory requirements.

Multilingual Customer Experience: 24/7/365 customer service and tech support across 35+ languages.

Digital Transformation services: RPA, NLP, Chatbots, automatic translation chatbots, sentiment analysis, voice-to-text, etc.

BlueTweak– a BPaaS agnostic platform that powers your CRM.

Globally disseminated in top BPO locations and with 18+ years in the industry, Conectys is large enough to be a safe partner and experienced enough to be a specialist yet of the right size to be flexible, dynamic and entrepreneurial.

Contact Us

Would love to connect with you to discuss how our team at Conectys can help enhance your Trust & Safety and Customer Experience services.

LET'S TALK

sales@conectys.com

info@conectys.com

rfp@conectys.com

☎ US - 1469 532 0215

UK - 44 203 318 1593

EU - 32 929 8011

HK - 852 800 930 130