OPEN LETTER

Considerations on user identity within metaverse environments

[version 1; peer review: 2 approved with reservations]

Kata Szita[1], Lauren Buck[2], Nicola Palladino[3], Qian Xiao[4], Pat Treusch[5], Dalila Burin[6], Jennifer O'Meara[7], Vincent Wade[8]

[1]School of Communications, Insight Research Ireland Centre for Data Analytics, Henry Grattan Building, Dublin City University, Dublin, Ireland
[2]Division of Games, The University of Utah, 332 South 1400 East, bldg. 72 Salt Lake City, Utah, UT 84112, USA
[3]Department of Management & Innovation Systems, University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano SA, Italy
[4]Maynooth International Engineering College, Computer Science, Maynooth University, Maynooth, Ireland
[5]Department of Design, Media and Educational Science, University of Southern Denmark, Campusvej 55, Odense, 5230, Denmark
[6]School of Computer Science and Statistics, O'Reilly Institute, Trinity College Dublin, Dublin, Ireland
[7]School of Creative Arts, Trinity College Dublin, Dublin, Ireland
[8]School of Computer Science and Statistics, O'Reilly Institute, Trinity College Dublin, Dublin, Ireland

v1 **First published:** 16 Jun 2025, **5**:162
https://doi.org/10.12688/openreseurope.20411.1
Latest published: 16 Jun 2025, **5**:162
https://doi.org/10.12688/openreseurope.20411.1

Abstract

The metaverse concept presents an immersive three-dimensional space for interpersonal connections, where people can socialize, learn, do business, and complete other activities. It is a digital system with its own economy, technological properties, and sensory and behavioral domains. While discourses often focus on the technological and economic feasibility of the metaverse, less is said about the implications for human identity. Identity in the metaverse is an amalgam of self-representation, branding, and behaviors, but is also dependent on technological features. This paper analyzes user identity in terms of behaviors and personal data collection and possible misuse. As such, it highlights technological, ethical, and psychological dilemmas and potential solutions before the realization of the metaverse or similar interoperable virtual networks. Specifically, we discuss questions regarding the representation of human identities, the collection and reuse of personal data, and the use of AI models for customizing user experiences. Based on our assessment of these, we propose a legal and ethical foundation for users and developers of the metaverse. Rather than averting future developments in technologies and use practices, our objective is to highlight elements where the protection of users and their experiences requires particular attention.

Open Peer Review

Approval Status ? ?

	1	2
version 1 16 Jun 2025	? view	? view

1. **Stavros Kaperonis**, Panteion University of Social and Political Sciences, Athens, Greece

2. **Dimitris Kogias**, University of West Attica, Egaleo, Greece

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
Metaverse, personal data, artificial intelligence, user experience, identity, AI ethics, AI legislation

This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.

This article is included in the Horizon 2020 gateway.

Corresponding author: Kata Szita (kata.szita@dcu.ie)

Author roles: Szita K: Conceptualization, Formal Analysis, Investigation, Resources, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Buck L**: Conceptualization, Formal Analysis, Investigation, Resources, Writing – Original Draft Preparation; **Palladino N**: Conceptualization, Formal Analysis, Investigation, Resources, Writing – Original Draft Preparation; **Xiao Q**: Conceptualization, Formal Analysis, Investigation, Resources, Writing – Original Draft Preparation; **Treusch P**: Conceptualization, Formal Analysis; **Burin D**: Formal Analysis, Investigation, Resources, Writing – Original Draft Preparation; **O'Meara J**: Conceptualization; **Wade V**: Conceptualization, Funding Acquisition, Investigation, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the HUMAN+ COFUND Marie Skłodowska-Curie grant agreement No. 945447.
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

How to cite this article: Szita K, Buck L, Palladino N *et al.* **Considerations on user identity within metaverse environments [version 1; peer review: 2 approved with reservations]** Open Research Europe 2025, **5**:162 https://doi.org/10.12688/openreseurope.20411.1

First published: 16 Jun 2025, **5**:162 https://doi.org/10.12688/openreseurope.20411.1

Introduction

The metaverse refers to the concept of a virtual multiuser environment that is accessible across platforms and allows for seamless communication via an iterative form of the internet. It operates based on the idea of a unique, persistent identity for all users. Currently, there is no scientific consensus on the exact form the metaverse will take (in case it will be fully or partially realized[1]). Nevertheless, previous research approaches it as a three-dimensional world where users interact through avatars: a sort of "walkable version of the Internet" (Ritterbusch & Teichmann, 2023, p. 12368) or, a system of permeable virtual worlds that resemble present-day social virtual reality platforms. As the review of definitions by Ritterbusch and Teichmann (2023) suggests, the metaverse is a digital system with its own economy, physical properties, and sensory and behavioral domains. Correspondingly, Matthew Ball (2022) formally defines the metaverse as a massively scaled and interoperable network of real-time rendered 3D virtual worlds that can be experienced synchronously and persistently by an effectively unlimited number of users with an individual sense of presence, and with continuity of data, such as identity, history, entitlements, objects, communications, and payments (p. 29).

A critical aspect to ponder is the impact of these digital systems on user identity. Identity in the metaverse is an amalgam of self-representation, branding, and behaviors, but is also dependent on technological features or affordances. Thus, we argue that digital identity in metaverse environments is defined by two dimensions—a *personal* and a *data-specific* dimension—that mark its novelty and call for further analysis (see Figure 1). On the one hand, identity in the metaverse combines modes of communication, technological affordances (representations or interaction mechanisms), and human factors (social or legal identities). On the other, it pertains to personal data collected during use, which includes personal identifiers (e.g., name, date of birth, address), demographic markers (e.g., age, ethnicity, gender), and physiological elements (e.g., biosignals, body features, voice). In this paper, we outline the representational, technological, ethical, and legal considerations of human identity in the metaverse. By this, we aim to provide grounds for understanding the human implications and the related risks of this prospective system—including aspects of behavior, identification, surveillance, and sensory rendering of virtual social networks. Based on our assessment of these risks, we propose a legal and ethical foundation for users and developers of the metaverse. Rather than averting future developments in technologies and use practices, our objective is to highlight elements where the protection of users and their experiences requires particular attention. We find this crucial in the context of a platform that promises to provide immersive experiences while encompassing a wide range of aspects of users' lives.

Research questions

Applications that fit the metaverse concept are anticipated to have a profound impact on online social systems—more so than social media and other digital environments. One unique trait of the metaverse is that the digital identity of a user will not only persist on current digital platforms (e.g., mobile devices and desktop computers), but will be integrated across extended reality platforms (XR, that includes virtual reality (VR), augmented reality (AR), and mixed reality (MR)). Extended reality platforms facilitate immersive experiences that can be deeply impactful on human behaviors and interactions (see for example, Freeman & Acena, 2021; van Brakel *et al.*, 2023). Thus, the immersive qualities and prospects of the metaverse for offering functions for all aspects of life (work, leisure, education, socializing, etc.) will have emotional, social, and behavioral impacts on users. This is especially true if we consider the nature of XR systems where AR and MR combine real-world and digitally created stimuli and objects, and VR

[1] It is worth acknowledging that the metaverse is often considered as a business concept based on existing technological means provided by virtual reality and remote communication technologies. While it is a feasible concept that won many investors over, its realization has been obstructed by the lack of a uniform technological and user experience outline, shared purpose, and the general accessibility of global populations.

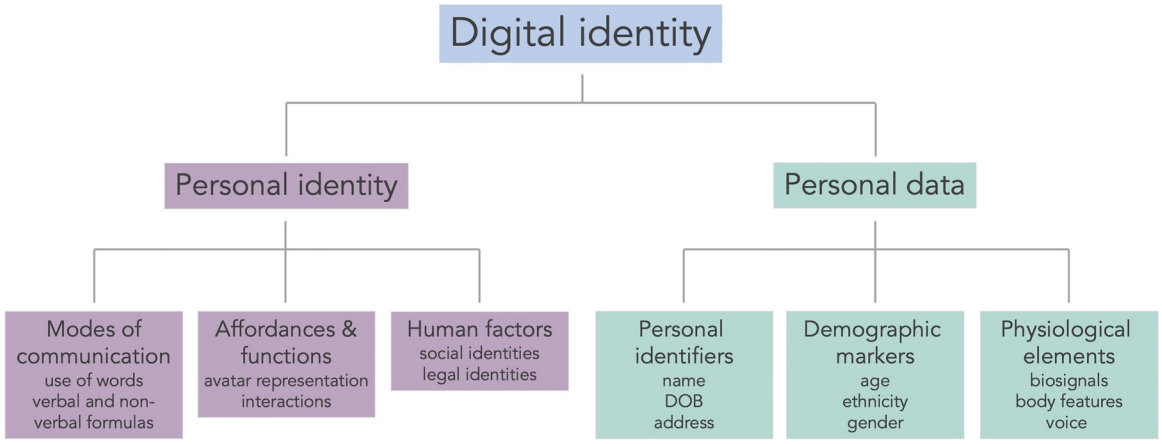

Figure 1. Digital identity in the metaverse.

provides a fully immersive simulation of environments. Based on these characteristics, XR raises the question of realism, anthropomorphism, and the effects thereof: namely, how much users accept XR experiences as part of reality and themselves and other users as unconditionally *real*.

As we detail below, digital identities created and used in the metaverse have peculiar links to real-world identities. While an avatar and other identifiers (e.g., name, voice) are connected to a single user, their digital twin (digital representation) may translate users' physical characteristics and behaviors according to the specificities or limitations of the platform. In addition, users may choose distinct self-representations. Irrespective of whether a user's avatar appears similar to or different from their real-life self, digital identities will define several aspects of life in the metaverse, including social interactions and communication, consumer behavior, interests, activities, and the ways in with commercial actors reach them, for example, in the form of targeted advertisements. This means that the identity a user creates in the metaverse is as relevant to their daily lives as their physical-world identity, which makes it a crucial subject of scholarship. We propose to address this issue by answering the following research questions.

1. How does XR technology, the continuous collection of personal data, and the use of generative AI impact the identity of a user?

2. What are the ethical and legal implications for user identity, and how does the widespread use of generative AI affect these?

Identity in virtual environments

Identity is a complex and multifaceted concept, encompassing how people perceive, define, and express themselves. Due to the proliferation of the internet, people's real-life selves became mirrored by and extrapolated into online selves, establishing a digital component of identity. Digital identity, particularly within a virtual environment like the metaverse, includes various aspects of one's virtual persona and how one interacts with others in these environments. It is comprised of many different components, including one's virtual bodily self-representation, social interactions, behavior and personality, cultural and subcultural associations, and so on.

In establishing a digital identity, one's virtual characteristics can be borne out of self-identification or one's choice to express oneself based on a set of experiences, knowledge, cultural background, and behaviors (Sparks & Shepherd, 1992). It can also emerge from one's social identity, which is adaptable and relative to one's assimilation with social groups chosen by or assigned to them (Abrams & Hogg, 2010; Tajfel *et al.*, 1971; Tajfel & Turner, 1986). Finally, digital identity can be linked directly to real-life bodily attributes. These elements of a user's identity are translated into digital personas that define the ways in which they are represented and perceived in virtual environments. A feature of the metaverse is that digital identities will be interoperable between multiple computing platforms, offering the opportunity for one's unique, defining features to be integrated into a single, distinct digital identity. In the following sections, we discuss the affordances of various XR systems and how they might contribute to and affect digital identities.

Bodily identity

Like the physical world, digital platforms offer sensory cues that transmit data about virtual environments which, in turn, allow their users to interpret and interact with those environments. The sensory cues available on current platforms are, however, generally limited to vision (sight), audition (hearing), and somatosensation (touch). Some consider visual cues to be the most important of the sensory cues (Hutmacher, 2019), and this sentiment has heavily influenced the development of traditional and emerging technologies like mobile, desktop, and XR devices. It is undeniable that the visual cues available on these platforms are rich and complex and are a substantial, primary contributor to a user's bodily identity.

Possessing a virtual bodily self-representation (an avatar) is a prerequisite for interacting and being interacted with in a virtual environment. Virtual bodies define the location of users in a space, provide them with a reference by which they can determine their own capabilities for actions (affordances), and give cues to others about one's actions and communication intents, to name but a few functions. The way users are represented can vary between a one-to-one replication of the user's physicality via a simulated 3D model that is constructed based on a photo or video recording to simple, cartoonish representation with a limited number of customization settings. Applications providing access to virtual environments often enable users to author, render, and realize the visual representation of their digital persona. This may include customizing their visible demographic, bodily characteristics, and looks. Such a feature challenges what is traditionally considered as identity, allowing users to reshape and recast their identities in many ways (Turkle, 1997).

When designing their avatars, users place importance on distinct characteristics and often express a diverse range of motivations for the design choices they make. Context and intrinsic desire often dictate the features attributed to a digital identity, and users often actualize idealized versions of themselves, create alter-egos, and push boundaries to explore identities that may normally be considered less socially desirable (Dengah & Snodgrass, 2020; Han *et al.*, 2023). These choices can be attributed to many factors, from the intrinsic desire to assimilate with a particular social group to the simplistic desire to try out and experiment with a new look (Freeman & Maloney, 2021; Szita, 2022).

Digital identities in online platforms evoke more than a superficial change in appearance. Users may adapt different behaviors to align with characteristics that they have conceptualized and assigned to their new virtual identity (O'Meara & Szita, 2021)—a phenomenon known as the proteus effect (Yee & Bailenson, 2007). XR systems enhance the proteus effect in that they enable users to embody their virtual

self-representation; move with them, act with them. This allows them to experience a profound sense of presence within a virtual environment and an embodied connection to their virtual selves. When a user embodies an avatar, they have the sense that the properties of their avatar's body are the properties of their own biological body (Burin *et al.*, 2019; Kilteni *et al.*, 2012). There are varying degrees to which users experience the sensations of embodiment and presence, but there is consensus that both can be experienced to a significant degree under particular conditions in XR environments (Genay *et al.*, 2022; Kilteni *et al.*, 2012). Nevertheless, the sensations of embodiment and presence offer a unique connection between the user, avatar, and virtual world. In the metaverse, one's digital identity will no longer solely exist as a profile within a screen-based desktop or mobile infrastructure, but will do as a simulated body that users can step into and use to interact with an environment that mimics certain properties of the physical world.

Risks associated with the visual bodily identity

The ability to choose one's visual bodily identity is not without risks. The concerns are associated with a user's relations to other—individual or commercial—actors in the metaverse. One such concern that must be addressed is the potential for comparison, unrealistic beauty standards, and negative body image, which may arise when users create idealized virtual bodies. Already, images of filtered faces are disproportionately engaged with on social media (Lavrence & Cambre, 2020), and concern has arisen following a growing trend of patients requesting AR filter-inspired plastic surgery procedures (Ramphul & Mejias, 2018). Regardless of the degree of physical trait manipulation, embodying an avatar or applying a filter allows a user to try on different physical characteristics and live out interactions while possessing them. Research demonstrates that VR applications can be deployed in clinical settings to alleviate eating and weight disorders (Riva *et al.*, 2021), but XR applications used in non-clinical settings have the potential to have the opposite effect. Already, much blame has been cast on social media applications for a rise in body dysmorphia and eating disorders among adolescents (Rizwan *et al.*, 2022).

Another risk to consider is the potential for users to experience dissociation, or disconnection, from their physical-world identities. Dissociation can result in identity confusion, and loss of control over behavior, thoughts, and emotions (Vanderlinden *et al.*, 1993). In an early work discussing online virtual worlds, Toronto (2009) noted that overinvolvement in these worlds can result in dissociation, and in a review, Guglielmucci *et al.* (2019) confirmed that excessive time spent in virtual worlds is linked to a variety of dissociative phenomena like depersonalization and escapism. Users may find that they enjoy the impactful experience of embodying one or many different identities in XR applications, but this identity (re)construction may lead to addictions or other behavioral issues (Huang *et al.*, 2021). Based on these findings, we may predict that involvement with a virtual identity over an increased proportion of one's daily live may lend itself to irreversible extents of dissociation.

The risk of dissociation is also linked to the potential for online disinhibition to occur when a new identity is adopted in the metaverse. This happens when individuals self-disclose or act out more frequently or intensely than they would in real life (Suler, 2004). According to Suler, online disinhibition can manifest as benign and toxic disinhibition, with the contributing factors including dissociative anonymity and imagination. While some may feel inclined to commit acts of kindness, show generosity, or share details of their personal life (benign disinhibition), others may espouse rude language, harsh criticisms, threats, or engage in salacious activities that they would not otherwise engage in in the physical world (toxic disinhibition). Many findings (e.g., by Banakou *et al.*, 2018; Gorisse *et al.*, 2021; Johnston *et al.*, 2023) suggest that adopting new personas in VR can influence one's perspective and actions in real life, thereby online disinhibition could potentially affect the way an individual acts offline.

Adopting a particular virtual identity and engaging in identity-play also has the potential to garner unwanted attention. The concept of the metaverse and consumer-grade extended reality technologies are relatively new, but research has already demonstrated the landscape of embodied harassment. A series of studies conducted by Freeman and Maloney (2021) and Sykownik *et al.* (2022) demonstrate that users experience varying degrees of harassment based on the perceived age, gender, and ethnicity of their avatars, yet they still tend to recreate the attributes of their physical self. This removes the layer of anonymity that XR systems afford and leaves users, particularly those of marginalized groups, susceptible to personal and disruptive attacks. While interactions in these spaces may resemble face-to-face interactions, these findings suggest that metaverse users can exhibit toxic disinhibition based on visible social identity markers, and with the increase of the time spent in virtual worlds, the effects of these on users—not least of children and adolescents' social and mental development—can increase.

Ethical and legislative considerations for user identity

As explained above, the process of identity-building in virtual environments occurs through the socio-technical infrastructure of the platform and is subject to its specifications and modes of operation: digital technologies embed their specific norms, values, and regulatory mechanisms that define user behaviors (Celeste *et al.*, 2023). To determine the ethical, social, and legislative implications of human identities in the metaverse, we can build on conclusions from social media use cases. These conclusions reflect that digital technologies and platforms are not just neutral intermediaries between users and the content generated by them. Rather, they contribute to shaping online communication, flow of opinions, trends, behavioral norms, and, importantly, the manifestations of identity (Gillespie, 2021; Grande Branger, 2023). This follows what van Dijck and Poell (2013) described as the social media logic; that is, the processes that govern users' contributions, creative practices, and outreach. It also highlights the role of datafication, the methods for quantifying users' social connections. This means

that interpersonal communication and engagement within digital social platforms generate data points displaying user networks and the trajectories of actions, messages, content, and the like.

To fully understand how technological affordances impact identity-building through the case of social media, we also need to consider online communication platforms' business models. A common pattern among major social media platforms is user data monetization, whereby user data is used for targeted advertising, predictive analyses, and other forms of influence on individual users (Gillespie, 2021; Zuboff, 2019). Data monetization implies maximizing user engagement and content moderation or curation practices. In order to amplify the collected data points and thus raise the value of their advertising spaces, platform providers need to increase the number of users and the amount of time they spend on the platform. This is achieved by providing engaging digital environments that can extend or even replace spheres of socialization, even at the expense of marginalized social groups.

The implications of platform-mediated identity building

The key risks of building identities and basing social interactions on them in digital environments like the metaverse will be analyzed through the following three key elements: the potential for manipulation, surveillance, and the limitation of users' freedom of expression.

Digital multiuser platforms hold the risk of manipulating users' identity-building to serve economic interests. Research on social media has already demonstrated how popularity metrics can have a detrimental effect on individuals' self-perception and expression (Gonzales & Hancock, 2011; Grande Branger, 2023). These metrics give rise to addictive reinforcement mechanisms, where individuals are motivated to repeat or adopt behaviors and appearances that are expected to elicit desired reactions. It follows, that the actions and looks of these users will also influence those who engage with them or follow them, leading to toxic behaviors or unrealistic beauty standards, which can increase the risk for low self-esteem or medical conditions, like eating disorders.

Design cues and system affordances in virtual environments can also lead users to think that they have a given set of options for customizing their experiences. This can happen through limiting choices or amplifying certain options, which can coerce people into decisions, for example, regarding their avatars' looks, behaviors, and communication. Such practices have been demonstrated to trigger filter bubbles or echo chambers and the polarization of attitudes (Spohr, 2017). They also contribute user data being used for personalized (commercial) recommendations that imbricate users in sophisticated stimulus-response mechanisms to steer consumer behaviors. This was demonstrated, famously, by the Cambridge Analytica case.

Trumping the way social media platforms' business models have adhered to *surveillance capitalism* by monitoring and profiling users (Zuboff, 2019), virtual environments like the metaverse may collect a hitherto unseen range of personal data for commercial purposes (Bibri & Allam, 2022). This extends demographic data and personal identifiers by biometric signals, voice, movement, and other potentially identifiable physiological data collected through the integration of smartphone apps, biometric wearables (e.g., smart watches), motion capture suits, or brain-computer interfaces. This data can be used for predictive analysis. As it was demonstrated by Miller *et al.* (2020) and Buck and McDonnell (2022), an individual in virtual reality can be identified based on data like body motion, and that data points collected from one's body motions can be used for predicting personal details, such as medical conditions or mental states. If users are unaware, such surveillance can lead to the unlawful collection of personal data. If they are aware, it may lead to behaviors that conform to situations where one is observed and hinder natural social interactions. But in both cases, such an extent of surveillance challenges the current definitions of personal data, and, consequently, reflects the urgent need for outlining new frameworks for data protection of XR users.

It has been confirmed in previous research that social media platforms' design and the governing companies' business models can define users' freedom of self-determination and expression (Jørgensen & Zuleta, 2020). In the case of virtual multiuser environments, it has also been pointed out that social interactions are outlined by an amalgam of interaction design, communicational affordances, and newly established social norms that impact the extent to which users can share their opinions, the range of verbal and non-verbal expressions, as well as outreach to other users (Kukshinov *et al.*, 2024). This means that, because of functions like muting a speaker or teleporting to another location (i.e., leaving a potentially uncomfortable social situation), users can curate others' expressions through technological functions in a virtual world. While this often happens to escape toxic behaviors, it also enables censoring other users. While these functions may have been designed to support the safe use of XR applications, they also provide space for developers, businesses, and investors (e.g., companies buying commercial space in the metaverse) to censor opinions that contravene their interests.

The role of developers in defining the range of customizations in virtual identities through avatar design or communication functions also holds the risk of delineating "legitimate" identities. Thus, for instance, if a platform only enables a binary choice of gender or lacks options for designing one's avatar with a wide range of body types, bodily abilities, skin colors, or religious expressions, it will strengthen social biases and stigma against marginalized groups. In addition, users won't be able to express the identities they desire to express, which will impact their social connections in virtual spaces, too. Based on this logic, businesses may encompass the power to define the boundaries of self-expression and limit the abilities of users to redress inequalities (Palladino, 2023).

The interplay between personal data and identity in the metaverse is another cause for disquiet. When individuals are granted unprecedented agency to craft and mold their digital personas, it can lead to questions about authenticity. The choice of

self-expression through avatars might empower individuals to explore and express various facets of themselves. This can be used for innocent purposes, like trying out a new gender identity or playing a different role in a virtual game. However, it can also reflect malicious practices such as catfishing (posing as someone else to deceive others), cyberbullying, or engaging in fraudulent activities. In addition, such fluidity of digital identities can also advance the spread of AI-driven fake avatars that can disseminate fake news or extremist propaganda, coerce users into buying certain products, or abuse personal data. Although this is far from a new phenomenon, having been observed on social media platforms, the unique affordances of the metaverse could nevertheless amplify the intensity of these malicious interactions, and could potentially increase their impact. Therefore, it is important to respond to this phenomenon on legal levels, for instance by compelling developers or commercial actors to disclose AI-generated identities.

Privacy concerns and solutions of AI use in the metaverse

In the context of the metaverse and other virtual worlds, AI provides diverse ways of creating digital personas. Recent developments in Large Language Models (LLMs), such as GPT 4.0, and generative AI techniques have created new levels of eloquence and conversation generation that were heretofore impossible. These AI-generated elements respond to user input in real-time, which can enhance user experience (for example, chatbots or translation services), but, as we explained above, can also be malicious.

Trained by fine-tuning, prompt engineering, or enhanced external information sources, AI-generated avatars, language-based user profiles, and synthetic audio and video content, are often fully or partially based on personal data collected from actual users. The handling of massive amounts of personal data can impact the overall performance of metaverse systems, which not only affects user experiences but also demands high energy consumption and imposes storage, processing, and networking requirements. Although AI models can effectively infer user actions while minimizing the amount of data accessed, they are still generally trained on personal data deriving from platform use. Therefore, we argue, it is important for metaverse developers to strike a balance between the demand for user data and the need to ensure a smooth and seamless user experience while minimizing environmental impact.

Besides system-based and environmental impacts, a significant issue with AI-generated profiles in the metaverse is the increased difficulty of distinguishing them from actual users, which highlights the importance of data security, transparency, accountability, and responsibility. This is because AI-generated profiles are versatile as they accurately simulate human-operated ones and can rapidly adapt to situations, actions, and even conversations. Therefore, in parallel to developments in AI and the metaverse, new regulatory and practice-based safeguards are needed.

AI-generated profiles operate on the continuous collection, analysis, and storage of personal data, which could pose serious risks to user privacy in the metaverse. The sensitivity of personal data cannot be overstated, and it is of utmost importance that user privacy is protected alongside, and in spite of the benefits of, any real-time data analysis. Since the process of creating digital identities involves personal data and AI technology, it raises important questions about privacy, security, and ownership of said personal data. Consent to the use of personal data has been long implemented in digital systems (e.g., cookie settings in web browsers, data collection policies to be approved in digital social networks). However, the amount, variety, and sensitivity of collected personal data are enhanced in XR-based systems, as we explain above, which require novel policies and technological solutions. A potential solution points to ways in which users can understand and manage their digital identities and maintain control over their personal data. Other, technology-based solutions include decentralized data sharing with federated learning and robust AI interference with privacy protection.

Federated learning is an innovative technique that allows machine learning to be performed on decentralized multimodal data (Huang *et al.*, 2024). This means that the training of models can take place on data that remains on a user's device, without requiring its transfer to a central server where it may become vulnerable to misuse or hacking. With the development of virtual reality systems, this technique has been assessed as a possibility to protect user privacy (Flores-Martin *et al.*, 2024). By allowing metaverse applications, such as generative avatars, to run on data that remains on users' devices, federated learning ensures that user privacy is preserved while also improving performance. It is particularly useful in the analysis and integration of biometric data, location, and gestures.

Conventional privacy-preserving mechanisms, such as anonymization or pseudonymization, do not provide a solid guarantee of user privacy. Meanwhile, the naive combination of advanced privacy mechanisms and neural network architectures may lead to unexpected model quality deterioration. For example, when using federated learning in an adversarial training process, the training of the AI model may fail to converge (Shen *et al.*, 2023). AI inference can enhance various functionalities of metaverse applications, including motion detection and face tracking. However, it is important to ensure that large amounts of precise user data are not stored and analyzed when it is not necessary. By integrating a privacy-preserving mechanism into the AI model training and inference process, we can ensure rigorous privacy protection throughout the entire AI model training and serving stage in the metaverse.

Conclusions

The development of new technologies—including extended reality technologies, language models, and many more—often happens in a vacuum. Although developers may have a clear idea of a technology's or application's public benefits, economic interests often overshadow usability principles and the *de facto* implications on users, their wellbeing, privacy, or social and economic interests. In addition, end-users are often viewed as entities that engage with technologies in pre-determined ways (determined by the technological

affordances of UX design) rather than a community of sovereign and autonomic entities that not only use but also shape technologies and use practices. Based on these problematics and challenges, we argue for adjusting the ratio of performance-based and user-centric design to highlight how technology can aid human users.

The metaverse concept presents a massive immersive multiuser space for interpersonal connections, where people can socialize, learn, do business, etc. It is a direct descendant (or little sibling) of social virtual reality applications, which encompass 3D immersive spaces but usually only for limited or specialized functions (e.g., concert venues, schools, cinemas). The metaverse promises a seamless connection between various immersive spaces and functions, where users appear behind a persistent virtual identity. As with all forms of digital presence, this identity inherently involves two key aspects: a digital representation and digitally stored personal data. In the existing social VR applications, digital representation is highly dependent on how an application is designed, and what adjustments and characteristics it enables. These limitations are often unavoidable or are necessary evils in enabling the presence of certain functions or seamless interactions. However, they often lead to the misrepresentation of certain bodily characteristics, abilities, or demographic markers, which may have negative impacts on user identity, marginalized groups, behaviors, and communication.

Privacy and data management in the metaverse carry profound human implications. As immersive multiuser environments encompass a growing range of people's daily activities, it is imperative that we prioritize and address these implications and design digital experiences that protect, rather than diminish, users' well-being and privacy. This is all the more important in light of the great number of questions raised by the possibility of the metaverse regarding the authenticity of digital identities: personal data (often inadvertently) shared during use can be subject to misuse in the form of data theft and be re-used for generating synthetic avatars. Even when they are not malicious in the sense of using personal data to cause harm, such AI-generated synthetic avatars can nevertheless contribute to the misrepresentation of both individual users and certain demographic groups, as well as to a general mistrust in online communication or social connections. Preserving the integrity of individuals' identities and protecting them from impersonation or manipulation is crucial. This is not only a matter of privacy but also of personal security and trust in digital interactions.

Ethics and consent

Ethical approval and consent were not required.

Disclaimer

The views expressed in this article are those of the author(s). Publication in Open Research Europe does not imply endorsement of the European Commission.

Data availability

No data associated with this article.

Acknowledgements

The authors thank Caitriona Curtis for managing this project and Órlaith Darling for the help in copyediting.

References

Abrams D, Hogg MA: **Social identity and self-categorization.** In: J. F. Dovidio, *et al.* (Eds.): *The SAGE handbook of prejudice, stereotyping and discrimination*. SAGE, 2010; 179–193.
Publisher Full Text

Ball M: **The metaverse: and how it will revolutionize everything**. Liveright, 2022.
Reference Source

Banakou D, Kishore S, Slater M: **Virtually being Einstein results in an improvement in cognitive task performance and a decrease in age bias.** *Front Psychol.* 2018; **9**: 917.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Bibri SE, Allam Z: **The Metaverse as a virtual form of data-driven smart cities: the ethics of the hyper-connectivity, datafication, algorithmization, and platformization of urban society.** *Comput Urban Sci.* 2022; **2**(1): 22.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Buck L, McDonnell R: **Security and privacy in the metaverse: the threat of the digital human.** In: *Proceedings of the 1st Workshop on Novel Challenges of Safety, Security and Privacy in Extended Reality*. 2022.
Reference Source

Burin D, Kilteni K, Rabuffetti M, *et al.*: **Body ownership increases the interference between observed and executed movements.** *PLoS One.* 2019; **14**(1): e0209899.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Celeste E, Palladino N, Redeker D, *et al.*: **The content governance dilemma: digital constitutionalism, social media and the search for a global standard**. Palgrave Macmillan, 2023.
Publisher Full Text

Dengah HJF, Snodgrass JG: **Avatar creation in videogaming: between compensation and constraint.** *Games Health J.* 2020; **9**(4): 265–272.
PubMed Abstract | **Publisher Full Text**

Flores-Martin D, Díaz-Barrancas F, Pardo PJ, *et al.*: **Privacy and performance in virtual reality: the advantages of federated learning in collaborative environments.** *J Web Eng.* 2024; **23**(8): 1085–1106.
Publisher Full Text

Freeman G, Acena D: **Hugging from a distance: building interpersonal relationships in social virtual reality.** In: *Proceedings of IMX '21: ACM International Conference on Interactive Media Experiences*. 2021; 84–95.
Publisher Full Text

Freeman G, Maloney D: **Body, avatar, and me: the presentation and perception of self in social virtual reality.** In: *Proceedings of the ACM on Human-Computer Interaction*. 2021; **4**: 239.
Publisher Full Text

Genay ACS, Lécuyer A, Hachet M: **Being an avatar "for real": a survey on virtual embodiment in augmented reality.** *IEEE Trans Vis Comput Graph.* 2022; **28**(12): 5071–5090.
PubMed Abstract | **Publisher Full Text**

Gillespie T: **Custodians of the Internet: platforms, content moderation, and the hidden decisions that shape social media**. Yale University Press, 2021.
Publisher Full Text

Gonzales AL, Hancock JT: **Mirror, mirror on my Facebook wall: effects of**

exposure to Facebook on self-esteem. *Cyberpsychol Behav Soc Netw.* 2011; **14**(1–2): 79–83.
PubMed Abstract | **Publisher Full Text**

Gorisse G, Senel G, Banakou D, *et al.*: **Self-observation of a virtual body-double engaged in social interaction reduces persecutory thoughts.** *Sci Rep.* 2021; **11**(1): 23923.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Grande Branger LA: **Mediated identities: how Facebook intervenes in the virtual manifestation of our identities.** In: V. Kannen & A. Langille (Eds.): *Virtual Identities and Digital Culture*. Routledge, 2023; 38–47.
Publisher Full Text

Guglielmucci F, Monti M, Franzoi IG, *et al.*: **Dissociation in problematic gaming: a systematic review.** *Curr Addict Rep.* 2019; **6**(1): 1–14.
Publisher Full Text

Han E, Miller MR, DeVeaux C, *et al.*: **People, places, and time: a large-scale, longitudinal study of transformed avatars and environmental context in group interaction in the metaverse.** *J Comput Mediat Commun.* 2023; **28**(2): zmac031.
Publisher Full Text

Huang J, Kumar S, Hu C: **A literature review of online identity reconstruction.** *Front Psychol.* 2021; **12**: 696552.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Huang W, Wang D, Ouyang X, *et al.*: **Multimodal federated learning: concept, methods, applications and future directions.** *Inform Fusion.* 2024; **112**: 102576.
Publisher Full Text

Hutmacher F: **Why is there so much more research on vision than on any other sensory modality?** *Front Psychol.* 2019; **10**: 2246.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Johnston T, Seinfeld S, Gonzalez-Liencres C, *et al.*: **Virtual reality for the rehabilitation and prevention of Intimate Partner Violence – from brain to behavior: a narrative review.** *Front Psychol.* 2023; **13**: 788608.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Jørgensen RF, Zuleta L: **Private governance of freedom of expression on social media platforms: EU content regulation through the lens of human rights standards.** *Nordicom Review.* 2020; **41**(1): 51–67.
Publisher Full Text

Kilteni K, Groten R, Slater M: **The sense of embodiment in virtual reality.** *Presence Teleop Virt Environ.* 2012; **21**(4): 373–387.
Publisher Full Text

Kukshinov E, Harley D, Szita K, *et al.*: **Disembodied, asocial, and unreal: how users reinterpret designed affordances of social VR.** In: *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 2024; 1914–1925.
Publisher Full Text

Lavrence C, Cambre C: **"Do I look like my selfie?": filters and the digital-forensic gaze.** *Soc Media Soc.* 2020; **6**(4): 2056305120955182.
Publisher Full Text

Miller MR, Herrera F, Jun H, *et al.*: **Personal identifiability of user tracking data during observation of 360–degree VR video.** *Sci Rep.* 2020; **10**(1): 17404.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

O'Meara J, Szita K: **AR cinema: visual storytelling and embodied experiences with augmented reality filters and backgrounds.** *Presence Virtual Augment Real.* 2021; **30**: 1–25.
Publisher Full Text

Palladino N: **A 'biased' emerging governance regime for artificial intelligence? How AI ethics get skewed moving from principles to practices.** *Telecommun Policy.* 2023; **47**(5): 102479.
Publisher Full Text

Ramphul K, Mejias SG: **Is "Snapchat dysmorphia" a real issue?** *Cureus.* 2018; **10**(3): e2263.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Ritterbusch GD, Teichmann MR: **Defining the metaverse: a systematic literature review.** *IEEE Access.* 2023; **11**: 12368–12377.
Publisher Full Text

Riva G, Malighetti C, Serino S: **Virtual reality in the treatment of eating disorders.** *Clin Psychol Psychother.* 2021; **28**(3): 477–488.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Rizwan B, Zaki M, Javaid S, *et al.*: **Increase in body dysmorphia and eating disorders among adolescents due to social media.** *Pakistan BioMed J.* 2022; **5**(3): 148–152.
Publisher Full Text

Shen Z, Ye J, Kang A, *et al.*: **Share your representation only: guaranteed improvement of the privacy-utility tradeoff in federated learning.** In: *The Proceedings of the Eleventh International Conference on Learning Representations*. 2023.
Publisher Full Text

Sparks P, Shepherd R: **Self-identity and the theory of planned behavior: assessing the role of identification with "green consumerism".** *Soc Psychol Quart.* 1992; **55**(4): 388–399.
Publisher Full Text

Spohr D: **Fake news and ideological polarization: filter bubbles and selective exposure on social media.** *Bus Inform Rev.* 2017; **34**(3): 150–160.
Publisher Full Text

Suler J: **The online disinhibition effect.** *Cyberpsychol Behav.* 2004; **7**(3): 321–326.
PubMed Abstract | **Publisher Full Text**

Sykownik P, Maloney D, Freeman G, *et al.*: **Something personal from the metaverse: goals, topics, and contextual factors of self-disclosure in commercial social VR.** In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022; 632.
Publisher Full Text

Szita K: **A virtual safe space? An approach of intersectionality and social identity to behavior in virtual environments.** *J Digit Soc Res.* 2022; **4**(3): 34–55.
Publisher Full Text

Tajfel H, Flament C, Billig MG, *et al.*: **Social categorization and intergroup behaviour.** *Eur J Soc Psychol.* 1971; **1**(2): 149–177.
Publisher Full Text

Tajfel H, Turner JC: **The social identity theory of intergroup behaviour.** In: S. Worchel & W. G. Austin (Eds.), *Psychology of Intergroup Relations.* Nelson-Hall, 1986; 7–24.
Reference Source

Toronto E: **Time out of mind: dissociation in the virtual world.** *Psychoanal Psychol.* 2009; **26**(2): 117–133.
Publisher Full Text

Turkle S: **Multiple subjectivity and virtual community at the end of the Freudian century.** *Sociol Inq.* 1997; **67**(1): 72–84.
Publisher Full Text

van Brakel V, Barreda-Ángeles M, Hartmann T: **Feelings of presence and perceived social support in social Virtual Reality platforms.** *Comput Human Behav.* 2023; **139**: 107523.
Publisher Full Text

van Dijck J, Poell T: **Understanding social media logic.** *Multidiscip Stud Media Commun.* 2013; **1**(1): 2–14.
Publisher Full Text

Vanderlinden J, Van Dyck R, Vandereycken W, *et al.*: **The Dissociation Questionnaire (DIS-Q): development and characteristics of a new self-report questionnaire.** *Clin Psychol Psychother.* 1993; **1**(1): 21–27.
Publisher Full Text

Yee N, Bailenson J: **The Proteus effect: the effect of transformed self-representation on behavior.** *Human Commun Res.* 2007; **33**(3): 271–290.
Publisher Full Text

Zuboff S: **The age of surveillance capitalism: the fight for a human future at the new frontier of power.** PublicAffairs, 2019.
Reference Source

Open Peer Review

Current Peer Review Status: ? ?

Version 1

Reviewer Report 17 October 2025

https://doi.org/10.21956/openreseurope.22088.r61637

?

Dimitris Kogias
University of West Attica, Egaleo, West Attica, Greece

The article deals with an examination of user identity in metaverse environments dealing with perspectives from ethics, psychology and technical view. Regarding the latter, it emphasizes on how AI can reshape the human presence and data in virtual spaces. While the technical section is descriptive and frames AI as both a facilitator of personalization and a potential threat to privacy, authenticity, and autonomy, the analysis remains conceptual. That is there is lack of concrete examples that could help the reader better understand the message of the authors. In general, there is lack of references to real life VR or AR applications that could help the reader better connect the dots and understand.
Overall, the paper is timely and relevant for *Open Research Europe*, providing a strong conceptual foundation that would benefit from moderate revision to integrate real-world examples on AI and Virtual Worlds.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
Partly

Does the article adequately reference differing views and opinions?
Partly

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Partly

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Yes

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I have a more technical background on blockchain and AI that allows me to focus more efficient on the technical details of the article.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 01 October 2025

https://doi.org/10.21956/openreseurope.22088.r60280

?

Stavros Kaperonis
Panteion University of Social and Political Sciences, Athens, Greece

The article examines identity in metaverse like spaces. More specific how people appear (avatars/voice/behavior) and how platforms collect and use their data. It flags benefits (expression, inclusion) and risks (harassment, body-image pressure, surveillance, AI misuse), and calls for a legal/ethical basis, privacy by design, and transparency around AI-generated identities.

Rationale and purpose
The abstract and the introduction clearly state the problem and the aim.

Differing views/opinions
It notes pros/cons (e.g., clinical VR benefits vs. social harms; safety tools vs. censorship risk) but doesn't fully develop opposing evidence or operator constraints.

Factual accuracy & support
Core claims cite prior work, but some forward-looking statements read as predictions and need tighter sourcing.

Accessible language
Mostly clear, but several specialist terms like, affordances, Proteus effect, datafication, interoperability, aren't briefly defined, and there's no mini-glossary.

Recommendations

it gives directions but not step-by-step guidance.

Must-fix points
1. use cautious wording (“may,” “under conditions”) and cite reviews/longitudinal evidence where available.
2. Add concrete XR examples when arguing that safety tools can slide into censorship.

Is the rationale for the Open Letter provided in sufficient detail? (Please consider whether existing challenges in the field are outlined clearly and whether the purpose of the letter is explained)
Yes

Does the article adequately reference differing views and opinions?
Partly

Are all factual statements correct, and are statements and arguments made adequately supported by citations?
Partly

Is the Open Letter written in accessible language? (Please consider whether all subject-specific terms, concepts and abbreviations are explained)
Partly

Where applicable, are recommendations and next steps explained clearly for others to follow? (Please consider whether others in the research community would be able to implement guidelines or recommendations and/or constructively engage in the debate)
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Yes, I can evaluate the article because my expertise in UX | UI Interaction Design, digital transformation, and the use of AI, lets me judge the clarity of the goals and arguments, the quality of evidence, the balance of benefits and risks for users, the practicality of the recommendations, and the accessibility of the language. For highly technical issues (e.g., advanced privacy/ML methods), long-term psychological effects in XR, and detailed legal interpretation (GDPR/DSA), I would complement my review with input from subject-matter experts.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.