frontiers | Frontiers in Virtual Reality

TYPE Policy and Practice Reviews
PUBLISHED 12 May 2025
DOI 10.3389/frvir.2025.1520655

Check for updates

OPEN ACCESS

EDITED BY
Mel Slater,
University of Barcelona, Spain

REVIEWED BY
Andres Chomczyk Penedo,
Vrije University Brussels, Belgium
Sara Louredo,
University of Vigo, Spain

*CORRESPONDENCE
Giovanni Sorrentino,
✉ giovanni.sorrentino5@unibo.it
Javier López-Guzmán,
✉ javier.lopez-guzman@ua.es

RECEIVED 31 October 2024
ACCEPTED 15 April 2025
PUBLISHED 12 May 2025

CITATION
Sorrentino G and López-Guzmán J (2025) Rethinking privacy for avatars: biometric and inferred data in the metaverse.
Front. Virtual Real. 6:1520655.
doi: 10.3389/frvir.2025.1520655

Rethinking privacy for avatars: biometric and inferred data in the metaverse

Giovanni Sorrentino[1,2]* and Javier López-Guzmán[3]*

[1]CIRSFID, University of Bologna, Bologna, Italy, [2]International School of Advanced Study, University of Camerino, Camerino, Macerata, Italy, [3]Global Innovation Law and Policy Research Group (GIP-LAW), University of Alicante, Alicante, Spain

As the metaverse expands, it introduces novel challenges to data privacy and protection, particularly in the handling of biometric and inferred data. This paper examines the implications of existing European legal frameworks, notably the General Data Protection Regulation (GDPR), on the processing of biometric information and other sensitive data within immersive virtual environments. In the metaverse, avatars—often designed to closely resemble their creators—serve as rich sources of both explicit and inferred personal data, raising significant concerns regarding user consent, data processing, and privacy. By comparing regulations from both the EU and the United States, including the newly proposed American Privacy Rights Act, this study highlights the gaps in current legal protections surrounding avatars and the biometric or inferred data they may reveal. The findings indicate an urgent need for updated regulatory approaches to address the unique data privacy challenges of virtual identities and underscore the need for transparency and user control over digital representations in evolving digital landscapes.

KEYWORDS

metaverse, biometrics, data, GDPR, privacy, avatars, digital identity, AI act

1 Introduction

Technology often moves faster than the legislative process. This will be, and already is, the most significant challenge that legislators will face soon. The European legislator is not exempt from this challenge. Despite the significant efforts made by the EU in technology regulation, there is a question as to whether these regulations are suitable when the technology they regulate evolves over time. The metaverse is an example of this. Considering the absence of specific European regulation that thoroughly governs the dissemination and use of the metaverse, including data protection aspects, it will be necessary to rely on the guidance provided by the General Data Protection Regulation (GDPR)[1]. Specifically, the present paper aims to analyze how the massive use of avatars has significant consequences for the processing of biometric data and inferred data.

In particular, the extensive use of avatars raises numerous questions. In recent years, we have witnessed a surge in avatar usage that shows no signs of slowing down (Global Market

[1] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Size, 2024). Technological advancements allow for the design of increasingly lifelike and realistic avatars. An example is the "Instant Codec Avatar" by Meta, a technology capable of reproducing our biometric data with extreme precision, or "Ready Player Me," a provider that enables users to create their own avatar and use it across various platforms.

The significance of avatars is driven by multiple factors. First, they are being used more frequently in the digital world. Additionally, an avatar can reveal important information about its creator. When designing their avatar, the user makes choices, attributing to their digital alter ego the characteristics they desire. Thus, the avatar becomes a reflection of these choices, with the user embedding a range of personal information in the creation process. For cases like these, regulation is still minimal, exposing various gaps.

For this reason, it will be essential to understand where the Data Protection framework is perfectly adaptable to the metaverse and where, instead, it will require interpretative activity. The regulation on the protection of personal data is understood in a wide sense in this work, starting on the broader privacy legislation developed in some jurisdictions, such as the United States of America, and the more specific aspects of EU legislation.

The paper aims first to analyze the regulations applicable to biometric and inferred data. Secondly, a brief literature review on avatar creation will be conducted to understand user behavior in creating their digital alter egos. We hypothesize that, with technological advancements, avatars will become increasingly lifelike and similar to reality. This raises the main research question that the paper tries to answer: how are privacy regulations related to biometric and inferred data structured? And are they sufficient to protect these types of data, considering the increasing use of avatars in the metaverse? This question drives the analysis of the adequacy of existing frameworks, such as the GDPR, in addressing the challenges posed by technological advancements and the growing use of avatars as digital representations. In particular, the paper aims to determine whether current privacy regulations can adequately address scenarios in which an avatar's appearance could precisely identify the user who created it or at least allow sensitive information to be inferred. This raises other questions, such as, for example,: can sensitive data be inferred from avatars that reflect real user characteristics and identities? And what type of legal protections should apply in such cases?

To try to answer these questions we used a methodology based on: i) regulatory analysis; ii) comparison between jurisdictions; iii) literature review; iv) analysis of platform privacy policies.

The paper examines the European and U.S. regulatory frameworks, analyzing, in particular the European Union's General Data Protection Regulation (GDPR) and other relevant regulations, such as the California Consumer Privacy Act (CCPA) and the Biometric Information Privacy Act (BIPA) in the United States. Through this analysis, the strengths and gaps in the protection of biometric and inferred data in the context of the metaverse are highlighted.

Extensive literature exists on the study and development of privacy policies, as highlighted by Zaeem and Barber (2020) and Malgieri (2021). A review of academic research is used to explore the discipline and to understand how users configure and use avatars, with a focus on the biometric and inferred data that can be extracted from them. Finally, we also analyzed the privacy policies of major metaverse platforms, such as Decentraland, Roblox, Epic Games, and Ready Player Me, to assess whether they comply with biometric and inferred data regulations and adequately inform users. The analysis reveals that none of these platforms explicitly seek consent for the collection of biometric data.

Clearly, this study does not purport to answer all these questions definitively. Indeed, such an endeavor would be lengthy and likely lead to hasty or erroneous conclusions. Instead, the intention is to highlight a potential issue, hitherto overlooked, and to initiate a discussion within the academic community.

2 Biometric data

The Oxford English Dictionary defines "biometrics" as referring to or describing physical traits that serve as unique identifiers for individuals, such as fingerprints and iris patterns (Biometrics, 2021). In general terms, biometrics involves the measurement of biological signals (Yannopoulos et al., 2008).

Previously, biometric identification was also described as: "rather than being something that an individual knows or has, it is something that they are" (Hopkins, 1999). This highlights the most crucial aspect of biometrics: its uniqueness. Indeed, out of billions of people, only one individual can have a specific fingerprint or iris pattern. In other words, biometrics refers to the measurement of physical aspects of the human body. This includes skin patterns, blood vessel networks beneath the skin, genetic code patterns, facial characteristics such as the distances between features like the eyes, nose, and mouth, as well as behavioral traits like gait (Smith et al., 2018). The use of biometrics has many potential applications in the metaverse. Some of these could enhance and secure the user experience. For example, biometric data can be used to create realistic avatars, access restricted areas, or prove identity.

We can categorize biometrics as: "strong," "weak," and "soft" identifiers. Strong identifiers, such as fingerprints, iris, and retina patterns, enable or confirm the unique identification of an individual. Weak biometrics, on the other hand, are less distinct or less consistent, such as body shape, behavioral patterns, voice, and body sounds.

Biometrics has moved beyond its early stages, once considered to be something out of science fiction. Today, it is being widely adopted for user identification and authentication in various fields, including information systems, border security, and healthcare. The technology's growing affordability and improved performance have made it suitable for both consumer use and government applications (Mordini and Petrini, 2007). As technology has progressed, the potential applications of biometric data across various fields have become increasingly apparent. However, the sensitivity and privacy concerns associated with handling this data have been recognized from the outset.

Recent technological advancements, particularly in data analytics and artificial intelligence, as well as improvements in hardware such as faster computers, high-resolution cameras, and IoT devices, have significantly enhanced the potential of biometric techniques and broadened their scope of application. As a result, this type of data has become highly appealing to private entities interested in processing or using it. In the last few years, the private sector has shown growing interest in the widespread use of advanced biometric technologies, which were previously limited to law enforcement applications.

This happens also in the metaverse. In this digital environment, a vast amount of data is collected. These data can, for instance, be simple "identifying data", "human characteristics data", or also so-called "inferred data". The first question to ask is whether "human characteristics data" includes only not particularly sensitive personal data, or whether it also includes a category of particularly sensitive data, such as biometric data. While the promise is that future developments will increase the usability, accuracy, and robustness of existing biometric technologies, technical capabilities have also given rise concerning trends and applications. Specifically, concerning avatars, the biometric data (or at least the inferred data) that can potentially be extracted and processed requires greater attention.

Given the sensitivity of biometric data, lawmakers around the world have paid special attention to regulating them. In order to determine the potential and impact of inferring biometric data in the Virtual Worlds, a comparative analysis is relevant. The orientation of the regulation in the future regarding biometric data may favour or limit very importantly the inference of personal data and the development of the Metaverse as a digital service.

2.1 Regulatory framework in EU and US

In particular, the processing of biometric data is subject to specific regulations worldwide. Operators that track physiological, mental, and biometric data are subject to laws governing the collection, use, and sharing of sensitive data and biometrics.

The EU GDPR is one of the most recent and powerful regulations passed to protect consumers' data. The GDPR has inspired sweeping new legislation in the US and continues to be the most widely referenced privacy regulation as new regulations are considered in the US and around the globe (Zaeem and Barber, 2020). It has a specific and restrictive regulation about this type of data. In the United States of America, several states have proposed and even passed legislation restricting the use of biometrics. This is the case of California, Colorado, Connecticut, Utah and Virginia and especially Illinois, which adopted the Biometric Information Privacy Act (BIPA)[2] in 2008 (Ball, 2022).

Currently in the United States, the legislative landscape regarding biometrics is dominated by state laws. The first state law devoted exclusively to biometrics regulation was Illinois' groundbreaking Biometric Information Privacy Act (BIPA), which was passed in 2008. The BIPA's definition of biometric does not consider "writing samples, written signatures, photographs, human biological samples used for valid scientific testing or screening, demographic data, tattoo descriptions, or physical descriptions" as forms of biometric data. Under BIPA, all private entities that obtain biometric information are required to publicize a policy regarding their methods of managing and destroying this type of collected information. Consequently, the protection granted to avatars will depend on the classification assigned to them. If avatars are defined similarly to photographs, they would not receive the more stringent protection reserved for biometric data. This contrasts with the scenario in which they are classified as "facial geometry," a category that would warrant stricter protection measures.

In California, the California Privacy Rights Act came into force In January 2023, which was implemented to amend the existing California Consumer Privacy Act (CCPA). It is worthwhile to examine Californian legislation for at least two reasons. First, California is home to the so-called "Silicon Valley," where the headquarters of the world's leading tech companies are located, making this legislation likely to have a substantial impact on them. Additionally, the CCPA is frequently presented as a potential U.S. data privacy law model. The regulation inspires many national laws outside the E.U., including Chile, Japan, Brazil, South Korea, Argentina, and Kenya. CCPA's definition of biometric data is a bit broader than that of GDPR: "an individual's physiological, biological or behavioral characteristics, including an individual's D.N.A., that can be used, singly or in combination with each other or with other identifying data, to establish individual identity". The GDPR, on the other hand, defines biometric data as "personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data".

As part of the definition, the CCPA provides several examples of biometric information protected by the law. In addition to keystroke patterns, gait patterns, and sleep, health, or exercise data that contain identifying information, it also includes images of the iris, retina, fingerprint, face, hand, palm, vein patterns, and voice recordings, from which an identifying template (such as a fingerprint, minutiae template, or voiceprint) can be extracted. It doesn't matter if a person is out in public when a company collects biometric data; it is still regarded as personal information. Biometric data is also expressly excluded from the definition of publicly available information as long as it is obtained by the company without the consumer's knowledge. Our hypothesis posits that an avatar that closely resembles the user or is generated from biometric data should be regarded as a form of biometric data, warranting similar treatment and consideration in terms of data protection and privacy regulations.

However, and maintaining the spotlight on the legal framework of the United States of America, it is worth noting that the comparison between legal frameworks does not incorporate the analysis of personal data protection as a fundamental right. This is purely a European perspective, that enshrined such right in article 8 (1) of the EU Charter of Fundamental Rights, and article 8 of the European Convention on Human Rights. Whereas the right to privacy is also a constant in European legislation, the American perspective chooses to avoid the use of the term "personal data protection" in the legislation analysed in this work. While it is, in our opinion, a very specific matter of semantics, it is true that the right to privacy covers wider aspects than the right to personal data protection, such as the domicile inviolability, the right to protect the honor and the self-image, and the preservation of intimacy. This

[2] (740 ILCS 14/) Biometric Information Privacy Act.

differentiation is unlikely to have any significant impact in the protection of privacy in the Metaverse.

Still regarding the legal framework of the United States of America, it is worth mentioning that no comprehensive federal privacy law regulates the right to privacy in the digital world and the processing of personal data. Several efforts have been developed in recent years to establish such a comprehensive law, with increasing popular support (Tale, 2022). Despite these efforts, a divided and increasingly polarized U.S. Congress has yet to reach the necessary consensus to enact a comprehensive federal law comparable to the EU's GDPR. Nevertheless, significant efforts have been made in this domain, the most mature and recent being the "American Privacy Rights Act of 2024"[3]. This Act, defines biometric information in Section 101 and excludes certain items from its definition of biometric information as: digital or physical photographs, audio or video recordings, and data derived from these that cannot be used to identify or authenticate a specific individual (Solove and Schwartz, 2024). Under this project of laws, biometric data cannot be considered publicly available data, following the path established by California'S CCPA, limiting the processing of these information and reinforcing the control of the individual over it. If enacted, this regulation would be of deep impact for the Metaverse.

One significant consequence of this regulation would be a strict limitation on the processing of biometric data. If biometric information is not considered publicly available, the creation of hyper-realistic avatars would be permissible. However, controllers—whether platforms or individual users—would not be authorized to process other users" biometric data solely because of the creation and use of avatars containing such data. For instance, automatic identity recognition based on avatar gait analysis would not be allowed. Similarly, large-scale facial recognition systems or any technology enabling automated identity recognition through avatar biometric data would be prohibited in virtual environments (Fazlioglu, 2024).

Furthermore, under the project of American Privacy Rights Act of 2024, biometric information is considered sensitive covered data. Moreover, additional protections for biometric information are foreseen, covering their collection, processing, retention and transfer. The interplay of this federal privacy law project with the existing state laws in the United States would be an interesting topic to analyze. Since this interplay could mean a potential downgrade of the existing protection for biometric data in the state laws. The state laws could be repealed if any federal privacy law were enacted. This analysis, unfortunately, exceeds the scope of this research.

In this regard, the EU already has some limitations imposed on a legal and jurisprudence level. The concept of publicly available information is not determinant for the application of GDPR. However, it is relevant to consider the allowance to process biometric data, which is permitted or prohibited considering the rules on the legal basis for its processing as special categories of personal data. Under art. 9 (2) (e) GDPR, biometric data may be processed if these have been "manifestly made public" by the data subject. Therefore, the processing of inferred data based on the biometrics originated in an avatar would not, in principle, be prohibited. However, the rules on purpose limitation are still applicable. Even if a person has made public some information about himself, these personal data are not available for any processing by any undertaking, without limitations. Under art. 5 (1) (b) GDPR, personal data should only be "collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes". Moreover, even if certain personal data have been made public by the data subject, other limitations may apply. For instance, pictures shared by users of social media cannot be simply reproduced by news agencies in their own websites or printed publications. Even if there is an informational and social relevance to the case, the pictures used to illustrate the pieces of news should respect the right to privacy of their protagonists, who, generally, did not share those pictures previously in their social media profiles for the benefit of the news media.[4] The same limitations, and additional, apply to special categories of personal data, including biometrics.

In light of the recent jurisprudence developed by the Court of Justice of the EU, special categories of personal data publicly available in the Internet cannot be processed in disregard of the purpose limitation principle. The Court ruled in this judgement that: art. 9 (2) (e) GDPR "must be interpreted as meaning that the fact that a person has made a statement about his or her sexual orientation on the occasion of a panel discussion open to the public does not authorise the operator of an online social network platform to process other data relating to that person's sexual orientation, obtained, as the case may be, outside that platform using partner third-party websites and apps, with a view to aggregating and analysing those data, in order to offer that person personalised advertising." Therefore, the fact that special categories of personal data are publicly available, does not allow an undertaking to process those data limitlessly. If the purpose limitation principle is to be applied in this fashion to data revealing the sexual orientation of a data subject, it seems proportionate to develop the same argument, by analogy, to biometric data. Likewise, if biometric data can be inferred from publicly available data, their obtention and processing cannot be developed without certain rules.[5]

In a virtual world platform, an example scenario would be as follows: a user introduces their avatar, which is often a realistic representation of the user. This avatar may have features that are identical or very similar to the user's actual appearance. The facial geometry of the avatar could allow inferences about the facial geometry of the user, thereby creating inferred biometric data within the Metaverse, originating from publicly available information. Since the user has consented (either explicitly or implicitly) to the platform's processing of their data to use the avatar, the platform might assume they are also authorized to process biometric data inferred from this virtual representation.

[3] H.R.8818, American Privacy Rights Act of 2024, 25 June 2024, Rep. Cathy McMorris Rodgers, R-Wash.

[4] Tribunal Constitucional de España, Sentencia 27/2020, de 24 de febrero de 2020, Sala Segunda, Recurso de amparo número 1369/2017.

[5] Case C-446/21, Maximilian Schrems v Meta Platforms Ireland Ltd., formerly Facebook Ireland Ltd. Judgement of the Court of Justice of the European Union, 4 October 2024, ECLI:EU:C:2024:834, 84.

However, this interpretation would not align with the jurisprudence established by the Court in Schrems v Meta Platforms Ireland.

Moving our focus specifically on the European regulation, the art. 4 GDPR defines Biometric data as "Personal data resulting from specific technical processing relating to the physical, physiological or behavioral characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data". Furthermore, according to the current wording of Article 9 (1) GDPR, the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited. It is immediately clear that the European legislators decided to strongly restrict the processing of this type of data. As a result, article 9 paragraph 1 of the GDPR establishes a general ban on the processing of biometric data for the purpose of uniquely identifying a natural person.

This general rule is subject to some derogation in the cases specified in the same Article 9 paragraph 2. The main derogation includes cases where the data subject has given explicit consent.

According to the European Data Protection Board (EDPB), to categorize data as biometric under Article 9, it is necessary to consider the following three elements: i) Nature of data, ii) Means and way of processing, iii) Purpose of processing (European Data Protection Board, 2019).

In general, when personal data is not technically processed for the specific purpose of uniquely identifying an individual, it should not be considered biometric data as defined by Art. 4, point 14) and regulated by Art. 9, paragraph 1 of the GDPR (Bolognini and Carpenelli, 2022).

Given this information, most biometric characteristics and their digital representations alone may not be considered personal data in most cases. Yet, this should be evaluated considering all objective factors. For instance, unique traits such as skin color, tattoos, or size can identify a person in a small sample set. As a corollary, determining whether the sensitive data regime applies to the digital representation of biometric characteristics requires the controller to take into account the nature and sensitivity of the data in question along with the specific circumstances applicable. The processing of data covered by special protection, including biometric data, will be allowed provided that the data subject has given a specific consent to their processing for one or more specific purposes. A contrario, it cannot be an "ordinary consent".

Upon closer examination, however, information such as physical appearance, age, gender, ethnicity, or even sexual orientation (see Schrems v Meta Platforms Ireland) are potentially accessible simply by the avatar. For instance, consider the case of an avatar that bears the same tattoo as the real user; this could be identified among a relatively large group of people. Indeed, even in the case in which they cannot be categorized as biometric data under Article 4, it is undeniable that such data would still fall within the scope of sensitive data, for which specific consent is nonetheless required.

Nevertheless, as we could see analyzing the privacy policy of the main platforms, in the vast majority of cases, explicit consent is not requested, nor is the user notified of the possibility that information may be inferred from their avatar, at the time of avatar creation or platform registration. This would likely influence the user's choices when configuring their avatar.

So, in conclusion, considering that GDPR aims to safeguard a fundamental and sensitive right, namely, personal data, with a particular emphasis on biometric data, it has to be observed that in the context of highly realistic avatars, this objective is not effectively achieved through existing regulatory frameworks or, at the very least, their current interpretations. Also, as some authors suggested, the first step to understanding whether biometric images fall under the sensitive data regime, and thus are afforded the same protection as biometric data should be clarifying the nature and sensitivity of the data in question, considering the technological realities (Sumer, 2022).

2.2 Limitations of the regulations

Despite the fact that European regulations on this matter are comprehensive and rather restrictive, they present at least two limitations. First, they can only be enforced within European borders. Second, as this article seeks to argue, the regulation concerning so-called inferred data is not as thorough as that pertaining to biometric data. This creates the possibility for potential circumvention of the rules. Moreover, it has direct or indirect impact on a number of fundamental rights, enshrined in the EU Charter of Fundamental Rights.

The fundamental rights protected within the European Union's legal framework apply in all instances governed by EU law, but not beyond those contexts. If EU law is not applicable, the protection of fundamental rights must rely on other legal instruments, such as the European Convention on Human Rights (ECHR) or the constitutions of the individual Member States.

2.3 The potential influence of AI act

In order to provide a comprehensive overview, it is pertinent to mention that in 2024, the European Union enacted the AI Act, the first regulatory framework specifically addressing artificial intelligence. A substantial aspect of this legislation is devoted to AI systems that use biometric data or data related to our unique physical characteristics. It has more detailed rules for these systems and it includes numerous definitions and specific rules. On the other hand, the General Data Protection Regulation (GDPR) only has one definition for biometrics. Of course the notion of 'biometric data' used in the AI Act should be interpreted in light of the notion of biometric data as defined in Article 4, point (14) of Regulation (EU) 2016/679 ("GDPR"), Article 3, point (18) of Regulation (EU) 2018/1725 and Article 3, point (13) of Directive (EU) 2016/680.

However, it is interesting to notice how the notion of "biometric identification" referred to in the AI Act is defined more widely than in the GDPR. Indeed, the AI Act defines "biometric identification" the automated recognition of physical, physiological and behavioural human features such as the face, eye movement, body shape, voice, prosody, gait, posture, heart rate, blood pressure, odour, keystrokes characteristics, for the purpose of establishing an individual's identity by comparing biometric data of that individual to stored biometric data of individuals in a reference database, irrespective of whether the individual has given its consent or not. The AI Act therefore addresses cases in which AI systems are used to identify the aforementioned characteristics. Upon closer examination, many of these characteristics could, in theory, also be extracted from a detailed analysis of an avatar created through advanced technologies, like those we are witnessing today. This highlights how the regulatory framework is inconsistent and reveals significant gaps.

3 Inferred data

3.1 Logic behind avatar configuration

As mentioned above, the avatar is the vehicle through which platforms, or third parties, can gain access to the data subject's sensitive information. During the avatar's creation process, customization options are extensive. There are many different types of avatars, and the specific characteristics of an avatar may vary depending on the context in which it is used. Some common types of avatars include (Blockchain Research Lab, 2023):

- Customizable avatars: allow users to personalize their appearance by choosing from various options, including clothing, hairstyles, and facial features, enabling them to create a distinct representation that reflects their individuality.
- Non-customizable avatars: Unalterable avatars, typically used to represent a specific character, persona or role, e.g., in video games or other types of interactive media (Ratan and Sah, 2015)
- Self-representational avatars: are designed to closely mirror the user's physical appearance, personality, or other traits. These avatars are often utilized in virtual or augmented reality settings to enhance the sense of immersion (McMahan, 2003).
- Non-human avatars: Avatars designed to represent non-human beings, such as animals, robots, or mythical creatures, e.g., in video games or other interactive media (Chen et al., 2019).
- Abstract avatars: designed without the intention of representing a specific individual or character. They take on a more symbolic or conceptual form and are often used in virtual or augmented reality environments to enhance immersion or convey particular ideas or themes (Yee et al., 2007).

Many researchers tried to understand the relation between avatar and the real person behind it. Three groups of theories explore the psychological factors that influence the choice of avatars in virtual environments: self-expression and identity, social comparison and group dynamics, and self-esteem and self-regulation. These theories suggest that people choose avatars that reflect their own identities, values, and social status, as well as those that help them navigate and fit in with the social dynamics of virtual environments.

According to some authors, elements like an avatar's facial expressions, body, clothing, gestures, and behaviors can influence user trust, sense of body ownership, group cohesion, as well as the perceived realism and presence within the virtual environment (Aljaroodi et al., 2019).

Another study (Lin and Wang, 2014) highlights that a significant portion of users (73%) create more than one avatar, ranging from 1 to 16 avatars per person, with an average of three. Interestingly, these avatars did not necessarily resemble the physical appearance of the players themselves. Over a third (35.6%) of respondents reported designing avatars that embodied non-human or non-organic forms. Furthermore, there was an even divide when it came to whether the avatar's personality reflected the player's own personality. Lin and Wang also identified four primary motivations driving participants to create avatars in virtual worlds: (1) exploring the virtual space to experience activities that are otherwise challenging or impossible in real life; (2) navigating social dynamics, using avatars to build friendships and social standing; (3) adapting to the cultural and social contexts within these virtual environments; and (4) representing identity, whether as a true reflection of themselves or as an idealized version.

Vasalou and Joinson (2009) found that avatars on blogging sites were created to accurately reflect their owners' physical appearance, lifestyle, and preferences. By contrast, participants on dating and gaming sites accentuated certain aspects of their avatars to reflect the tone and perceived expectations of the context. For instance, avatars in dating were made to look more attractive while avatars in gaming were made to look more intellectual. This indicates that how users choose to present themselves via avatars depends on different communication goals and purposes (Huffaker and Calvert, 2005; Riegelsberger et al., 2006; Toma et al., 2008; Vasalou and Joinson, 2009).

Avatars are one of the six major pillars of the metaverse and avatar creation is a widespread activity nowadays. In today's context of digital worlds and the metaverse users interact through avatars. They are the digital image of a user's presence in a virtual world and their creation may, to a greater or lesser extent, reflect the actual appearance of the user. It is even possible to create avatars directly based on one's biometric data, a concept theorized as early as 1998 by M. J. Lyons and his associates in their study, "Avatar Creation using Automatic Face Recognition". Lyon and his team outline the precise procedures and processing methods required to produce an avatar almost mechanically from a human face. The procedure outlined in the aforementioned article is essentially the biometric synthesis process. Users have also observed that avatars frequently mimic the creator's features, including body type, accessories, and attire, in addition to facial features.

The metaverse is likely to become increasingly realistic, so avatars will also become highly similar to the real user. For instance, consider the level of realism that some video games have achieved. Video games such as FIFA or Call of Duty can serve as an example. This is even more evident in the case of Instant codec Avatar or FaceChain. Recently, for instance, Alibaba group lauched FaceChain, a toolchain that combines deep learning with facial recognition, allowing users to craft their digital twin and generate personalized images in various contexts (Yang et al., 2023).

Alibaba Group, the Chinese company heading this project, is subject to a regulatory framework—the Chinese one—that is significantly less restrictive than that applicable in a European context. Nevertheless, this highlights that the risk of identification (or, more broadly, the use) of biometric data through avatars is not science fiction but a present reality.

Furthermore, if an avatar can be created from biometric data, the reverse process may also be possible. This raises certain concerns when considering the significant technological advancements from 1998 to the present.

A player is immediately recognizable, even in their game representation. Excluding, for now, the numerous legal differences, this comparison is useful for understanding the level of quality and similarity that avatars could reach.

Based on the discussion so far, supported by relevant literature, it is evident that various types of information about the user can be inferred from analyzing their avatar. At times, however, the user is not even aware of providing this information to the platform or to third parties. This is because, as explained above, if data derived from avatars cannot be classified as "biometric data," they do not receive the special, stricter protections granted to this category. Consequently, data potentially derived from avatar analysis are treated as ordinary data. However, this approach leads to the handling of data that fall outside the user's control—a scenario typical of "inferred data".

3.2 Inferring data from avatars

Although the GDPR does not provide a definition, "Inferred Data" refers to information derived from the analysis or interpretation of other data, such as human characteristic data. Should these data reveal sensitive information, including data concerning health, the applicable legal framework would be the one set out in Art. 9 of the GDPR with the relevant restrictions and conditions of processing.

The first case that comes to mind is related to users with some form of disability, which can be inferred from the analysis and interpretation of certain elements, such as, for example, the appearance of the avatar or the behavior of the user in the metaverse. However, this contribution aims to bring to attention another hypothetical case. The massive use of avatars, in fact, raises several potential legal issues. As a matter of practice, during registration on a platform, users are typically asked to create their own avatar. There are no specific requirements or guidelines for avatar design, and users may choose to use an avatar that looks very different from their real-life appearance. However, in practice, most users tend to create an avatar that resembles their real-life appearance (see Vasalou and Joinson, 2009).

Furthermore, as previously discussed, some research confirms in many cases a strong resemblance of the avatar to its human creator, which makes it possible to use the results of successful avatar recognition for human recognition, and *vice versa*. This will, in turn, open a new area of virtual biometrics, or augmenting the actual biometric with results of recognition in virtual world. Early results of the research show that avatars for the most parts resemble their "owners" rather than being completely virtual creations. As the physical and the virtual worlds seem to come increasingly close to each other, the distinction between the two begins to fade, and the need arises for security systems capable of working in the contexts of interreality and augmented reality (Gavrilova and Roman, 2011).

The combination of available and inferred data is a powerful tool for the development of the Metaverse. The fact that the identity of a person may be in the reach just by looking at his/her avatar opens a variety of possibilities in terms of targeted marketing, customization of services and networking capabilities. The brokering of biometric data is a somewhat shady, but extremely profitable area, which could be increased in the Metaverse by the inference of biometric data using the avatars. The external appearance of an avatar may reveal the facial geometry of the person behind it. But also, some other connected special categories of personal data, such as race, or sensitive information such as gender (sensitivity depending on the context), or health information if the person suffers from a disease or syndrome that are externally visible.

The argument, therefore, revolves around the potential of avatars, which can be significantly different from the physical person controlling them, or can be very similar. This choice depends on the user who creates them. Especially with the passage of time and resulting technological development, avatars may become incredibly similar to the user controlling them. In such a case, it cannot be ruled out that the avatar's image may be sufficient to precisely identify the person to whom it belongs. This is especially true considering that biometric identification is possible even if the subject is identified among a small group of people. This could occur at the hands of the platform itself or even third parties. However, there is no reference to this in privacy notices, which raises transparency issues. In fact, the user may not foresee the consequences of configuring and using their avatar. Sensitive data, such as racial or ethnic origin, health data, or other physical characteristics, could be inferred from this.

3.3 legal Implications

Based on experience to date, within the main platforms of the metaverse, there is no mention of biometric data, nor is consent required under Article 9 of the GDPR. However, this is still an unexplored area, as until now there has been no digital environment capable of simultaneously involving so much personal data as in the metaverse, and it will certainly be a critical issue in the future. As the technical capabilities to process personal data grow, the legal limitations to the processing of special categories of personal data increase as well. In the legislative international landscape, more and more jurisdictions are adopting privacy laws that protect specific types of data, such as biometrics. In the case-law scenario, these limitations are also increasing in importance. The previously mentioned case Schrems v Meta Platforms Ireland is a clear exponent of these growing limitations. The processing of biometric data in the Metaverse may be further limited. The possibilities for inference, although increasingly easier to develop from a technical capability point of view, may not be a safe course of action for companies who develop their services in the virtual worlds. Publicly available data, or evident information available in the Metaverse are not necessarily legally in the reach of any controller or third party. Further efforts in compliance to demonstrate the appropriate legal basis for these operations will be necessary. The processing of these sensitive information may be very limited if the Court of Justice of the EU maintains the line marked in Schrems v Meta Platforms Ireland for other special categories of personal data.

Excluding, for the time being, the challenges related to biometric data, it is important to note that a vast amount of inferred data can be deduced and processed by the platform itself and by third parties, without an adequate basis of consent. The methodology employed in this study for comparing and analyzing privacy policies is rooted in a direct comparison of terms across platforms, adopting a European perspective on the definitions of biometric data previously outlined. The assessment framework drew methodological inspiration from Yu et al. (2023), who conducted a brilliant comparative analysis of privacy policies in virtual worlds. Regarding the impact of insufficiently transparent privacy policies on users and consumers, this work incorporated comparative insights from Bottis et al. (2019).

From the analysis of the privacy policies of the main metaverse platforms, it is shown that the platforms claim no collection of consent on the processing of biometric data. Not only that, no reference is made to biometric data, nor to other types of data that fall under the discipline of Article 9 of the GDPR. The policies examined are those of Decentraland, Roblox, Epic Games (parent company of Fortnite) and especially Ready Player Me. In the cases that have been examined, there is no apparent request for consent to process biometric data. It is worth noting that, according to the organization's privacy policy, the images associated with the avatars are not utilized for specific purposes. This may explain why there is no requirement for specific consent for the processing of biometric data. However, the crux of the matter is distinct: if the data related to the avatars were to be processed and utilized, would they be classified as "personal data" or as "special categories of personal data" as per Article 9 of the GDPR?

To be precise, even if it does not fall under Article 9 (2) (a), it could fall under Article 9 (2) (e), i.e., the case where the data are manifestly made public by the data subject. However, the disclosure would have to be more comprehensive in order to be free from criticism. Above all, it should take into account not only biometric data, but also inferred data, which instead seem to be totally ignored. This classification could have significant implications for the handling and processing of such data. Although biometric data is not being processed with the aim of identifying individuals, it seems that platforms are not adequately taking into account the data that can be inferred during user activities. While actual consent as per Article 9 GDPR may not be strictly necessary, it would be appropriate to inform users of the consequences of their avatar configurations, as the principle of transparency requires.

The possibility of inference within different platforms merits a mention. As an example of this possibility: a user that registers in a virtual world platform may want to use an already existing avatar that he created and already used in other platforms. Ready Player Me ("RPM"), for instance, offers a service to cover this gap. The use of this platform is fairly direct, since it creates an avatar based on the processing of a picture. By using automated means, the avatar is created from the picture of the person who uploaded it. This automated processing may allow the avatar to be created in a hyper-realistic appearance (including biometric data such as facial geometry). However, their privacy policy does not mention at all the processing of biometric data, which is fairly limited under art. 9 GPDR. Not only it does not inform well the users about this potential processing, but also the biometric data is shared and transferred to third parties without the explicit consent of the data subject. Their service is based on the sharing of those data, since Ready Player Me is not a virtual worlds platform as such, but an enabler for the use of other metaverse operators. If biometric data is to be shared in this way, additional information and protection is merited, regarding the legal basis for its processing, privacy by design and by default measures and transfer of the data.

The receiving platform which acts as a processor of the service of RPM may process subsequently the digital information of the avatar. But it may also exceed the original processing, becoming a controller of the operation. For instance, by inferring the identity of the user based on the processing of the facial geometry of the avatar. This possibility raises questions on the pertinence of the limitation of the inference of biometric data in the virtual worlds, particularly when the biometric data belongs to minors, or people who belong to social minorities who could easily be discriminated in the virtual worlds by their race and broad physical appearance.

In any case, for the sake of completeness, it is essential to note that Article 13 and 14 GDPR stipulate a specific obligation to provide clear information on the processing of personal data. Regardless of the source of the information, when data is obtained or not directly from the data subject. These precepts outline a detailed list of information to be disclosed to the data subject. However, in the specific case analyzed in this paper, we believe that such information alone is insufficient. Indeed, for data subjects to be genuinely aware of the consequences associated with the creation and use of their avatar, they should be informed at the very moment of creation. Only then can they make informed decisions about the types and extent of information they choose to disclose through their avatar's configuration. If this avatar is shared and used in different platforms, every service provider in the Metaverse should inform the data subject about the processing of personal data. Including the possibility of inference of biometric data by the platform itself or by other users. As it stands, this does not appear to occur on any platform. Consequently, data subjects create a virtual alter ego, often embedding many aspects of their digital identity within it, without fully understanding the ramifications. The average user, in fact, is generally unaware of the range of information that could be deduced—or more precisely, inferred—from their avatar.

Privacy policies of service providers in the Metaverse generally disregard the possibility of inference of personal data. They are not deep enough in their assessment of the possibilities of the avatars and the processing of biometric data. Inferred data generally escape the obligations imposed to protect special categories of personal data, leaving the door open for potential abuses or disregard of vulnerabilities that are better assessed for other recollection of personal data. The most basic of these is the due obligation to inform the data users of the processing. The right to be informed is the first step in the "active empowerment" of the data subjects for the control of their personal information. In this case, the users of the Metaverse. A complete information about the processing of inferred data should cover, in light of art. 13 and 14 GPDR:

- The identity and the contact details of the potential controller and, where applicable, of the controller's representative in the case of the inferred data.

- The purposes of the processing for which the inferred data may be used, such as automated identification through facial geometry.
- The potential claim of the use of art 6 (1) GDPR, with the legitimate interest claim of the potential controller or third party who could infer the data.
- The potential recipients of the inferred personal data, when communication activities may be developed.
- The existance of international personal data transfers of the inferred data, if Chapter V GDPR is of application due to the activities developed in the Metaverse.

A separated discussion is merited for the due compliance with article 6 GDPR on the infered data. The legal basis for the processing of the inferred data could be of specific controversy. The analysis of these legal basis, however, exceeds the objective of the research exposed in this text.

The bare minimum level of compliance with GDPR demands a review of the application of art. 13 and 14 GDPR, in order to assess inferred data in light of the possibilities of breach of the purpose limitation principle, the data minimization principle, and the due respect to the right to be informed of the users of these platforms. There should be a paragraph in every privacy policy of the Metaverse service providers dedicated to inform about the possibility of inference of biometric data through avatars.

4 Conclusion

The considerations in this article aim to open up a discussion on the adequacy of the current legal framework with regard to the metaverse. Until now, the most widely used technologies had not manifested the capacity to collect and process data that the metaverse is highlighting.

In the era of Web 3.0 and the imminent 4.0, an increasing number of real-world elements will be integrated into a digital context. In particular, data that were not processed until recently are now destined to be collected, processed, or otherwise subject to the attention of other entities. If the objective is to bring as many elements as possible from the real world into the virtual dimension, there is a political and legal need to protect users from possible negative consequences. For instance, avatars are still being created and used with little regard for their implications. Currently, avatars lack a unified legal definition, specific regulations, and the necessary attention from platforms and users. This article highlights just some of the many potential issues that users may encounter, and on which platforms have yet to provide clear guidance.

The primary objective of this paper is to highlight certain critical aspects of the current legal framework concerning personal data protection. Specifically, we focus on the relationship between avatars and biometric or sensitive data that can be derived from them. Our research draws upon studies showing that avatars are frequently created to resemble their respective users. Moreover, due to significant technological advancements in recent years, avatars can potentially be highly similar, even nearly identical, to the real user who controls them. With the advent and spread of artificial intelligence, then, all these problems may become more acute. Consequently, the avatar becomes a conduit for information, including sensitive data. In particular, data types that can be extracted and processed from avatars include biometric data or a wide range of other sensitive information. Additionally, a significant amount of data and information could potentially be inferred from one's digital twin. Furthermore, most users are unaware of the breadth of information that can be inferred from their avatars.

According to the current legal framework, data inferred from avatars are not categorized as biometric data and therefore do not receive the special, stringent treatment afforded to this type of data. However, it is apparent that a considerable amount of sensitive data can be inferred from the characteristics assigned by the user when configuring the avatar. This includes data such as age, physical appearance, ethnicity, stylistic preferences, and, in certain cases, even disabilities. Additionally, users are not adequately informed about the real risks associated with creating a digital alter ego, which raises critical concerns.

Through a comparative analysis of data protection regulations and privacy policies of leading platforms, it has become evident that a new interpretation of biometric data legislation is urgently needed. Moreover, the discipline will necessarily have to be coordinated with the newly-arrived AI Act, which pays particular attention to biometric data.

This situation calls for reflection and opens the discussion on two key issues. Firstly, the European (as well as American) data protection frameworks are ill-suited for this level of technological change and thus require either modification or at least a more updated interpretation. Secondly, this analysis seeks to underscore not only the risks but also the potential that digital environments present. This calls for heightened awareness among legislators and, even more crucially, among users themselves, who remain the first line of defense for their own rights.

Author contributions

GS: Conceptualization, Investigation, Methodology, Supervision, Writing – original draft, Writing – review and editing. JL: Conceptualization, Investigation, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The author declare that the payment of the publication fee has been made by University of Bologna.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Generative AI was used in the creation of this manuscript. To review some grammar errors in English.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Aljaroodi, H. M., Adam, M. T. P., Chiong, R., and Teubner, T. (2019). Avatars and embodied agents in experimental information systems research: a systematic review and conceptual framework. *Australas. J. Inf. Syst.* 23. doi:10.3127/ajis.v23i0.1841

Ball, M. (2022). *The metaverse: and how it will revolutionize everything*. Liveright Publishing Corporation, a Division of W.W. Norton and Company.

Biometrics (2021). Oxford English dictionary (OED) online edition. Available online at: https://www.oed.com/view/Entry/19233?rskey=LqDtRj&result=1#eid.

Blockchain Research Lab (2023). Avatars: shaping digital identity in the metaverse. Available online at: https://www.blockchainresearchlab.org/wp-content/uploads/2020/05/Avatars-Shaping-Digital-Identity-in-the-Metaverse-Report-March-2023-Blockchain-Research-Lab.pdf.

Bolognini, L., and Carpenelli, F. (2022). Il futuro dei dati personali nel metaverso. in *Diritto, economia e tecnologie della privacy*. doi:10.5281/zenodo.6802662

Bottis, M., Panagopoulou-Koutnatzi, F., Michailaki, A., and Nikita, M. (2019). The right to access information under the GDPR. *Int. J. Technol. Policy Law* 3 (2), 131–142. doi:10.1504/IJTPL.2019.104950

Chen, Z.-H., Lu, H.-D., and Lu, C.-H. (2019). The effects of human factors on the use of avatars in game-based learning: customization vs. Non-customization. *Int. J. Human Computer Interact.* 35, 384–394. doi:10.1080/10447318.2018.1543090

European Data Protection Board (EDPB) (2019). Guidelines 3/2019 on processing of personal data through video devices. Available online at: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201903_video_devices_en_0.pdf.

Fazlioglu, M. (2024). *US federal privacy legislation tracker introduced in the 118th congress (2023-2024)*. International Association of Privacy Professionals. Available online at: https://iapp.org/media/pdf/resource_center/us_federal_privacy_legislation_tracker.pdf.

Gavrilova, M. L., and Roman, Y. (2011). "Applying biometric principles to avatar recognition," in *Transactions on computational science XII. Lecture notes in computer science*. Editors M. L. Gavrilova, C. J. K. Tan, A. Sourin, and O. Sourina (Berlin, Heidelberg: Springer), 12–26. doi:10.1007/978-3-642-22336-5_8

Global Market Insights (2024). AI avatars Market size. Available online at: https://www.gminsights.com/industry-analysis/ai-avatars-market.

Hopkins, R. (1999). An introduction to biometrics and large-scale civilian identification. *Int. Rev. Law, Comput. Technol.* 13, 337–363. doi:10.1080/13600869955017

Huffaker, D. A., and Calvert, S. L. (2005). Gender, identity, and language use in teenage blogs. *J. Computer-Mediated Commun.* 10 (2). doi:10.1111/j.1083-6101.2005.tb00238.x

Lin, H., and Wang, H. (2014). Avatar creation in virtual worlds: behaviors and motivations. *Comput. Hum. Behav.* 34, 213–218. doi:10.1016/j.chb.2013.10.005

Malgieri, G. (2021). In/acceptable marketing and consumers' privacy expectations: four tests from EU data protection law. *J. Consumer Mark.* 40 (2), 209–223. doi:10.1108/JCM-03-2021-4571

McMahan, A. (2003). "Immersion, engagement, and presence: a method for analyzing 3-D video games," in *The video game theory reader*, 1–20.

Mordini, E., and Petrini, C. (2007). Ethical and social implications of biometric identification technology. *Ann. Dell'Istituto Super. Sanità* 43, 5–11.

Ratan, R., and Sah, Y. J. (2015). Leveling up on stereotype threat: the role of avatar customization and avatar embodiment. *Comput. Hum. Behav.* 50, 367–374. doi:10.1016/j.chb.2015.04.010

Riegelsberger, J., Counts, S. J., Farnham, S. D., and Phillips, B. C. (2006). "Sounds good to me: effects of photo and voice profiles on gaming partner choice," in *Proceedings of computer-supported cooperative work*, 159–162. doi:10.1145/1180875.1180899

Smith, M., Mann, M., and Urbas, G. (2018). *Biometrics, crime and security*. New York: Routledge.

Solove, D. J., and Schwartz, P. M. (2024). *Privacy law fundamentals*. 7th ed. IAPP. Available online at: https://iapp.org/resources/article/privacy-law-fundamentals-2/.

Sumer, B. (2022). "When do the images of biometric characteristics qualify as special categories of data under the GDPR?: a systemic approach to biometric data processing," in 2022 international conference of the biometrics special interest group (BIOSIG), 1–6.

Tale, C. (2022). *More than half of voters back a national data privacy law*. Morning Consult Pro. Available online at: https://pro.morninMoreThanHalfofVotersBackaNationalDataPrivacyLawgconsult.com/instant-intel/federal-data-privacy-legislation-polling (Accessed October 28, 2024).

Toma, C. L., Hancock, J. T., and Ellison, B. N. (2008). Separating fact from fiction: an examination of deceptive self-presentation in online dating profiles. *Personality Soc. Psychol. Bull.* 34 (8), 1023–1036. doi:10.1177/0146167208318067

Vasalou, A., and Joinson, A. N. (2009). Me, myself and I: the role of interactional context on self-presentation through avatars. *Comput. Hum. Behav.* 25 (2), 510–520. doi:10.1016/j.chb.2008.11.007

Yang, H., Zheng, M., Feng, W., Huang, H., Yu-Kun Lai, H., and Wan, P. (2023). Towards Practical Capture of High-Fidelity Relightable Avatars. In SIGGRAPH Asia 2023 Conference Papers (SA '23). New York, NY: Association for Computing Machinery. *Article* 23, 1–11. doi:10.1145/3610548.3618138

Yannopoulos, A., Andronikou, V., and Varvarigou, T. (2008). "Behavioural biometric profiling and ambient intelligence," in *Profiling the European citizen*. Editors M. Hildebrandt, and S. Gutwirth (Springer), 89.

Yee, N., Bailenson, J. N., Urbanek, M., Chang, F., and Merget, D. (2007). The unbearable likeness of being digital: the persistence of nonverbal social norms in online virtual environments. *CyberPsychology and Behav.* 10, 115–121. doi:10.1089/cpb.2006.9984

Yu, B., Liu, Y., Ren, S., Zhou, Z., and Liu, J. (2023). METAseen: analyzing network traffic and privacy policies in Web 3.0 based Metaverse. *Digital Commun. Netw.* 11, 13–25. doi:10.1016/j.dcan.2023.11.006

Zaeem, R. N., and Barber, K. S. (2020). The effect of the GDPR on privacy policies: recent progress and future promise. *ACM Trans. Manag. Inf. Syst.* 12 (2), 1–20. doi:10.1145/3389685