Privacy and Cybersecurity in the Metaverse

Gaspare Mattarella - 965461 - gaspare.mattarella@studenti.unimi.it

Matteo Biglioli - 938199 - matteo.biglioli@studenti.unimi.it

Abstract

This paper provides a birds eye view on the role of privacy and cybersecurity in the next future, with a focus on the idea of Metaverse. We start by explaining what is the Metaverse and how it will permeate our everyday life; we then address the ongoing conflict between private companies interests in big data and full information versus public interest in preserving citizens' privacy with a particular focus on the latest technological advancements such as Big Data and the Metaverse. We then discuss the state of the art of Privacy and Data Protection focusing on the latest lawsuits regarding infringements of GDPR policies and presenting a brief overview of data regulation policies in different countries. Finally we explore how privacy and data protection could evolve in the Metaverse by presenting internal risks and privacy threats that user could face in the Metaverse, from private companies and from other malicious users/bots, and by presenting different approaches to address the issues, both technological-wise and policy-wise.

1 What is the Metaverse

While being an extremely trending topic in the last months[1], it is still not that easy to find accurate and reliable information regarding what exactly the Metaverse is.

The term metaverse[2] was originally coined in 1992 by the American writer Neal Stephenson in his science fiction novel Snow Crash[3], even if the idea of virtual worlds began in the late 1970s with the birth of tabletop role-playing games like *Dungeons and Dragons*. A few years later we witnessed the birth of video-games; in this field the *virtual world* concept was firstly applied by Lucasfilm in their Habitat application (1984) and it was then immediately implemented by other companies in their products[4]. Nowadays the game-focused metaverse industry is extremely advanced, it proposes products which are so immersive that keep players engaged for hours at a time [2] and it has developed sophisticated free-to-play business models generating billions in profits [3].

While definitely the gaming world represent one of the keystones of the metaverse's evolution, there are also a lot of other different sectors in which this idea was applied: from movies[5] to social networks[6] we already grew familiar with the concept of a metaverse, we are just not used to call it in this way.

One of the best formal definitions of Meta-

[1] Mainly thanks to the 2021 Connect conference held by Meta (ex Facebook) [1]

[2] The lowercase m is actually intended, we will later explain the difference between metaverse and the Metaverse.

[3] It was used to describe a virtual shared space, similar to a virtual reality video-game, in which people can experience a second life.

[4] The most important being *Worlds Inc* and *Second Life*.

[5] The most popular are *The Matrix* and *Ready Player One*.

[6] From Facebook to Instagram we are moving towards near real-time interaction, the most recent one being Clubhouse.

verse is the one proposed by Dionisio et al. [4]: *Virtual worlds are persistent online computer-generated environments where multiple users in remote physical locations can interact in real time for the purposes of work or play.*

This phrasing allow us to grasp all the different concepts that are commonly overlooked in formulations that lean mostly towards the recreational aspect of the term. In their article Dionisio et al. [4] offer four key features that are considered the core of a viable Metaverse: Realism, Ubiquity, Interoperability and Scalability.

The first one is strictly related to the technology underling the application, and it is what make a metaverse stand out between basics *virtual-life* projects: with this term we put the focus on the interaction of the user with the other players and with the virtual world itself. Mark Zuckerberg called it a *feeling of presence* in the 2021 Connect conference [1], and he went on describing the following: *"You're going to really feel like you're there with other people. You'll see their facial expressions, their body language, maybe figure out if they're actually holding a winning hand, all the subtle ways we communicate that today's technology can't quite deliver."*.

At the second place we find ubiquity in both access and identity sense. In this feature we group all the different psychological aspects and perceptions that are the foundation of one's person identity and feeling of belonging to a world. We expect the Metaverse to be accessible from every one of our connected devices and our virtual identity to persist throughout different accesses; these two concepts make the Metaverse something more than a common video-game.

We then find the feature that make us differentiate between *a metaverse* and *the Metaverse*, that is interoperability. Broadly speaking anyone can create a metaverse; we previously stated that a lot of different metaverses, mainly gaming-focused[7], are already part of our society; but when we speak about the Metaverse what we are referring to is a completely open architecture where different entities connect to a single shared metaverse and operate based on agreed-upon standards. As we expect our virtual identity to persist, we also expect to have **one** virtual identity that can transition through all the different environments - or metaverses - that are part of the main Metaverse.

While it seems that Meta Inc. itself is moving its first steps towards this goal, allowing users to log in their Oculus[8] devices via non-Facebook account and proposing catchphrases like *"You want to know that you own your items, not a platform"*, we expect the path leading to a shared and open-source architecture to be long and tough given that all major big tech companies will fight to promote their corporate-controlled product as the Metaverse.

Lastly but not less important we find scalability. This feature, similarly to realism, is strongly linked with the technology underling the Metaverse, and refers to the ability of the application *to enable a massive number of user to occupy the Metaverse without compromising the efficiency of the system and the experience of the users* [4].

Nowadays the Metaverse industry is moving its first steps towards the general population, and it is doing that at an incredible rate. As discussed before the gaming sphere is the cutting edge, as it has already introduced experiences that a few years ago would have been inconceivable, like real-time worldwide in-game concerts performed by top artists [5].

Meta was the first company to actually develop a 360 degree plan[9] to develop their Metaverse-related product, presenting it in the October 2021 Connect conference [1], in which its CEO, Mark Zuckerberg, spoke about a *"need to bring that same imagination and commitment to building for interoperability, openness, safety and privacy as we do for all the other product aspects of the Metaverse"*. A few days later during the Ignite 2021 Conference [6] Meta was followed by Microsoft, as the company

[7] To cite some of the most known we have *Minecraft*, *Fortnite* and *World of Warcraft*

[8] Oculus Quest is a virtual reality headset developed by Oculus, a division of Meta.

[9] With an attached $10B investment for this year alone.

CEO, Satya Nadella, introduced their first Metaverse-related product while describing the metaverse as something that *"enables us to embed the computing into the real-world, and the real-world into computing"*. While we expect other big tech companies to follow this path in the next future, we must acknowledge that they are not alone in the journey: also South Korea announced a plan to develop a public metaverse [7] defined as *a virtual public service center*.

Unfortunately, in the last years, mainly because of video-games and movies, people began to grasp the concept of *virtual world*, but they did so in a way that is slightly wrong or at least extremely focused in a single aspect of the whole picture. In the popular opinion the Metaverse, commonly referred to as *virtual reality*, is either a complex 3D immersive video-game or part of a dystopian future which resemble the movie *The Matrix*, but, as shown in the 2021 Connect conference [1], it can be much more than this.

The past two years were extremely influenced by the COVID pandemic, and, as already discussed in different papers and articles, the isolation we were forced to helped us realize how important it is for our society to leverage technological innovations, based on the internet, in order to stay connected with each other, both for work and social reasons, even if we are physically apart. Companies and governments realized the impact that a *smart* way of working has on different aspects of our everyday life, such as productivity, pollution, traffic and much more. The majority of the working population agrees on the fact that the future should move towards allowing us to maintain the work-life balance that we were able to experience in the pandemic period, and the Metaverse will be the perfect tool to achieve this goal.

One of the key points that must be addressed nowadays is the approach that private companies and governments will have in building this Metaverse as a safe place, both with advanced cybersecurity and privacy-related policies. Unfortunately we know that private and public interests are not always aligned on these kind of decisions, with private companies focusing on reaching the maximum gain from every single data entry and governments concerned about the privacy of their citizens.

2 Role of private companies as privacy controllers vs Public institutions

The conflict between private companies interest in big data and full information versus public interest in preserving citizens' privacy is long time rooted in history but the relevance that this debate has gained for every single person has exponentially increased in the last decade, making it now probably the "The Biggest Public Policy Challenge of Our Time", as Prof. Omer Tene [28] affirmed.

It calls for momentous choices to be made between weighty policy concerns such as medical research, urban planning and efficient use of resources, on one hand, and individuals' rights to privacy, fairness, equality and freedom of speech, on the other hand. It requires deciding whether efforts to cure fatal disease or eviscerate terrorism are worth subjecting human individuality to omniscient surveillance and algorithmic decision-making.

Advances in data mining and analytics and the massive increase in both computing power and data storage capacity have expanded by orders of magnitude the scope of information available for businesses and government. Data are now available for analysis in raw form, escaping the confines of structured databases and enhancing researchers' abilities to identify correlations and conceive of new, unanticipated uses for existing information.

In addition, the increasing number of people, devices, and sensors that are now connected by digital networks has revolutionized the ability to generate, communicate, share, and access data. Data creates enormous value for the world economy, driving innovation, productivity, efficiency, and growth.

At the same time, the "data deluge" presents privacy concerns which could stir a regulatory

backlash dampening the data economy and stifling innovation[29]. The extraordinary societal benefits of big data—including breakthroughs in medicine, data security, and energy use must be reconciled with increased risks to individuals' privacy.
As is often the case, technological and business developments in big data analysis have far outpaced the existing legal frameworks, which date back from an era of mainframe computers, predating the Internet, mobile, and cloud computing. Now that all the biggest nations in the world seem to have updated their privacy regulations, a new unknown challenge appears at in front of us with the incoming "Metaverse".

Thus, although almost everyone would agree with the fact that "Big Data techniques will lead to significant, new, life-enhancing (even life-saving) benefits that we would be ill advised to electively forego"[32], many academics would also contest that too many commentators have too often overstated the benefits of Big Data, inflating studies and praising the merely trivial, highlighting as Big Data's touted benefits are often less significant than claimed and less necessary than assumed.

The risks that Big Data poses are several. The harvesting of large sets of personal data and the use of state of the art analytics implicate growing privacy concerns. Protecting privacy will become harder as information is multiplied and shared ever more widely among multiple parties around the world. As more information regarding individuals' health, financial, location, electricity use, and online activity percolates, concerns arise regarding profiling, tracking, discrimination, exclusion, government surveillance, and loss of control.

This has abundantly been made clear in the last few years with Facebook[10] right in the eye of the storm but the promise of even more interconnection within the whole realm of the "internet experience" would inevitably lead to an increased hardship multiplier. Just to recapitulate what scholars refer to when speaking about risks, these extraordinary perils can be rapidly summed up in:

- Incremental Effect: once data are linked to an identified individual, they become difficult to disentangle. *Narayanan & Shmatikov* explained, "Once any piece of data has been linked to a person's real identity, any association between this data and a virtual identity breaks anonymity of the latter" [30]. Paul Ohm warned that this incremental effect will lead to a "database of ruin," chewing away, byte by byte, on an individual's privacy until his or her profile is completely exposed.

- Automated Decision-Making is the second identified threat by [29] to public safety inducing users and society to be compartmentalized into pockets (or "echo chambers") of like-minded individuals.

- Predictive analysis, facilitating unlawful activity such as "redlining" or distopic ethical dilemmas à la "Minority Report", where a "PreCrime" police department apprehends "criminals" based on foreknowledge of their future misdeeds.

- Lack of access and exclusion

- The so called "Chilling Effect" which is the inhibition or discouragement of the legitimate exercise of natural and legal rights by the threat of being spied on[11].

Given both the outstanding benefits that could enhance our everyday life and the distopic alike threats that Big Data collection and their unethical use pose on us, we are able to find very different positions among academics on what to do on this regards.

Opinions and stands may vary from very conciliatory, business-friendly positions to positions much more prone to restrict big techs and call for a unilateral action of the ruler against them.

Tane & Polonetsky [29], which we can use as champions of the first above position, were proposing a legal model where the benefits

[10] Meta Platforms, Inc.

[11] The Potential Chilling Effects of Big Data

of data for organizations and researchers are shared with individuals.

“If organizations provide individuals with access to their data in usable formats, creative powers will be unleashed to provide users with applications and features building on their data for new innovative uses. In addition, transparency with respect to the logic underlying organizations’ data processing will deter unethical, sensitive data use and allay concerns about inaccurate inferences”[29].

They, for example, individuate two main problems modern privacy regulatory systems should address. Firstly an updated definition of Personally Identifiable Information (PII). Since “re-identification” techniques were proven to disrupt privacy policy landscape by undermining the faith that we have placed in anonymization and de-identification [31] the stands for limiting at the bare minimum the use/collection of PII and the enlargement of the definition of the concept of PII itself have rapdly grown up[31].

Yet, *Tane & Polonetsky* fear the perverse incentives for organizations to forgo de-identification altogether and therefore increase, not alleviate, privacy and data security risks. Not to mention about the fact that enlarging the definition of PII to all data would make the privacy framework become all but unworkable [29]. The major concern, thus, lay on the possibility that many beneficial uses of data would be severely curtailed if information, ostensibly not about individuals, comes under full remit of privacy laws based on a remote possibility of being linked to an individual at some point in time through some conceivable method, no matter how unlikely to be used [29].

Thus, they propose PII to be instead “defined based on a risk matrix taking into account the risk, intent, and potential consequences of re-identification” [29].
As the second important pillar of privacy policy, Data Minimization, grow to be impossible and antithetical to the new big data business model, they keep proposing a coherent framework that is based on a risk matrix that weighs the value of data against potential privacy risks up to the point of stating that “Where prospective data uses are highly beneficial and privacy risks minimal, the legitimacy of processing should be assumed even if individuals decline (or are not asked) to consent.”[29]

On the other side, scholars with views much more prone to restrict big techs and call for a unilateral action of the ruler against them though that on PII problem one possible conclusion is that all data should be treated as PII and subjected to the regulatory framework[32] with all the drawbacks and troubles that this would cause, as we have seen above.
Furthermore, they were much more concerned on the consent. They required firms to need informed, specific, and unambiguous consent from consumers in order to process their personal data, which requires consumers to explicitly opt into data collection as well for the limited time and quantity of data to keep in their archive.
Researchers have tried to quantify and study the impact of such a policy, finding how, although protecting the opting-out users, those kind of norms make the opt-in consumers who share their data more trackable and possibly more predictable to the firms with which they share data. If this increased trackability makes up for decreased data (resulting from opt-outs), then the firms using consumer data could also come out as winners[33].

We can definitely find tracks of thees late instances in the European Union GDPR that were evidently taken in consideration during the drafting period but only executed for the first time in 2018, years later all of the contradictions mentioned above were already clear cut. Technological and business developments in big data analysis have far outpaced the existing legal frameworks before and there is no evidence it won’t happen again with the Metaverse “revolution” we foresee ahead.

> *“Metaverse technologies like VR and AR are perhaps the most data-extractive digital sensors we’re likely*

to invite into our homes in the next decade" [12]

Are we going to be ready for that?

3 State of the art of Privacy and Data Protection

In order to be able to understand how privacy and data protection could progress and evolve in the novel metaverse, we first need to comprehend the state of the art of this field. In this section we propose a brief overview of the way in which different countries approach this matter. We mainly focus on analyzing some of the main concepts of the current European regulation[8] by presenting a few interesting trials that happened since its entry into application and then discuss how different extra-EU countries (US, China) are approaching this sector. We point out that the goal of this chapter is not to provide a complete and accurate review of all the different regulations, but to focus on the leading aspects that are still debated and for which a lot of companies have been found guilty.

3.1 General Data Protection Regulation

The current European regulation for what it concerns privacy and data protection is the General Data Protection Regulation (GDPR [8]), adopted on 14 April 2016 and enforceable since May 2018. We can immediately notice how both this document and its predecessor [11] are written with a special focus on the protection of natural persons' rights[13], and they do that by enforcing a variety of policies that aim to balance the gap of power between individuals and big companies.

While redefining its material scope to *"any information which are related to an identified or identifiable natural person"* (Art. 4), the GDPR is one of the first documents that extends the territorial scope not only to data controllers and processors in the EU, but to any entity that process personal data of data subjects who are in the Union (Art 3); this is one of the reason why this regulation is crucial also for extra-EU countries, because it affects even big-tech companies operating in EU.

Before addressing some of the main concepts, we point out that the GDPR lays down new fines to companies found in contrast with its policies; specifically *"up to 20 million euros, or in the case of an undertaking, up to 4% of their total global turnover of the preceding fiscal year, whichever is higher"* (Art. 83). The most expensive fine so far, €746 Million, has been imposed recently to Amazon by the Luxembourg DPA [12]; unfortunately due to the fact that the company is appealing the decision we do not know precisely which are the allegations in question but only that the eventuality of data breaches was denied by internal sources.

In the following subsections we separately address three main cores of the document

3.1.1 Principles relating to processing of personal data

In the second chapter of the regulation we find a list of principles [9] relating to the processing of personal data and different conditions which apply to special categories of personal data. While all of these concepts are obviously equally relevant on a theoretical level, in the past years we witnessed different trials focusing specifically on the sixth article, the one defining the *Lawfulness of processing.*

It is easy to understand the motives pushing companies towards these kind of improper behaviours: it is very tempting to exploit datasets obtained for different purposes to perform customized and aggressive marketing campaigns that can increase revenues exponentially with respect to standard advertisement. This is the case of TIM [19], which was issued a €27,8 million fine in 2020 after the GDPR received several hundreds complaints regarding aggressive promotional campaigns in the previous 3 years. We shall highlight that the unlawfulness

[12] -Marcus Carter, University of Sydeny

[13] *This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data* (GDPR Art 1.2)

of processing was not the only infringement that the Italian Data Protection Authority found when analyzing the company, in fact the fine includes different offences from excessive data retention to faulty manage of data breaches, but the main reason for which the consumers gained the DPA attention was obviously related to the unlawfulness of processing.
The other case that is worth mentioning is the one involving the Swedish DPA against Google in 2019 [18]; in this trial a particular focus was put on the legal basis for ads personalization processing, specifically on the fact that the consent collected by the company was defined *"neither 'specific' nor 'unambiguous'"*[18].

While we addressed two of the most relevant cases related to this topic, in the recent years a lot of other companies were found guilty of similar charges. As explained above it is obvious that there are incentives behind these decisions, and for that reason it is crucial that DPAs address this issues with particular strength. The Metaverse should not lay its foundations in these questionable conducts: we cannot expect people to provide their personal data to companies that do not respect privacy agreements.

3.1.2 Data Subject's rights

The third chapter of the regulation focuses on the concept of transparency and offers a list of rights of the data subject [10]. The idea of transparency is one of the key subject of the GDPR, and for this reason it is involved in the second biggest fine ever issued by a DPA (so far): €225 million [13]. The company in case, WhatsApp, has been found guilty of lack of clarity with respect to how it process data and how it shares information between the other companies of the same group[14]. In this specific trial the Ireland's Data Protection Commission (DPC) stated that the messaging service had not provided enough information about how data was collected *"in a concise, transparent, intelligible and easily accessible form, using clear and plain language"*, and, in addition to the fine, imposed a reprimand along with an order for WhatsApp to bring its processing into compliance by taking *"a range of specified remedial actions"*. A second lawsuit that can help us understand DPAs focus regarding Data Subject's rights is the one involving Vodafone Spain [14], which ended in March 2021 with a fine slightly over €8 Million. In this case the Spanish DPA, between other counts, found the telecommunications company unable to *avoid advertising actions to those citizens who had exercised their rights of opposition or erasure of their personal data*, two of the rights mentioned in the GDPR.

These trials seems similar to the ones mentioned in the section above[15] but there is a main difference: in these cases the companies are not trying to leverage personal data without consent, but they are providing the Data Subjects with information that are effectively useless due to their complexity.
As a society moving towards an everyday more entangled relationship with technology, we must provide final users both with the information they need to be able to fully understand their actions and the related consequences and the rights to control their *second life* on the growing metaverse.

3.1.3 General Obligations of Data Controller and Data Processor

Finally in chapter 4 we find a list of requirements related to both the data controller and the data processor. In this matter the GDPR approach revolves around the concept of accountability, in the sense that both entities must be able to present and explain to the DPA all the steps of the data flow, from the purpose of the processing to the processing itself, and they can be held accountable over their decisions. In addition to maintaining a record of processing

[14] Meta Platforms, Inc.

[15] And to some extent they are; as we discussed these events are not quick and concise trials but final steps of complex processes in which the DPA in question addresses multiple domains related to Data Protection.

activities, they are required[16] to carry out impact assessment prior to the processing, to designate a Data Protection Officer (DPO), to implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk and to notify eventual data breaches to the authority in a span of 72 hours.

The most popular trial regarding these policies is the one involving the Information Commissioner's Office (ICO) agains British Airways [17], in which the airline company was fined £20 Million for failing to protect personal and financial details of more than 400,000 of its customers. In this case the ICO focus has been on the lack of adequate security measures, specifically flaws where found in the absence of rigorous testing, multi-factor authentication and in bad designed organizational policies that did not restrict access to applications based on users' role. Other trials focuses more on the correct notification of data breaches, like the ones against Booking [15] and Twitter [16], and are concluded with lower but still significant fines, in these cases respectively €475 and €450 thousand. In these processes penalties are lower because the company did nothing wrong that caused the breach[17], but still delay in notifying a breach can cause other chained attacks, in which the attackers will use the data recently uncovered to try and breach other companies.

The key concept of this chapter is *accountability*, both inside a company and in the collaboration with the others, as a mean to avoid and, in the worst cases, promptly stop cyber threats.

3.2 Regulation in extra-EU Countries

In this second section we will briefly present how extra-EU contries are approaching the Data Protection sector.
The reason why we started our discussion by going through the GDPR is that other countries are actually taking the European regulation as a model to follow, and they are therefore developing policies similar to the ones we already addressed.
The first example of this is found in the US, specifically the state of California which is currently in the lead for what it concerns Data Protection thanks to its regulation, the California Consumer Privacy Act (CCPA) [21]. Unfortunately at this time the US framework is still definitely fragmented, both due to the co-presence of federal and state laws [22], which sometimes even overlap each other, and due to the kind of approach of the policies, which tend to address specific use-cases instead of proposing a broader picture on the subject. While there have been several attempts to unify the different regulations, in order to provide an environment that can improve the experience of both consumers and businesses; the government seems to look at the cybersecurity sector as another line of defense against terrorism, as indicated by the $1.9 billion dollar investment in national cybersecurity contained in the Infrastructure Investment and Jobs Act (Infrastracture Bill) [23] recently passed by the Congress.

A second country deeply concerned with cybersecurity is China, which recently released, on 12 July 2021, its three-year action plan for the high-quality development of the Cybersecurity Industry 2021-2023 aiming to create a nearly $39 billion market [24]. However, in this case, the People's Republic of China managed to provide a unified regulation officially adopted in august 2021, the Personal Information Protection Law (PIPL) [25]. From the translated text we can easily gather the resemblance with GDPR, which is seen in concepts like *"The principles of openness and transparency shall be observed in the handling of personal information"* or *"Obtaining individuals' consent"*.

Other countries have recently developed or improved their laws with regards to data protection with the aim of keeping up to the pace of innovation in the sector of Big Data and AI, which processed every day more information with as-

[16] In specified cases.

[17] We know from experience that nothing is impenetrable in the cybersecurity sector.

tonishing granularity.
There are also different entities that are keeping track of the state of these laws worldwide, like the United Nations Conference on Trade and Development [26] or DLA Piper, a multinational law firm [27].

4 The next big challenge: regulating the Metaverse

In this last chapter of the essay we are going to explore the internal risks and the privacy threats that users could face from other malevouls users/bots and how they could be tackled. Furthermore, we are going to explore which kind of threats the Metaverse itself could instead represent for users and how they could protect themselves from it.
Firstly, in the light of what we have already seen in the previous chapters, we are going to analyze how and why the next Metaverse “builders” will take care of public institutions privacy concerns. Secondly, we are going to explore some practical techniques those risks could be avoided with.

Up to now we have discussed about the privacy concerns that big companies represent for users around the world but, with the enlargement of the *Web 3.0* and the Metaverse becoming a “reality”, most of the threats could also come from other malevouls users or bots inside the Metaverse itself. As there’s already evidence this will be a serious matter [34], it will also be one of the first problem to be addressed by the “builders” if they do not want to incur in heavy sanctions.

We will begin exploring this very threats as they have been deeply studied through first Metaverse attempts such as Second Life[18] that can represents a good proxy for the next generations of Metaverses.

The study by *Ronald E. Leenes* [34] shows that players do mainly behave in the Metaverse, such as Second Life, similarly to real life and therefore their privacy and security behaviours are similar to the real ones if not enhanced. *”People are curious and nosy in Second Life as they are in real life”.* Whenever “Residents” - that is how users are called in Second Life - are interacting, by text or voice chat, there will also be Residents around that can listen in on the conversation and based on what they hear or see conclude that they want to know more. The knowledge that can be acquired by listening in on conversations will often be confined to inworld knowledge. Judging from users’ answers to the researchers, most of the people use Second Life as a virtual market square where conversations will cross the SecondLife/RealLife border. Many people are (also) interested in other Resident’s real lives. The distinction between Second Life and real life easily blurs, hence information about a Resident’s real life can often also be obtained easily.
It also has to be taken in consideration that in the Metaverse many other form of non-Real-life eavesdropping could be possible, from AI-based bots trained to retrieve information to hacking threats like the illegal “spying bugs” that infested Second Life, potentially making the Metaverse a jungle for privacy concerned users. In these virtual scenarios, the use of deep-fakes and alternate representations can have a direct effect on users’ behaviours. In the Metaverse, the generated virtual worlds can open potential threats to privacy more substantial than in the real world. For example, ‘deep-fakes’ can have more influence in users’ privacy behaviours. Users can have trouble differentiating authentic virtual subjects/objects from deep-fakes or alternate representations aiming to ‘trick’ users. The attackers can use these techniques to create a sense of urgency, fear, or other emotions leading users to reveal personal information.

As pointed out in this paper from *Lee, Tristan et al.*[35], “the attacker can create an avatar that looks like a friend of the victim to extract some personal information from the latter. In other cases, the victim’s security can be at

[18] Released in 2003, Second Life is an application that allows people to create an avatar for themselves and have a second life in an online virtual world, or Metaverse.

stake, such as physically (in the virtual world) assaulting the victim."

Other advanced techniques such as dark patterns can be used to influence users into unwanted or unaware decisions by using prior logged observations in the Metaverse.
Always in *Lee, Tristan et al.*[35], the researchers highlight how *"for example, the attacker can know what the users like to buy in the metaverse, and he/she will design a similar virtual product that the user will buy without noticing it is not the original product the user wanted. Moreover, machine learning techniques can enable a new way of chatbots and gamebots in the metaverse. These bots will use the prior inferred users' traits (e.g., personality) to create nudged social interactions in the metaverse."*

A common solution to overcome these kind of privacy and security threats, pointed out in various papers[19], could be the use of multiple avatars and privacy copies in the Metaverse. The first technique focuses on creating different avatars with different behaviours and freedom according to users' preferences. These avatars can be placed in the Metaverse to confuse attackers as they will not know which avatar is the actual user.

The second approach creates temporary and private copies of a portion of the Metaverse (e.g., a room) where attackers can not eavesdrop on the users. In the case the private portion use resources from the main fabric, the Metaverse API should address the merge accordingly from the private copy to the main fabric of the Metaverse. "Techniques that address the parallel use of items in the Metaverse should be implemented to avoid inconsistencies and degradation of the user experience (e.g., the disappearance of items in the main fabric because they are being used in a private copy)"[35]. Finally, following the creation of privacy copies, the users can also be allowed to create invisible copies of their avatar so they can interact in the Metaverse without being monitored. Although very practical, there's already a great deal of literature trying to address those kind of problems proposing entire frameworks of "Privacy Plans" [36].

Probably more relevant or, at least, definitely more popular is the user's privacy concern over the threat the company itself or the Metaverse "builders" represent for the user. Nowadays, as virtual reality (VR) applications increase in popularity and fidelity they also threaten to erode our privacy in new ways ranging from knowing how we physically move around to the patterns of our neural activities. Internet-connected devices such as wearables allow monitoring and collect users' information. This information can be interpreted in multiple ways. In most situations, such as in smart homes, we are not even aware of such ubiquitous and continuous recordings, and hence, our privacy can be at risk in ways we cannot foresee. These devices can collect several types of data: personal information (e.g., physical, cultural, economic), users' behaviour (e.g., habits, choices), and communications (e.g., metadata related to personal communications). In many situations, users accept the benefits in comparison with the possible privacy and security risks but, while all these biometric data can render more immersive experiences, they also open up to new privacy threats to users.

Let's begin defining the kind of Metaverse that we could face in the next years. There are three possibilities:

1. **Walled Garden**: In a 'walled garden' environment, the creator of that environment sets the policies by which the service is governed. Your data doesn't leave that creator's Metaverse, and it is not interoperable with other metaverses unless connections are enabled on the creator's terms. This is most akin to subscription or account-based digital publishers today such as Facebook, Amazon, and Twitter.
2. **Open House**: An 'open house' concept

[19] *(Lee, Tristan et al., 2021), (Falchuk, Loeb & Neff, 2018)*

of the metaverse is the idea that once the technology is more commoditised, anyone could build their own metaverse and connect it loosely to those of other creators. This model is most akin to the 'open web' notion of digital publishing where access is not restricted, standards for publication are minimal if they exist at all, and data can move freely between metaverses.

3. **Data Portability Hybrid**: The 'hybrid' option allows for creators to set their own governance policies, however key data privacy, community, and security standards are governed by the state (as the least bad governance option). The chief standard among these would be to require firms hosting metaverses to enable data portability and interoperability between other metaverses so as to ensure consumers are not "locked in" to one provider. This forces providers to compete on quality and services rather than relying on high switching costs to keep users in their networks.

Given the increasing request for data portability and the current *"oligopolistic"* structure of the tech world, the hybrid model, among those, seems to be the way to go.

It also represents an opportunity for the state to provide more effective data privacy governance than we experienced at the onset of online digital networks. The hybrid model still benefits from the privacy and security infrastructure expertise of what we assume will be the major metaverse firms like Facebook, Microsoft, Amazon and Apple — But, it takes away the potential for total monopoly power over user interactions in the metaverse. Now, despite the novel potentials which could be enabled by the metaverse ecosystem, it will need to address the issue of potential privacy leakage in the earlier stage when the ecosystem is still taking its shape, rather than waiting for future when the problem is so entrenched in the ecosystem that any solution to address privacy concerns would require redesign from scratch. An example of this issue is the third-party cookies based advertisement ecosystem, where the initial focus was to design for providing utilities. The entire revenue model was based on cookies which keep track of users in order to provide personalised advertisements, and it was too late to consider privacy aspects. "Eventually, they were enforced by privacy regulations like GDPR, and the final nail to the coffin came from Google's decision to eliminate third-party cookies from Chrome by 2022, which have virtually killed the third-party cookies based advertisement ecosystem." we read on [35].

They continue, "Also, we have some early signs of how society might react to the ubiquitous presence of technologies that would enable the metaverse from the public outcry against the Google Glass, when their concerns (or perceptions) are not taken into account" [35]. We have also witnessed the privacy "paradox" of users not paying attention to how their public data are being used by other parties, but showing very strong negative reactions when the difference between the actual use of their data and the perceived use of data become explicit and too contrast. For example, many people shared their data on Facebook willingly. Still, the Facebook and Cambridge Analytica Data scandal triggered a public outcry to the extent that Facebook was summoned by the U.S. Congress and the U.K. Parliament to hearings, and Cambridge Analytica went bankrupt soon after. Yet, on the other side, there is also a growing concern of overtrust. "Users tend to trust products from big brands far too easily, and rightly so, since human users have often relied on using reputation as predominant metric to decide whether to trust a product/service from the given brand" [35]. However, in the current data-driven economy where user's information are a commodity, big brands have been repeatedly reported to engage in practices aimed to learn about the user as much as possible[38], [37]. One other important aspect that should be faced as soon as possible has been recently showcased to the general public thanks to the former Facebook employee Francis Haugen bringing tons of papers to American newspapers [39] and testifying before the UK and the UE Parliaments[40]. Among all, Francis Haugen urged the lawmakers to intervene in child and young teen protection. As she claimed clearly

[39] the “old generation” social media platforms are loosing the battle over the kids and young teen because they already prefer hanging out in a Metaverse-like videogame experience. We mentioned in chapter 1 the success of world wide live concert inside Fortnite[20] and how new generations use videogames spaces to hang out rather than to actually play. As a demonstration of it, Fortnite released just few days ago ”Party Worlds”[41], a product/feature that is purely social experience made for the Metaverse. The focus is on friendship, not combat. Those generations will obviously represents the big target of the imminent Metaverse as well as every new generation to come. They are traditionally less aware of the risks involved in the processing of their data. From a practical standpoint, it is often difficult to ascertain whether a user is a child and, for instance, valid parental consent has been given. Therefore, since minors constitute a wide portion of increasingly sophisticated and tech-savvy XR users, handling their sensitive information will become the next big challenge for both privacy concerned lawmakers and big companies.

5 Conclusions

The goal of this paper was to provide a general overview of different contemporary privacy and data protection concepts, and to discuss how they could evolve in order to meet the requirements needed to build this new environment referred to as the Metaverse.
First of all we clarified the concept of Metaverse: we understood, using a brief historical overview, how this term was born, what does it mean in today’s technological framework, which are the key features by which we can define a viable Metaverse and, most importantly, that this *virtual world* concept is definitely something more than just a fancy 3D reinterpretation of The Matrix.

We then addressed concerns regarding the conflictual relationships between public and private interests, presenting an excursus on scholars positions taking one side or the other, pushing for more ”liberalization” of data usage to ensure innovation and growth or pushing for more strict regulations in order to avoid eventual dystopian scenarios. We therefore elaborated on this last position, identifying major risks and threats. Therefore we retrace the framework holes that lawmakers needed to address in order to be comply with these risks and some possible solutions proposed by academics.

We then presented the State Of The Art of data regulation, focusing on the main trials involving the GDPR, in order to better understand the real-world implication of its policies, and briefly going over the approach that two other major countries (US and China) are using to address this matter. The aim of this chapter was to understand which are the key concepts that Data Protection Authorities are enforcing nowadays and how much *strength* is needed in order to keep the correct balance between private companies interests and the public right to privacy.

Finally we explored how privacy and data protection could evolve in the Metaverse exploring studies on proto-versions of the Metaverse such as *Second Life* identifying several possible threats and possible solutions. Thus, we tried to delineate conformation and characteristics of the next Metaverse, providing insights on some of the next challenges that both Metaverse builders and lawmakers will probably have to deal with in the near future.

From the information we gathered and presented in this paper, we can state that the Data Regulation world is still not ready to a big step such as the Metaverse. While some countries, like Europe, have a comprehensive, clear and unified regulation, others still do not manage to properly address nowadays privacy infringements.

History has taught us that bureaucracy did never manage to be a step ahead to technology, and every new technological breakthrough is

[20] Ariana Grande’s Fortnite Concert Opens Up the Metaverse, Adrian Pennington

followed by a period where government are confused and do not know how to approach the new subject[21].

We hope that, since the Metaverse is definitely here to stay and will shortly begin to permeate our everyday activities, this time the bureaucracy will move faster, providing big-tech companies with a legal framework on which they could safely build without the worry of having to go through a complete (expensive) remodeling.

[21] The last example of this is given by the birth of cryptocurrencies, that are still not properly regulated in different countries.

References

[1] https://www.youtube.com/watch?v=Uvufun6xer8t=4sab_channel=Meta.

[2] https://www.statista.com/statistics/882113/time-spent-playing-fortnite/.

[3] https://www.forbes.com/sites/mattperez/2020/01/06/fortnite-earned-an-estimated-18-billion-in-2019-leading-all-free-to-play-games/?sh=78b3d35f658f.

[4] Dionisio et al. (2013) *3D Virtual Worlds and the Metaverse: Current Status and Future Possibilities*.

[5] https://www.theguardian.com/games/2019/feb/03/marshmello-fortnite-in-game-concert-edm-producer

[6] https://www.youtube.com/watch?v=PraEcNDGSqYab_channel=MicrosoftIgnite.

[7] http://english.seoul.go.kr/seoul-first-local-govt-to-start-new-concept-public-service-with-metaverse-platform.

[8] https://gdpr-info.eu/.

[9] https://gdpr-info.eu/art-5-gdpr/.

[10] https://gdpr-info.eu/chapter-3/.

[11] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046from=EN.

[12] https://www.bbc.com/news/business-58024116.

[13] https://www.bbc.com/news/technology-59348921.

[14] https://edpb.europa.eu/news/national-news/2021/spanish-dpa-fines-vodafone-spain-more-8-million-euros_en.

[15] https://edpb.europa.eu/news/national-news/2021/dutch-dpa-fines-bookingcom-delay-reporting-data-breach-0_en.

[16] https://www.bbc.com/news/technology-55317207.

[17] https://ico.org.uk/action-weve-taken/enforcement/british-airways/.

[18] https://www.cnil.fr/en/cnils-restricted-committee-imposes-financial-penalty-50-million-euros-against-google-llc.

[19] https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9256486.

[20] https://gdpr-info.eu/art-6-gdpr/.

[21] https://oag.ca.gov/privacy/ccpa.

[22] https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/.

[23] https://www.congress.gov/bill/117th-congress/house-bill/3684/text.

[24] https://cset.georgetown.edu/wp-content/uploads/t0381_cyber_3_year_plan_draft_EN.pdf.

[25] https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021/.

[26] https://unctad.org/page/data-protection-and-privacy-legislation-worldwide.

[27] https://www.dlapiperdataprotection.com/.

[28] https://iapp.org/news/a/privacy-and-big-data-making-ends-meet/.

[29] Omer Tene and Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics,*
11 Nw. J. Tech. Intell. Prop. 239 (2013). .

[30] Arvind Narayanan Vitaly Shmatikov, *Robust De-anonymization of Large Sparse Datasets,*
2008 IEEE SYMP. ON SECURITY PRIVACY 111.

[31] Paul Ohm, 2013, *The Underwhelming Benefits of Big Data,*
161 University of Pennsylvania
Law Review Online 339
(responding to Paul M. Schwartz, Information Privacy in the Cloud)

[32] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,*
57 UCLA L. REV. 1701, 1748 (2010) .

[33] *The Effect of Privacy Regulation on the Data Industry: Empirical Evidence from GDPR,*
Guy Aridor, Yeon-Koo Che, and Tobias Salz
NBER Working Paper No. 26900
March 2020, Revised May 2021
JEL No. K24,L0,L5,L81.

[34] Leenes, R.E. (2007). *Privacy in the Metaverse - Regulating a complex social construct in a Virtual World. FIDIS.*

[35] Lee, Lik-Hang Braud, Tristan Zhou, Pengyuan Wang, Lin Xu, Dianlei Lin, Zijun Kumar, Abhishek Bermejo, Carlos Hui, Pan. (2021). *All One Needs to Know about Metaverse: A Complete Survey on Technological Singularity, Virtual Ecosystem, and Research Agenda.*
10.13140/RG.2.2.11200.05124/8.

[36] B. Falchuk, S. Loeb and R. Neff, *"The Social Metaverse: Battle for Privacy,"*
in IEEE Technology and Society Magazine, vol. 37, no. 2, pp. 52-61, June 2018,
doi: 10.1109/MTS.2018.2826060.

[37] Abhishek Kumar, Tristan Braud, Young D. Kwon, and Pan Hui, Aquilis, 2020 *Using contextual integrity for privacy protection on mobile devices.*
Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.,
4(4), December 2020.

[38] Anthony Cuthbertson. *Google admits giving hundreds of firms to your gmail inbox. The Independent, 2018.*

[39] https://www.nytimes.com/2021/10/24/business/media/facebook-leak-frances-haugen.html

[40] https://www.nytimes.com/2021/10/24/business/media/facebook-leak-frances-haugen.html

[41] Fortnite's new 'Party Worlds' put the focus firmly on socializing