Telefónica

Social and ethical challenges of the metaverse

Opening the debate

Richard Benjamins | Chief AI & Data Strategist Telefónica
Yaiza Rubio | Chief Metaverse Officer Telefónica
Chema Alonso | Chief Digital Officer Telefónica

1. Abstract 3
2. Introduction 3
3. How do we / should we define the metaverse 4
3.1 Key ingredients of the metaverse 5
4. A pragmatic vision of the metaverse 6
4.1 Key evolutions 6
4.2 Different roles companies can play in the metaverse 7
5. Societal and ethical challenges of the metaverse 8
5.1 Learn from the challenges associated to artificial intelligence 8
5.1.1 Direct challenges 8
5.1.2 Indirect challenges 11
5.2 New challenges 12
5.2.1 Health challenges, including mental ones 12
5.2.2 Privacy challenges 13
5.2.3 Liberty challenges 13
5.2.4 Dual world challenges 14
5.2.5 Equality challenges 14
5.3 Challenges stemming from a deliberate malicious use of the metaverse 14
5.4 Approaches to tackle the challenges and mitigate social and ethical risks 15
6. Conclusion 16
6.1 Towards values and ethical principles of the metaverse 17
7. Acknowledgements 18

1. Abstract

The metaverse is attracting a huge amount of attention as the next version of the internet, and companies are starting to explore the many new business opportunities it provides. However, based on experience with artificial intelligence, we know that there are also potential negative ethical and social consequences of the massive use of technology that must be dealt with. In this white paper, we briefly describe what the metaverse is, what technologies are part of its ecosystem, and where it comes from. We then focus on the potential societal and ethical risks of the metaverse. We argue that companies that are implementing the responsible use of AI are well prepared for the social and ethical risks of the metaverse. Not because they know the future, but because they have the right governance and culture in place to deal with such risks.

2. Introduction

Who hasn't heard of the Metaverse? The future Internet with an immersive experience in 3-D where we live in virtual worlds through avatars. A digital world completely integrated with our physical world. There are many discussions and opinions about what the metaverse is, or better, will be, or what apps it will provide. And like Internet 30 years ago, nobody knows how important it will be, but most agree that it is something to watch carefully. Think about a new way of teleworking where you can virtually sit down with colleagues in the same room and have physical-like conversations and interactions. Or about new marketing experiences like driving a car that you might want to buy from your home, rapidly choosing between different alternatives. Or psychological treatments to help cure people from phobias. And new, immersive ways of education or instructions training where students can “live” in the Middle Ages, and technical engineers can practise expensive maintenance operations without risks. The metaverse might be as big as the Internet with huge, though mostly still unknown, opportunities. There is even already a standards organisation (the Metaverse Standards Forum) like the one for the web (the Worldwide Web Consortium, W3C).

The opportunities of the metaverse abound, and while we believe that the potential benefits will, by far, outweigh the risks, it is also important to reflect, in advance, on those potential risks, with the goal to mitigate them before they actually happen. The objective of this white paper is to review the *ethical and social risks* that might be associated to the metaverse, and give guidance on how organizations can prepare for them. There are other risks, including, among others, legal, tax, crime, and IP that fall outside the scope of this white paper, though we might allude to them in case they have a strong societal aspect.

The metaverse is only just beginning and this gives us a unique opportunity to build a metaverse that we want to live in or with. It will have profound impact on our lives and therefore it is very important -from the start- to think about potential consequences and act accordingly. This will help moving away from the current "break, apologise and fix" approach to a more proactive one.

3. How do we / should we define the metaverse

Many people associate the metaverse with a 3D immersive experience of playing video games. And while this is definitely included, it is not just that. It will be much more. Just as it was hard to define the Internet 30 years ago, it is also hard to define in advance what the metaverse will be. Parisi (2021), in his much-referenced post[1], discusses seven rules of the metaverse, that help define what the metaverse is and, also, what it is not:

- There is only one metaverse.
- The metaverse is open.
- The metaverse is for everyone.
- No one controls the metaverse.
- The metaverse is a network.
- The metaverse is hardware independent.
- The metaverse is the Internet.

This definition of Analysis Group[2] is a good attempt to define the metaverse in a descriptive way. *"The metaverse as conceived today is considered by many as a "successor" of the Internet. While there is no agreed upon definition of the metaverse, one way to think about it is as an expansive network of digital spaces, including immersive 3D experiences in augmented, virtual, and mixed reality, that are interconnected and interoperable so you can easily move between them, and in which you can create and explore with other people who aren't in the same physical space as you. Some have referred to the metaverse as an "embodied internet" in which individuals will feel as if they are actually "present" in experiences and not simply looking at experiences through their screens. This means that interacting with the Internet (and the devices that provide access to the Internet) has the potential to be much more natural, incorporating modes of communication that include gesture and voice, such that individuals are not limited to typing or tapping. In addition, the metaverse is envisioned to be able to host almost all the activities we currently take*

[1] https://medium.com/meta-verses/the-seven-rules-of-the-metaverse-7d4e06fa864c
[2] https://www.analysisgroup.com/globalassets/insights/publishing/2022-the-potential-global-economic-impact-of-the-metaverse.pdf

part in (e.g., socializing, work, learning, entertainment, shopping, content creation, etc.) and make new types of activities possible as well."

Some additional examples of applications in the metaverse described elsewhere: a space where you'll be able to go shopping, play games, meet friends, attend concerts, work, and generally build a virtual life; revolutionizing the way we work, with ultrarealistic avatars and virtual conference rooms full of international staff members; similar to traditional property and real estate, virtual landowners can earn rent fees through predetermined terms that are negotiated with renters and enforced by smart contracts; royalties that creators can earn when NFTs are sold or resold on secondary markets; decentralized autonomous organizations (DAOs), representing a pool of people acting together through well-defined, encoded rules. DAOs may act as crowdsourced community moneylenders, creating the next generation of decentralized financial services; transforming economic sectors such as education, health care, manufacturing, job training, communications, entertainment, and retail.

3.1 Key ingredients of the metaverse

The metaverse is and will be enabled by major innovations in hardware, human-computer interface, network infrastructure, creator tools and digital economies, including, but not limited to, virtual and augmented reality, artificial intelligence, data, holographic displays, IoT, fibre deployment, 5G stand-alone, edge computing, blockchain, smart contracts, tokenomics/NFTs, IFPS, cybersecurity, and "softwarization" of networks[3].

The metaverse, like any society, needs a functional economy. In a virtual society, the economy depends on verification of digital properties, such as one's metaverse home, car, farm, books, clothing, and furniture. To flourish, it also needs the ability to travel and trade freely between realms of the metaverse that might have different laws and rules[4]. Non-fungible tokens (NFTs) – records of digital ownership stored in a blockchain – will be an essential part of the metaverse economy, by enabling verification of possessions, property and even identity. Since each NFT is secured by a cryptographic key that cannot be deleted, copied, or destroyed, it enables the robust, decentralised verification – of one's virtual identity and digital possessions – necessary for the metaverse society to succeed and interact with other metaverse societies. NFTs, thus, enable the property transactions that drive the metaverse.

[3] https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/telecoms-future-in-the-web3-era-jose-maria-alvarez-pallete-lopez
[4] https://www.ft.com/partnercontent/crypto-com/nfts-the-metaverse-economy.html

Another key ingredient of the metaverse will be data, given the expected increased virtual activity and interactivity. Data will be able to float between different realms of the metaverse rather than sitting in company silos as in web2.0. Ownership of user data will transfer from companies (back) to users who sovereignly decide what will happen with their data (exploit, delete, port, access, etc.) in a privacy-preserving way. Artificial intelligence and machine learning will be used massively to operate and steer the metaverse while at the same time optimising the user experience; even more than today is the case.

4. A pragmatic vision of the metaverse

Because nobody knows how the metaverse will be, in order to move forward, being pragmatic might help: rather than waiting for what the future will be, it is better help creating it. We believe that the immersive 3D experience is only the tip of the iceberg of the metaverse. One way to understand the genesis of the metaverse, is to view it as the culmination of several major evolutions that have taken, and are still taking, place in the past decades in, respectively, *user interface*, *programming*, *connectivity*, and *internet economics*.

4.1 Key evolutions

User interface. Over the past 20 to 30 years the user interface paradigm has gradually shifted from people adapting to technology to technology adapting to people. It evolved from simple to richly-formatted text to graphical interfaces using windows, to browsers and apps, and to voice using cognitive interfaces. 3D interfaces are just the next evolution of the same concept exploiting new technologies (VR, AR, MR) and capabilities (e.g., HPC, cloud and high-speed connectivity). In the future, in addition to sight and hearing, the user interface may also include elements of the other three senses such as touch (to some extent already available through haptic interfaces), taste and smell. Would you imagine “printing” the smell of a forest or the taste of a cafe latte?

Programming (coding) has moved from simple procedural languages to complex rules and object-oriented approaches where people write clear instructions (algorithms) to “tell” machines what to do. The capabilities of visual programming (no/low code) have also increased significantly in the past few years. And today it is artificial intelligence / machine learning that “write” those machine instructions, based on processing huge amounts of data, for carrying out complex tasks. Some large AI models are even capable of converting voice instructions into formal programming code such as Copilot or GPT-3.

Connectivity has also experienced major evolutions, starting with digital data connections that connected to hosts, and moving to the client-server architecture over LANs. Then the internet came with its three-tiered architecture of ISPs. The dot.com era was enabled by the increasing availability of DSL and fiber broadband, and Web2.0 saw the evolution in mobile connectivity from 2G to 3G and 4G. Web3, that will underly the metaverse, will see another evolution including Wifi6, 5GSA, FTTR, edge computing, network slicing, SDN, and NaaS (Network as a Service).

The **economic model** of the first web (web 1.0) was based on publishers publishing content and consumers consuming content. The economic driver behind everything was paid content or advertising. Web 2.0 was (and is) the web of user-generated content with social media being the most popular phenomenon. But also other services such as online video and ecommerce have skyrocketed. Activity is concentrated in a few players (e.g., FANG[5]) dominating the market in a monopoly-like way. While users significantly contribute to the success of those companies, they do not participate fully in the benefits. Moreover, users' personal data is exploited at massive scale to improve services and advertising revenues. Web3 promises to change this model by giving users more benefits of the ecosystem in which they take part. Apart from being consumers of web3 services, users can become owners and decision makers. Notions of tokenomics and decentralized autonomous organizations (DAO) are born.

Together, those four evolutions naturally lead to what is called the metaverse: a new internet with improved experience and to possibility to blur the power and economic borders between companies and users.

4.2 Different roles companies can play in the metaverse

As with any new technology, companies must decide what role(s) they want to play in the metaverse. They can try to build major virtual worlds and attract massive amounts of users (probably the focus of companies like Meta and Microsoft). They can build technology components of the metaverse (e.g., start-ups for virtual or augmented reality, NFTs, Artificial Intelligence, or blockchain). They can create enablers for other businesses to take advantage of the metaverse (e.g., high-speed connectivity for families and hospitalized patients to meet realistically). They can use the metaverse as a new channel to interact with stakeholders, including customers such as marketing & sales experiences (see how clothes look on you, how new furniture fits your home, drive that new car, etc., but also customer aftersales like complaints and repair). Applying the notion of tokenomics, companies can involve their customers and

[5] FANG: Facebook, Amazon, Netflix, Google.

stakeholders in new ways in their business providing a more balanced distribution of benefits.

In this way, companies can actively take part in the development of the metaverse - each in their own way- learning much during the process, rather than adopting a wait and see attitude.

5. Societal and ethical challenges of the metaverse

There are many different types of challenges that the metaverse faces, including technical, legal, business, tax, value chain, security, user experience and economic challenges. Moreover, risks can stem from deliberate malicious -usually forbidden- actions, or from unintended consequences of supposedly innocent actions. The focus of this white paper is mostly on the societal and ethical challenges of the metaverse stemming from unintended consequences of innocuous actions, but it also introduces some challenges from deliberate malicious use of the metaverse.

5.1 Learn from the challenges associated to artificial intelligence

There are few ICT technologies that have provoked more discussions on ethical and societal challenges than artificial intelligence. Since many of the risks associated to AI might even be more present in the metaverse, it is a good place to start. We distinguish between *direct* and *indirect* societal and ethical challenges. “Direct” refers to attributable and immediate consequences of a certain use of AI or the metaverse, such as undesired discrimination. “Indirect” refers to consequences that are not obvious nor foreseen at the start.

5.1.1 Direct challenges

There are two types of direct challenges of AI, and therefore also of the metaverse: those that companies can handle individually, and those that require government involvement.

Challenges to be dealt with individual organizations

Challenges to be dealt with by individual organizations include bias and undesired discrimination, explainability and black-box algorithms, human intervention and proper system autonomy, privacy, security and safety, and the carbon footprint of AI algorithms. If an organization takes measures regarding those challenges, it might be possible to prevent or mitigate them. But if those challenges are not even considered, then the likelihood of a negative impact occurring will be higher. Following, we will briefly explain each of those challenges and how they relate to the metaverse.

Bias and undesired discrimination. Hundreds of articles have been written about this topic where artificial intelligence systems, based on bias, might lead to undesired discrimination of vulnerable groups. Black people treated differently than white people in the American justice system[6], women receiving smaller loans from banks[7] and getting hired less by companies[8] just because of their gender, etc. The metaverse will be full of applications that use artificial intelligence for prediction and classification, and therefore this challenge is also a challenge for the metaverse.

Explainability and black-box algorithms. AI systems can take decisions without human intervention and the metaverse is expected to be full of such autonomous decisions. Currently, deep learning algorithms provide the best results, but their inner workings are very hard to understand for people. If decisions have a significant impact on people's lives, then we need to understand how the algorithm comes to a certain conclusion or decision; we need white-box algorithms.

Human intervention and proper system autonomy. When decisions have an impact on people's life, we may not want to leave everything to a machine. For example, we wouldn't want to diagnose a patient with a serious disease without the involvement of a medical professional especially if this implies intrusive treatments. Again, many AI algorithms will run in the metaverse, and it is important to take explicit decisions on how much autonomy we assign to an algorithm.

Privacy. Privacy is a challenge for any digital system accessible over the Internet. Artificial intelligence has increased this challenge because data, which includes personal data, is the fuel of AI. Some business models exploit artificial intelligence to analyze enormous amounts of personal data, sometimes not in a transparent or even in an illegal way. In the metaverse, this challenge is expected to exacerbate because of the increasing opportunity for registering any kind of digital and virtual interaction involving personal data.

Security and safety. Artificial intelligence algorithms are never 100% correct. It is therefore important to assess in advance the potential security and safety damages when algorithms make errors, and to act accordingly to prevent or mitigate them. Deep learning AI systems are vulnerable to so-called adversarial attacks, and hackers managed to fool Tesla's Autopilot system to confuse a stop sign for a speed limit sign[9]. In addition to the AI challenge, the metaverse introduces a whole new set of other security and safety risks. Think for example about people moving vigorously their arms and legs while wearing 3D glasses.

---

[6] https://www.propublica.org/datastore/dataset/compas-recidivism-risk-score-data-and-analysis
[7] https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html
[8] https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G
[9] https://medium.com/self-driving-cars/adversarial-traffic-signs-fd16b7171906

Carbon footprint of algorithms. A more recent challenge that has been identified for AI is related to the energy consumption of very large (language) models. Models like GPT-3 consume millions of dollars in electricity and therefore have a significant carbon footprint. This has led to the term “green AI” where electricity consumption consideration is an integral part of the algorithm’s design. NFTs are based on blockchain, which uses consensus algorithms for adding new blocks to the chain. We all know from Bitcoin that some of those algorithms incentivize the consumption of electricity, in particular the “proof of work” consensus algorithm: the more “mining” work is performed, the higher the likelihood of being compensated. Another type of consensus algorithm called “proof of stake” is much less energy intensive and therefore preferred for large-scale NFTs that will underpin web3 and the metaverse. In addition, bringing computation closer to the end application through edge computing may also reduce the carbon footprint of the computation required for immersive experiences in the metaverse. We expect to see the term “*green metaverse”* more in the future.

Challenges to be dealt with by governments

Challenges to be dealt with by governments include fake news & deep fakes, the relation between people and machines, copyright, the concentration of data & wealth and increasing inequality, and AI for defense and warfare. To prevent or mitigate those challenges, government actions are required. Companies cannot deal satisfactory with those challenges at an individual level.

Fake news and deep fakes. Artificial intelligence is relevant in two ways for fake news and deep fakes. Firstly, by using large language models and deep learning it is respectively possible to automatically generate news from some keywords and fake videos of people saying things they haven’t said. If there is a malicious intent, large amounts of fake news and deep fakes can be produced with little effort. Secondly, recommendation algorithms can spread fake news and deep fakes in a viral way thereby reaching millions of users. In the metaverse, this risk is amplified because the immersive experience makes it harder to distinguish between real and fake content.

The relation between people and machines. Can a machine be your boss? Can people marry robots? To what extent are robots allowed to take care of elderly people? Is there a need for rules to regulate the relation between people and machines? All those questions, and similar, become even more important in the metaverse because of avatars. Can avatars engage in a formal, legal relationship? Can one avatar be held liable for an action against another avatar?

Copyright. While the notion of copyright is a legal concept, it resonates much with the society at large as was evidenced by the furious complaints of artists in an art contest

that was won with important help from an AI system[10]. Who has the copyrights of a piece of art generated by an AI system such as a song, a painting generated by Dalle2[11], text by LaMDA[12] or even a piece of programming code by Copilot[13]? The metaverse will consist of millions of lines of code and will host millions of pieces of content. Generated by people, by AI?

The concentration of data & wealth and increasing inequality. Artificial intelligence is currently not helping to reduce the gap between the rich and the poor, both at the level of countries, and within countries. A few big companies are concentrating a significant share of all the digital data in the world and use this data to train and build the best AI algorithms. At the macro level, productivity and wealth are increasing, but the benefits are not fairly distributed. The risk exists that this trend will increase with the metaverse, moreover, the better situated citizens will be able to take earlier advantage of the metaverse then the less well-situated citizens.

AI for defense and warfare. Killer robots or lethal autonomous weapon systems are delicate topics in debates about AI for defense and warfare. Governments and international organizations are struggling with this debate as there is a risk of an AI sarms race. But what will be the role of the metaverse in warfare? While it is not (yet) called war, many battles between nations are already taking place in cyberspace. How will that work out in the metaverse? Could or should we imagine a war fought in the metaverse? As a digital addition to a traditional war? As a fully independent war?

5.1.2 Indirect challenges

Indirect challenges refer to problems that start surfacing after some time, as a secondary effect, mostly unforeseen at the time of market launch. The most well-known, all due to social media, include the increase of polarization in societies, the increase of anorexia and bigorexia in adolescents through social pressure, and technology addiction to gaming or social media apps.

Increase of polarization in societies. Recommendation algorithms can create filter bubbles where people only see what they are interested in, reinforcing their own thinking, and shutting them off from alternative perspectives. As we have seen in some elections and other important democratic events, AI algorithms can make the difference when elections are a close race. Given the immersive experience of the metaverse, this risk is likely to increase.

Technology addiction to gaming or social media apps. Some recommendation algorithms are so good that users are not able to disconnect from the content offered

[10] https://edition.cnn.com/2022/09/03/tech/ai-art-fair-winner-controversy/index.html
[11] https://openai.com/dall-e-2/
[12] https://blog.google/technology/ai/lamda/
[13] https://github.com/features/copilot

by them. Specific examples include Instagram and Tiktok which present video after video perfectly personalized to the interests of each user. Some, especially young, persons may become addicted to those social media applications. They can spend many hours a day on them and when they finally disconnect, they feel anxiety and they need to reconnect as soon as possible. Gaming is providing a similar challenge, when artificial intelligence optimizes the experience so much, that people cannot stop playing and when they do, they feel unhappy. While this may happen to any person, the risk is most severe for young adolescents. Again, due to the immersive experience of the metaverse, this risk is likely to increase[14].

The increase of anorexia and bigorexia in adolescents through social pressure. Connected to the previous risk on addiction to social media, those applications can also exercise profound social pressure and creating a distorted experience of reality. The combination of hours of exposure and the social pressure of being physically perfect, is causing already an increase in anorexia and bigorexia in young adults, who are vulnerable for those risks to occur. In the metaverse, this risk is likely to increase because people might more easily dissociate from their physical image and start preferring their avatar image.

5.2 New challenges

The metaverse will also introduce new challenges[15,16], although some of them are an extension of the challenges we are already facing with AI, they acquire a new dimension in the metaverse. While we cannot foresee all those challenges in advance, some will probably include:

5.2.1 Health challenges, including mental ones

- We repeat technology addiction here because one shouldn't underestimate the power of engagement (recommendation) algorithms in an immersive environment like the metaverse. Recommendations can be so good that, beyond a certain point, people may find it (too) hard to leave the metaverse when they should/want.

- The obsessive use of the metaverse to escape from reality: people who are not happy in the real world, may find an attractive alternative in the virtual world where they can be who they want to be. Rather than being stimulated to

[14] https://www.sciencedirect.com/science/article/pii/S0261517721001138
[15] https://www.makeuseof.com/reasons-to-worry-about-facebook-metaverse/
[16] https://www.makeuseof.com/metaverse-worsen-effects-social-media/

improve their real life, they escape from it, perceiving real life increasingly worse.

- Children are especially vulnerable for immersive technologies as they are more likely to confuse reality with a virtual world.

- Cyberbullying will probably increase[17] and have a larger negative impact through the enhanced digital, immersive experience, which is almost perceived as reality.

- Post-VR sadness[18] – the real world becomes disappointing, and people experience feelings of sadness.

- Virtual reality “hangovers” or cybersickness, denoting feelings of nausea, fatigue, dizziness, and bodily disorientation[19].

5.2.2 Privacy challenges

- Extended reality technology has an even bigger capacity of capturing personal data such as locations, movements, reflexes, eye movements and voice patterns. What happens with this information and how can users remain in control?

- Targeted ads and the associated data privacy concern will skyrocket and can become very intrusive.

5.2.3 Liberty challenges

- Find the right balance between freedom of expression and avoiding malicious content to be published and spread in the metaverse.

- As we have seen, a key element of the metaverse is mixed reality (MR), a blend of the digital and real worlds by using virtual reality (VR) and augmented reality (AR) technologies. Eventually, this blend may become so immersive and pervasive that people's virtual and real lives become tied together and indistinguishable. If this happens, whoever controls (a significant part of) the metaverse could control a significant portion of reality.

---

[17] Metaverse beyond the hype: https://www.sciencedirect.com/science/article/pii/S0268401222000767
[18] https://www.theatlantic.com/technology/archive/2016/12/post-vr-sadness/511232/
[19] The Ethics of Virtual Reality Technology: https://link.springer.com/article/10.1007/s11948-017-9979-y

5.2.4 Dual world challenges

- Are there acts considered illegal in the real world, that are allowed in the metaverse? What are those acts and who should regulate them?

- Spending hours and hours in virtual worlds, may cause people to become less sensitive for consequences of actions in the real world with all potential negative consequences (dissociation).

- As we saw, a decentralized autonomous organization (DAO) is a computer program that incorporates governance and decision-making rules that work on top of a decentralized network (blockchain). A DAO provides a system that transfers those rules to a smart contract, so that there is no need to trust people and institutions (there is no such thing as a governing body in a DAO), avoiding potential manipulation or corruption. However, although in theory it seems like an efficient system without the need to relegate a decision to a committee or to specific people, there are potential risks such as buying of votes to bypass the programmed rules, or potential bias being encoded in smart contracts that could end up harming the organization. DAOs such as Decentraland[20] do not even have a company behind it, creating uncertainty in how to virtual and real world interact in case of problems.

5.2.5 Equality challenges

- Expensive hardware may exclude people with less purchasing power enhancing the digital divide.

- Access to the metaverse and its virtual worlds should be fair in the sense that commissions of the creators should not be abusive.

- There is an accessibility and inclusiveness risk if the metaverse is only developed for the majority, leaving out specific groups such as older people who are less digital savvy (digital divide) or people with limited sight or hearing.

5.3 Challenges stemming from a deliberate malicious use of the metaverse

So far, we have discussed challenges of the metaverse stemming from legitimate intentions. However, there are also challenges because of deliberate malicious use of the metaverse. While most of those acts are forbidden by law, they still happen and are likely to increase with the metaverse.

[20] https://decentraland.org/

- Impersonation. In the metaverse, you can access different virtual worlds with the same identity. The negative consequences of impersonation are therefore much more severe than is currently the case in digital services where each service has a different username.

- If your Identity is hijacked, you are more vulnerable to ransomware and extortion. With digital services, you can reset your password. However, in the metaverse you cannot simply change your avatar as it is directly connected to your virtual existence. And how will play deep fakes a role in identity theft?

- Theft of virtual properties.

- Should virtual aggressions in the metaverse be considered a crime? May this even imply physical damage in the real world, for instance, if using a haptic suit?

- Manipulation of the voting process of DAOs with malicious intent through voting buying.

5.4 Approaches to tackle the challenges and mitigate social and ethical risks

As currently is being experienced with AI, there are different ways to deal with the foreseen ethical and social risks of the metaverse: self-regulation, recommendations of international institutions or regulation. Notice that those are not necessarily mutually exclusive.

Self-regulation means that companies consider, detect, and mitigate the potential societal and ethical risks of their metaverse activity on a voluntary basis. In the AI world, the notion of a Responsible use of AI by Design is being adopted by increasingly more organizations. Ethical and societal risks are considered during the full lifecycle of the system, and, when detected, are mitigated through specific actions. If it is impossible to mitigate the foreseen risks, companies could and should refrain from deploying it until risks are better understood and can be avoided. Self-regulatory actions related to societal and ethical risks can be managed through ESG initiatives (Environment, Social and Governance) or codes of conduct. Based on the experience with AI and ESG, we believe that companies will start working also on the responsible use of the metaverse by design. Self-regulation is very important with new technology like the metaverse since there is little real experience of its impact in society.

Recommendations of international institutions and public-private partnerships. Relevant stakeholders of several nations can come together to discuss how technology can form a societal or ethical risk and issue recommendations to companies and governments for dealing with them. Examples of such initiatives for AI include the

OECD AI principles[21] and UNESCO's AI recommendation[22]. Individual organizations can then adhere to the guidelines. We strongly emphasize the importance of such interdisciplinary and global debates on the social and ethical risks of the metaverse, including public-private partnerships. Those debates are an essential input of potential regulation.

Regulation means that some uses of the technology are governed by law. The GDPR (European data protection regulation[23]) and the AI Act (upcoming European AI regulation[24]) are examples. A risk-based approach to regulation implies that the higher the risk, the more rules apply. Given the potential far-reaching impact of the metaverse, it would be wise to start thinking about important risks we -as societies- want to surely avoid. There is a risk if regulation of the metaverse comes too early when there is still little real experience with its social and ethical impact. As said, interdisciplinary, global debates are essential here.

6. Conclusion

Everybody talks about the metaverse, but today it is still more a promise then a reality. In this paper, we have shown that the metaverse is not just appearing out of the blue but is a logical consequence of several important evolutions. We have explored potential social and ethical challenges of the metaverse based on a similar experience with artificial intelligence.

Few people doubt that the metaverse will have a significant impact on our lives, society, and economies. Given that expectation, it is important to think in advance (now) about potential social and ethical challenges, and design strategies to mitigate them. Only in this way, humanity will be able to fully enjoy the opportunities, while at the same time reduce potential, negative consequences. We can and should learn from the experience with artificial intelligence, a transformational technology with also a huge impact on society, whose use needs to be managed carefully. As we have seen, there are currently many initiatives underway to ensure a good use of artificial intelligence, including self-regulation, international recommendations, and regulations.

Nobody knows how the potential social and ethical risks of the metaverse will play out. But companies that have started to implement the ethical use of AI, will be in a good position to meet them head on. In Telefonica, we are using a methodology called "the responsible use of AI by design", which consists of AI principles[25], awareness and

[21] https://oecd.ai/assets/files/OECD-LEGAL-0449-en.pdf
[22] https://unesdoc.unesco.org/ark:/48223/pf0000381137
[23] https://gdpr-info.eu/
[24] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206
[25] https://www.telefonica.com/wp-content/uploads/sites/7/2021/11/principios-ai-eng-2018.pdf

training to employees, a questionnaire with the right questions to ask, technical tools and a governance model[26] that defines roles and responsibilities. We have identified a new role called the *responsible AI champion*[27] who is the go-to person for questions related to the ethical use of AI. As part of our governance model, we have created an AI ethics committee consisting of multidisciplinary experts. We also have an agreement with UNESCO with regards to their Recommendation on the Ethics of Artificial Intelligence[28]. It is our intention to use a similar approach for the metaverse, leveraging also Telefónica's ample experience in cybersecurity and privacy by design.

6.1 Towards values and ethical principles of the metaverse

Given the many open challenges associated to the metaverse, there is a need for values and ethical principles now -before it is growing big- to ensure a fair metaverse in the future. To open the public debate of what those values and ethical principles should be, we suggest taking as a starting point the values identified in UNESCO's recommendation on the ethics of artificial intelligence[29], including:

- Respect, protection and promotion of human rights and fundamental freedoms and human dignity**.**
- Environment and ecosystem flourishing
- Ensuring diversity and inclusiveness
- Living in peaceful, just and interconnected societies

Ethical principles for a just metaverse will probably include notions such as: fairness, equality, responsibility & accountability, privacy, security, safety, green & sustainability, children & other vulnerable groups, do-not-harm business models, transparency, inclusiveness, and liberty.

Welcome to the debate!

[26] https://www.telefonica.com/es/wp-content/uploads/sites/4/2021/06/ia-responsible-governance.pdf
[27] https://business.blogthinkbig.com/a-new-organizational-role-for-artificial-intelligence-the-responsible-ai-champion/
[28] https://www.telefonica.com/es/sala-comunicacion/la-unesco-y-telefonica-se-comprometen-a-promover-impulsar-e-implementar-la-recomendacion-sobre-la-etica-de-la-inteligencia-artificial/
[29] https://unesdoc.unesco.org/ark:/48223/pf0000380455

7. Acknowledgements

We would like to thank the following people for their feedback on earlier versions: Francisco Montalvo, Joaquina Salado, Eusebio Felguera, Irene Bernal, and Silvia Díaz.

www.telefonica.com